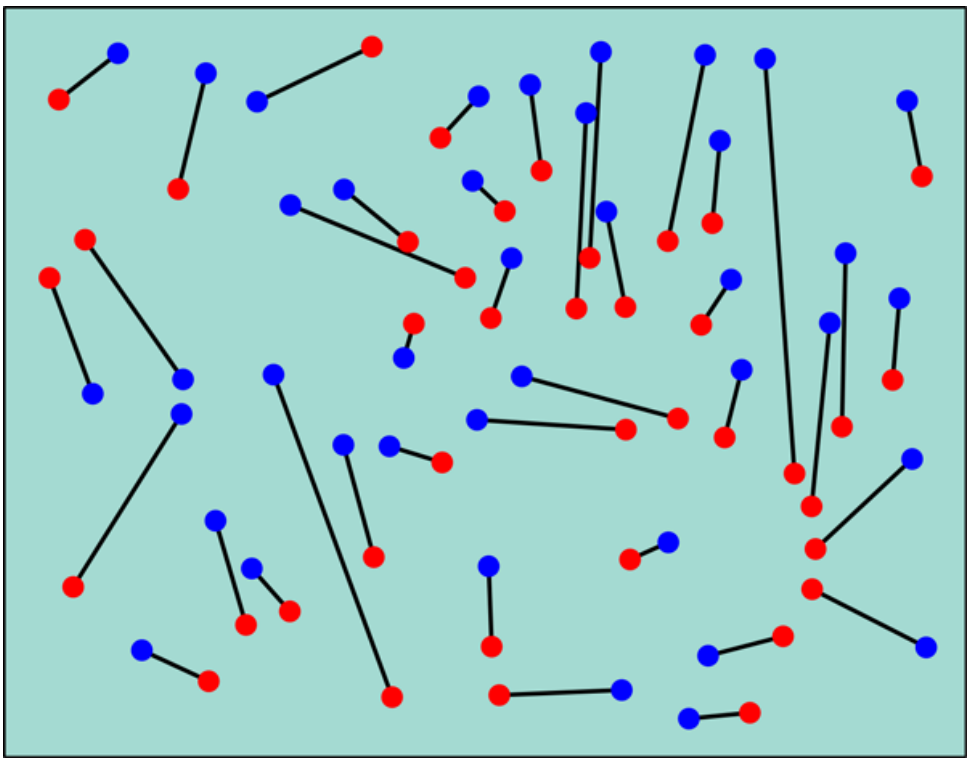
CS598
Topics in Graph Algorithms

Pingbang Hu

September 27, 2024

Abstract

This is an advanced graduate-level graph algorithm course taught by Chandra Chekuri at University of Illinois Urbana-Champaign.



This course is taken in Fall 2024, and the date on the cover page is the last updated time.

Contents

Chapter 1

Introduction

Lecture 1: Overview

Throughout the course, we consider a graph $G = (V, E)$ such that $n := |V|$ and $m := |E|$. Let's see some examples about the recent breakthroughs. 27 Aug. 11:00

Example (Shortest paths with negative length). The classical algorithm runs in $O(mn)$. In 2022, [BNW] came up with an algorithm $O(m \log^3 nC)$, where C is the largest absolute value of the *integer* length.

This is not a strongly polynomial time algorithm. In 2024 [Fineman] come up with $\widetilde{O}(mn^{8/9})$, and soon after 2024 [HJQ] improve this to $\widetilde{O}(mn^{4/5})$.

cite

cite

cite

Example (s-t max-flow). The tradition running time is $O(mn \log m/n)$, and it's later improved to be $O(m\sqrt{n} \log nC)$. Recently, [Chen et-al] improve to $O(m^{1+o(1)})$, which is almost-linear.[a]

cite

[a] This can be also applied to min-cost flow and quadratic-cost flow.

1.1 Minimum Spanning Tree

Finding the minimum cost spanning tree (MST) in a connected graph is a basic algorithmic problem that has been long-studied. We introduce the problem formally.

Definition 1.1.1 (Spanning tree). A *spanning tree* T of a connected graph $G = (V, E)$ is an induced subgraph of G which spans G, i.e., $V(T) = V$ and $E(T) \subseteq E$.

Then, the problem can be formalized as follows.

Problem 1.1.1 (Minimum spanning tree). Given a connected graph $G = (V, E)$ with edge capacity $c \colon E \to \mathbb{R}_+$, find the min-cost spanning tree.

Remark. The edge costs need not be positive, but we can make them positive by adding a large number without affecting correctness.

Standard algorithm that are covered in most undergraduate courses are Kruskal's algorithm, Jarnik-Prim's (JP) algorithm,[1] and (sometimes) Borůvka's algorithm. There are many algorithms for MST and their correctness relies on two simple rules (structural properties). The first one is about cuts:

Lemma 1.1.1 (Cut rule). If e is a minimum cost edge in a cut $\delta(S)$ for some $S \subseteq V$, then e is in some MST. In particular, if e is the unique minimum cost edge in the cut, then e is in every MST.

[1] This is typically attributed usually to Prim but first described by Jarnik

Definition 1.1.2 (Light). An edge e is *light* or *safe* if there exists a cut $\delta(S)$ such that e is the cheapest cost edge crossing the cut. We also say that e is *light* w.r.t. a set of edges $F \subseteq E$ if e is light in (V, F).

Another one is about cycles:

Lemma 1.1.2 (Cycle rule). If e is the highest cost edge in a cycle C, then there exists an MST that does not contain e. In particular, if e is the unique highest cost edge in C, then e cannot be in any MST.

Definition 1.1.3 (Heavy). An edge e is *heavy* or *unsafe* if there exists a cycle C such that e is the highest cost edge in C. We also say that e is *heavy* w.r.t. a set of edges $F \subseteq E$ if e is heavy in (V, F).

Corollary 1.1.1. Suppose the edge costs are unique and G is connected. Then the MST is unique and consists of the set of all light edges.

Remark. Without loss of generality, we can assume that the cost are unique by, e.g., perturbation or consistent tie-breaking rule.

1.1.1 Standard Algorithms

Let's review the basic algorithms, the data structures they use, and the run-times that they yield.

Kruskal's Algorithm

Intuitively speaking, Kruskal's algorithm sorts the edges in increasing cost order and greedily inserts edges in this order while maintaining a maximal forest F at each step. When considering the i^{th} edge e_i, the algorithm needs to decide if $F + e_i$ is a forest or whether adding e creates a cycle.

Algorithm 1.1: Kruskal's Algorithm

Data: A connected graph $G = (V, E)$ with edge capacity $c\colon E \to \mathbb{R}_+$
Result: A MST $T = (V, F)$

1 Sort the edges such that $c(e_1) \leq c(e_2) \leq \cdots \leq c(e_m)$ // $O(m \log n)$[a]
2 $F \leftarrow \varnothing$ // Initialize the tree
3 **for** $i = 1, \ldots, m$ **do**
4 **if** $e_i + F$ *has no cycle* **then**
5 $F \leftarrow F + e_i$
6 **return** (V, F)

[a] Since the graph is connected, $O(m \log m) = O(m \log n)$ as $n/2 \leq m \leq n^2$.

Theorem 1.1.1. Kruskal's algorithm takes $O(m \log n)$.

Proof. Sorting takes $O(m \log n)$ time. The standard solution for line 4 is to use a union-find data structure. Union-find data structure with path compression yields a total run time, after sorting, of $O(m\alpha(m, n))$ where $\alpha(m, n)$ is inverse Ackerman function which is extremely slowly growing. Thus, the bottleneck is sorting, and the run-time is $O(m \log n)$. ■

Jarnik-Prim's Algorithm

Jarnik-Prim's algorithm grows a tree starting at some arbitrary root vertex r while maintaining a tree T rooted at r. In each iteration it adds the cheapest edge leaving T until T becomes spanning. Thus, the Jarnik-Prim's algorithm takes $n - 1$ iterations.

Algorithm 1.2: Jarnik-Prim's Algorithm

Data: A connected graph $G = (V, E)$ with edge capacity $c\colon E \to \mathbb{R}_+$
Result: A MST $T = (V, F)$

```
1 r ←uniform(V)                                          // Sample a root
2 V' ← {r}, F ← ∅                                        // Initialize the tree
3 while V' ≠ V do
4 |   e ← arg min_{e=(u,v)∈δ(V'),u∈V'} c(e)
5 |   F ← F + e, V' ← V' + v                             // Update the tree
6 return (V, F)
```

Theorem 1.1.2. Jarnik-Prim's algorithm takes $O(m + n \log n)$.

Proof. To find the cheapest edge leaving T (line 4), one typically uses a priority queue where we maintain vertices not yet in the tree with a key for v equal to the cost of the cheapest edge from v to the current tree. When a new vertex v is added to T the algorithm scans the edges in $\delta(v)$ to update the keys of neighbors of v. Thus, one sees that there are a total of $O(m)$ `decrease-key` operations, $O(n)$ `extract-min` operations, and initially we set up an empty queue. Standard priority queues implement `decrease-key` and `extract-min` in $O(\log n)$ time each, so the total time is $O(m \log n)$. However, Fibonacci heaps and related data structures show that one can implement `decrease-key` in amortized $O(1)$ time which reduces the total run time to $O(m + n \log n)$. ■

Remark. The Jarnik-Prim's algorithm runs in linear-time for moderately dense graphs!

Borůvka's Algorithm

Borůvka's algorithm seems to be the first MST algorithm, which has very nice properties and essentially uses no data structures. The algorithm works in phases. We describe it recursively to simplify the description, while refer to Algorithm 1.3 for the real implementation. In the first phase the algorithm finds, for each vertex v the cheapest edge in $\delta(v)$. By the cut rule this edge is in every MST.

Note. An edge $e = uv$ may be the cheapest edge for both u and v.

The algorithm collects all these edges, say F, and adds them to the tree. It then shrinks the connected components induced by F and recurses on the resulting graph $H = (V', E')$. It's easy to see that Borůvka's algorithm can be parallelized, unlike the other two algorithms.

Algorithm 1.3: Borůvka's Algorithm

Data: A connected graph $G = (V, E)$ with edge capacity $c\colon E \to \mathbb{R}_+$
Result: A MST $T = (V, F)$

```
1  F = ∅                                                 // Initialize the tree
2  S ← {S_v = {v}}                                       // Collection of all sets
3  while |S| > 1 do
4  |   S' ← S                                           // Make a copy
5  |   for S ∈ S do
6  |   |   e_S = (u, v) ← arg min_{e∈δ(S)} c(e)
7  |   |   S' ← S' − {S_u, S_v} + S_u ∪ S_v             // Merge (i.e., shrink)
8  |   |   F ← F + e_S                                 // Update the tree
9  |   S ← S'                                          // Update S
10 return (V, F)
```

Notation. In line 7, S_u and S_v both refer to $S := S_u \cup S_v$ later in the algorithm.

Theorem 1.1.3. Borůvka's algorithm takes $O(m \log n)$.

Proof. The first phase needs $O(m)$ from a linear scan of the adjacency lists, and also computing H (i.e., shrinking) can be done in $O(m)$ time. The main observation is that $|V'| \leq |V|/2$ since each vertex v is in a connected component of size at least 2 as we add an edge leaving v to F. Thus, the algorithm terminates in $O(\log n)$ phases for a total of $O(m \log n)$ time. ■

1.1.2 Faster Algorithms

A natural question is whether there is a linear-time, i.e., $O(m)$, MST algorithm. The following is the history of fast MST algorithms:

- Very early on, Yao, in 1975, obtained an algorithm that ran in $O(m \log\log n)$[Yao75], which leverages the idea developed in 1974 for the linear-time Selection algorithm.
- In 1987, Fredman and Tarjan [FT87] developed the Fibonacci heaps and give an MST algorithm which runs in $O(m \log^* n)$.[2] This was further improved to $O(m \log\log^* n)$ [Gab+86].
- Karger, Klein, and Tarjan [KKT95] obtained a linear time randomized algorithm that will be the main topic of this lecture.
- Chazelle's algorithm [Cha00] that runs in $O(m\alpha(m, n))$ is the fastest known deterministic algorithm.

Note. Pettie and Ramachandran gave an optimal deterministic algorithm in the comparison model without known what its actual running time is [PR02]!

Perhaps an easier question is the following.

Problem 1.1.2 (MST verification). Given a graph G and a tree T, decide T is an MST of G or not.

One can always use an MST algorithm to solve the verification problem, but not necessarily the other way around. Interestingly, there is indeed a linear-time MST verification algorithm based on several non-trivial ideas and data structures and was first developed in the RAM model by Dixon, Rauch, and Tarjan [DRT92] with insights from Komlós [Kom85]. Simplification is done by King [Kin97].

Note (RAM model). The RAM model allows bit-wise operation on $O(\log n)$ bit words in $O(1)$ time.

Theorem 1.1.4 (MST verification). There is a linear-time MST verification algorithm in the RAM model. In fact, the algorithm is based on a more general result that we will need: Given a graph $G = (V, E)$ with edge costs and a spanning tree $T = (V, F)$, there is an $O(m)$-time algorithm that outputs all the F-heavy edge of G.

Proof. The original complicated algorithm has been simplified over the years. See lecture notes of Gupta and Assadi for accessible explanation, also the MST surveys [Eis97; Mar08]. ■

Fredman-Tarjan's Algorithm

Here we briefly describe Fredman and Tarjan's algorithm [FT87; Mar08] via Fibonacci heaps, which is reasonably simple to describe and analyze modulo a few implementation details that we will gloss over for the sake of brevity. First, we develop a simple $O(m \log\log n)$ time algorithm by combining Borůvka's algorithm and Jarnik-Prim's algorithm.

As previously seen. Jarnik-Prim's algorithm takes $O(m + n \log n)$ time via Fibonacci heaps where the bottleneck is when $m = o(n \log n)$. On the other hand, Borůvka's algorithm starts with a graph on n nodes and after i^{th} phases, reduces the number of nodes to $n/2^i$; each phase takes $O(m)$ times.

[2] Formally, it runs in $O(m\beta(m, n))$, where $\beta(m, n)$ is the minimum value of i such that $\log^{(i)} n \leq m/n$, where $\log^{(i)} n$ is the logarithmic function iterated i times. Since $m \leq n^2$, $\beta(m, n) \leq \log^* n$.

Intuition. Suppose we run Borůvka's algorithm for k phases and then run Jarnik-Prim's algorithm once the number of nodes is reduced. We can see that the total run time is $O(mk)$ for the k phases of Borůvka's algorithm, and $O(m + n/2^k \log n/2^k)$ for the Jarnik-Prim's algorithm on the reduced graph. Thus, if we choose $k = \log \log n$, we obtain a total run-time of $O(m \log \log n)$.

Tarjan and Fredman obtained a more sophisticated scheme based on the Jarnik-Prim's algorithm, but the basic idea is to reduce the number of vertices. The algorithm runs again in phases. We describe the first phase here.

Intuition (First phase). Start growing the tree. If the heap gets too big, we stop.

Consider an integer parameter t such that $1 < t \leq n$. Pick an arbitrary root r_1 and grow a tree T_1 via Jarnik-Prim's algorithm with a Fibonacci heap. We stop the tree growth when the heap size exceeds t for the first time or if we run out of vertices. All the vertices in the tree are marked as visited. Now pick an arbitrary, unmarked vertex as root $r_2 \in V - T$ and grow a tree T_2, and we stop growing T_2 if it touches T_1, in which case it merges with it, or if the heap size exceeds t or if we run out of vertices. The algorithm proceeds in this fashion by picking new roots and growing them until all nodes are marked.

Note. While growing T_2, the heap may contain previously marked vertices. It is only when the algorithm finds one of the marked vertices as the cheapest neighbor of the current tree that we merge the trees and stop.

It's easy to see that the first phase of Fredman-Tarjan algorithm correctly adds a set of MST edges F. After this, we simply shrink these trees and recurse on the smaller graph.

Algorithm 1.4: Fredman-Tarjan's Algorithm

Data: A connected graph $G = (V, E)$ with edge capacity $c\colon E \to \mathbb{R}_+$
Result: A MST $T = (V, F)$

```
V' ← V, F ← ∅                                              // Initialize the tree
while |V| > 1 do
    T ← Grow(G)                                            // First phase
    F ← F ∪ E(T)                                           // Update the tree
    Shrink G w.r.t. T, update V and E                      // Second phase
return (V', F)

Grow(G):
    V' ← ∅, F ← ∅, T ← (V', F)                             // Initialize the forest
    while V' ≠ V do
        r ← uniform(V − V')                                // Pick an unmarked vertex
        T' ← ({r}, ∅)                                      // Initialize a tree
        while |N(T')| < t or V(T') ∩ V' ≠ ∅ do
            Run one more step of Jarnik-Prim(r, T')        // Starting at r, maintaining T'
        V' ← V' ∪ V(T')                                    // Mark
        F ← F ∪ E(T')                                      // Update the forest by merging the tree
    return (V, F)                                          // Return a forest of G
```

Note. This can be seen as a parameterized version of Borůvka's algorithm.

The difficult part is to determine its runtime. We have the following.

Theorem 1.1.5. Fredman-Tarjan's algorithm takes $O(m\beta(m, n))$.

Proof. Firstly, the total time to scan edges and insert vertices into heaps and do `decrease-key` is $O(m)$ since an edge is only visited twice, once from each end point. Since each heap is not allowed to grow to more than size t, the total time for all the `extract-min` operations take $O(n \log t)$. With the fact that the initialization of each data structure is easy as it starts as an empty one, hence, the

first phase takes $O(m + n \log t)$. We claim that it also reduces the number of vertices to $2m/t$.

Claim. The number of connected components induced by F is $\leq 2m/t$ after the first phase.

Proof. Let $C_1, \ldots, C_h$ be the connected components of F. If for every C_i, $\sum_{v \in C_i} \deg(v) \geq t$,

$$2m = \sum_{v \in V} \deg(v) = \sum_{i=1}^{h} \sum_{v \in C_i} \deg(v) \geq ht \Rightarrow h \leq \frac{2m}{t}.$$

To see why the assumption holds, consider the growth of a tree T' in line 14:

- If we stop T' because heap size $|N(T')|$ exceeds t, then each of the vertex in the heap is a witness to a unique edge incident to T', hence the property holds.
- If T' merged with a previous tree, then the property holds because the previous tree already had the property and adding vertices can only increase the total degree of the component.

The only reason the property may not hold is if line 17 terminates a tree because all vertices are already included in it, but then that phase finishes the algorithm. ⊛

The question reduces to choosing t.

Intuition. We want linear time in the first phase, i.e., $n \log t$ to be no more than $O(m)$, leading to $t = 2^{2m/n}$. If we do this in every iteration, then this leads to $O(m)$ time per iteration.

We now bound the number of iteration. Consider $t_1 \coloneqq 2^{2m/n}$ and $t_i \coloneqq 2^{2m/n_i}$,[a] where n_i and m_i are the number of vertices and edges at the beginning of the i^{th} iteration, with $m_1 = m$ and $n_1 = n$. From the previous claim, $n_{i+1} \leq 2m_i/t_i$, which gives

$$t_{i+1} = 2^{2m/n_{i+1}} \geq 2^{\frac{2m}{2m_i/t_i}} \geq 2^{t_i}.$$

Thus, t_i is a power of twos with $t_1 = 2^{2m/n}$, and the Fredman-Tarjan's algorithm stops if $t_i \geq n$ since it will grow a single tree and finish. Thus, the algorithm needs at most $\beta(m, n)$ iterations, giving the total time $O(m\beta(m, n))$. ■

[a] Technically, we need to choose $t_i \coloneqq 2^{\lceil 2m/n_i \rceil}$, but we will be a bit sloppy and ignore the ceilings here.

Lecture 2: MST and Tree Packing

Linear-Time Randomized Algorithm

29 Aug. 11:00

Using randomization, it's possible to derive a linear-time algorithm for MST.

Theorem 1.1.6 ([KKT95]). Karger-Klein-Tarjan's algorithm takes $O(m)$ time that computes the MST with probability at least $1 - 1/\mathsf{poly}(m)$.

Karger-Klein-Tarjan's algorithm relies on the so-called sampling lemma, which we first discussed.

Lemma 1.1.3 (Sampling lemma). Given a graph $G = (V, E)$, and let $E' \subseteq E$ be obtained by sampling each edge e with probability $p \in (0, 1)$. Let F be a minimum spanning forest[a] in $G' = (V, E')$. Then the expected number of F-light edge in G is less than $(n - 1)/p$.

[a] As G' can be disconnected.

Proof. The proof is based on the *principle of deferred decisions* in randomized analysis. Let A be the set of F-light edges. Note that both A and F are random sets that are generated by the process of sampling E'. To analyze $\mathbb{E}[|A|]$, we consider Kruskal's algorithm to obtain F from E', where we generate E' on the fly:

Algorithm 1.5: Sampling Process

Data: A connected graph $G = (V, E)$ with edge capacity $c\colon E \to \mathbb{R}_+$, probability $p \in (0, 1)$
Result: A minimum spanning forest F and the set of F-light edges A

Sort the edges such that $c(e_1) \le c(e_2) \le \cdots \le c(e_m)$
$A \leftarrow \varnothing$, $F \leftarrow \varnothing$, $E' \leftarrow \varnothing$
for $i = 1, \ldots, m$ **do**
    $r \leftarrow$ `Ber(`p`)` `// Toss a biased coin`
    if $r = 1$ **then**
        $E' \leftarrow E' + e_i$
        if $F + e_i$ *is a forest* **then**
            $F \leftarrow F + e_i$
            $A \leftarrow A + e_i$
    else if e_i *is* F*-light* **then**
        $A \leftarrow A + e_i$
return F, A

The following is exactly the same as the above, but easier to analyze:

Algorithm 1.6: Sampling Process with Tweaks

Data: A connected graph $G = (V, E)$ with edge capacity $c\colon E \to \mathbb{R}_+$, probability $p \in (0, 1)$
Result: A minimum spanning forest F and the set of F-light edges A

Sort the edges such that $c(e_1) \le c(e_2) \le \cdots \le c(e_m)$
$A \leftarrow \varnothing$, $F \leftarrow \varnothing$
for $i = 1, \ldots, m$ **do**
    if e_i *is* F*-light* **then** `// Sorting implies` $F + e_i$ `is a forest` $\Leftrightarrow$ e_i `is` F`-light`
        $A \leftarrow A + e_i$
        $r \leftarrow$ `Ber(`p`)` `// Toss a biased coin`
        if $r = 1$ **then**
            $F \leftarrow F + e_i$
return F, A

The second algorithm makes the following observation clear.

Intuition. An edge e_i is added to A implies that it is added to F with probability p.

Hence, $p\mathbb{E}[|A|] = \mathbb{E}[|F|] \le n - 1$, hence $\mathbb{E}[|A|] \le (n-1)/p$. ■

With the sampling lemma, we know that when $p = 1/2$, the number of F-light edges from E is at most $2n$. Hence, we can eliminate most of the edges from $E \setminus E'$ from consideration given the fact that we can efficiently compute the F-heavy edges via the MST verification theorem. It's worth noting that to work with the sampling lemma via the natural recursion that it implies means that we need to work with potentially disconnected graph. That is, we will need to consider disconnected graph. Hence, we make the following generalization.

Definition 1.1.4 (Spanning forest)**.** A *spanning forest* T of a graph $G = (V, E)$ (potentially disconnected) is an induced subgraph of G which spans G, i.e., $V(T) = V$ and $E(T) \subseteq E$.

Problem 1.1.3 (Minimum spanning forest)**.** Given a graph $G = (V, E)$ (potentially disconnected) with edge capacity $c\colon E \to \mathbb{R}_+$, find the min-cost spanning forest.

Note. MST and MSF are closely related and one is reducible to the other in linear time, and the cut and cycle rules can be generalized to MSF easily.

Now, consider the following natural recursive divide and conquer algorithm for computing MSF.

Algorithm 1.7: Natural Recursive Algorithm from Sampling Lemma

Data: A graph $G = (V, E)$ with edge capacity $c\colon E \to \mathbb{R}_+$
Result: A MSF $T = (V, F)$

```
1 if |V| < n_0 then                                   // n_0 is some constant
2     return Standard-MST(G, c)              // Use a standard deterministic algorithm
3
4 Sample each edge i.i.d. from Ber(1/2) to obtain E_1 ⊆ E
5 (V, F_1) ←Karger-Klein-Tarjan((V, E_1))                  // Recursively compute MSF
6 E_2 ←Light-Edge(G, F_1)            // Compute all F_1-light edges with Theorem 1.1.4
7 (V, F_2) ←Karger-Klein-Tarjan((V, E_2))                  // Recursively compute MSF
8 return (V, F_2)
```

The correctness of Algorithm 1.7 is clear from the cut and cycle rules. The issue is the running time:

Claim. Algorithm 1.7 is not efficient enough.

Proof. The expected number of edges in $G_1 := (V, E_1)$ is $m/2$, and the expected number of edges in $G_2 := (V, E_2)$, via the sampling lemma, is at most $2n$. We see that the algorithm does $O(m+n)$ work outside the two recursive calls (line 5, line 7). Let $T(m, n)$ be the expected running time of the algorithm on an m-edge n-node graph. Informally, we see the following recurrence:

$$T(m, n) \le c(m + n) + T(m/2, n) + T(2n, n).$$

If we take the problem size to be $n+m$, then Algorithm 1.7 generates two sub-problems of expected size $m/2 + n$ and $2n + n$, with the total size being $4n + m/2$. If $m > 10n$, say, then the total problem size is shrinking by a constant factor, and we obtain a linear-time algorithm. However, this is generally not the case. ⊛

The problem becomes reducing the graph size, which is the trick of Karger-Klein-Tarjan's algorithm: we run Borůvka's algorithm for a few iterations as a preprocessing step, reducing the number of vertices:

Algorithm 1.8: Karger-Klein-Tarjan's Algorithm [KKT95]

Data: A connected graph $G = (V, E)$[a] with edge capacity $c\colon E \to \mathbb{R}_+$
Result: A MSF $T = (V, F)$

```
1  if |V| < n_0 then                                  // n_0 is some large constant
2      return Standard-MST(G, c)              // Use a standard deterministic algorithm
3
4  G' = (V', E'), T' = (V', F') ←Borůvka(G, c, 2)   // Run two iterations with |V'| ≤ |V|/4.
5
6  Sample each edge in G' i.i.d. from Ber(1/2) to obtain E_1 ⊆ E'
7  (V', F_1) ←Karger-Klein-Tarjan((V', E_1))                 // Recursively compute MSF
8  E_2 ←Light-Edge(G_1, F_1)                 // Compute F_2-light edges with Theorem 1.1.4
9  (V', F_2) ←Karger-Klein-Tarjan((V', E_2))                 // Recursively compute MSF
10 return (V, F' ∪ F_2)
```

[a] Assume no connected component of G is small.

Now, we provide the proof sketch of Theorem 1.1.6, which can be made precise with expectation.

Proof Sketch of Theorem 1.1.6. The correctness is easy to see as before. As for the running time, we see that Borůvka's algorithm takes $O(m)$ time for each phase, so the total time for the preprocessing (line 4) is $O(m)$. Then, the recurrence for $T(m, n)$ is

$$T(m, n) \le c(m + n) + T(m/2, n/4) + T(2n/4 + n/4),$$

i.e., the resulting sub-problem is of size $n/4 + m/2 + n/4 + n/2 = n + m/2$, which is good enough assuming $m \ge n - 1$.[a] By a simple inductive proof, we can show that $T(m, n) = O(n + m)$. ■

[a] Since we eliminate small components including singletons.

Remark. A more refined analysis of the sampling lemma can be used to show that the running time is linear with high probability as well.

Many properties of forests and spanning trees can be understood in the more general context of *matroids*. In many cases this perspective is insightful and also useful. The sampling lemma applies in this more general context and has various applications [Kar95; Kar98]. Obtaining a deterministic $O(m)$ time algorithm is a major open problem. Obtaining a simpler linear-time MST verification algorithm, even randomized, is also a very interesting open problem.

1.2 Tree Packing

We turn to another interesting problem, tree packing.

Problem 1.2.1 (Tree packing). Given a multigraph $G = (V, E)$, find all the edge-disjoint spanning trees in G. In particular, find the maximum number, $\tau(G)$, of edge-disjoint spanning trees of G

1.2.1 Bound on the Tree Packing Number

There is a beautiful theorem that provides a min-max formula for this. We first introduce some notation.

Notation. Let $\mathcal{P}$ be the collection of partitions of V, and E_P is the edge between connected components induced by a partition $P \in \mathcal{P}$, i.e., $e \in E_P$ if its endpoints are in different parts of P.

It's easy to see that any spanning tree must contain at least $|P| - 1$ edges from E_P. Thus, if G has k edge-disjoint spanning trees, then

$$k \leq \frac{|E_P|}{|P| - 1}.$$

More generally, we have the following.

Theorem 1.2.1. The maximum number of edge-disjoint spanning trees in a graph G is given by

$$\tau(G) = \left\lfloor \min_{P \in \mathcal{P}} \frac{|E_P|}{|P| - 1} \right\rfloor.$$

Remark. Theorem 1.2.1 is a special case of a theorem on matroid base packing where it is perhaps more natural to see [Sch+03].

A weaker version of the theorem is regarding fractional packing. In fractional packing, we allow one to use a fraction amount of a tree. The total amount to which an edge can be used is at most 1 (or $c(e)$ in the capacitated case). Clearly, an integer packing is also a fractional packing. The advantage of fractional packings is that one can write a linear program for it, and they often have some nice properties. Let $\tau_{\text{frac}}(G)$ be the fraction tree packing number. Clearly, we have $\tau_{\text{frac}}(G) \geq \tau(G)$.

Corollary 1.2.1. Given a graph G, we have

$$\tau_{\text{frac}}(G) = \min_{P \in \mathcal{P}} \frac{|E_P|}{|P| - 1}.$$

Proof. Assuming Theorem 1.2.1, then with $c \coloneqq |P^*| - 1$ for $P^* = \arg\min_{P \in \mathcal{P}} |E_P|/(|P| - 1)$,

$$\tau(G_c) - \min_{P \in \mathcal{P}} \frac{c|E_P|}{|P| - 1} = \lfloor |E_{P^*}| \rfloor - |E_{P^*}| = 0,$$

where G_c is with edge capacity scaled up by c. This implies that $\tau_{\text{frac}}(G_c) = \tau(G_c)$. As this holds for every c (with different graphs), this can only happen if $\tau_{\text{frac}}(G) = \min_{P \in \mathcal{P}} |E_P|/(|P| - 1)$. ■

The second important corollary that is frequently used is about the min-cut. We see that while the min-cut size $\lambda(G)$ of G is upper-bounding $\tau(G)$, i.e., $\tau(G) \leq \lambda(G)$, this is not tight at all.

Corollary 1.2.2. Let G be a capacitated graph and let $\lambda(G)$ be the global min-cut size. Then

$$\tau_{\text{frac}}(G) \geq \frac{\lambda(G)}{2} \frac{n}{n-1}.$$

Proof. Let P^* be the optimum partition that induces $\tau_{\text{frac}}(G)$. Then, $\tau(G) = |E_{P^*}|/(|P^*| - 1)$. Since for every connected component induced by P^*, at least $\lambda(G)$ edges are going out, hence

$$\tau_{\text{frac}}(G) = \frac{|E_{P^*}|}{|P^*| - 1} \geq \frac{\lambda(G)/2 \cdot |P^*|}{|P^*| - 1} \geq \frac{\lambda(G)}{2} \frac{n}{n-1},$$

where we use the fact that $|P^*| \leq n$ and $i/(i-1)$ is decreasing. ■

We first see a tight example.

Example (Cycle). Consider the n-node cycle C_n. Clearly, $\tau(C_n) = 1$, and $\tau_{\text{frac}}(C_n) \leq n/(n-1)$ since each tree has $n-1$ edges and there are n edges in the graph. Indeed, we have $\tau_{\text{frac}}(C_n) = n/(n-1)$. Finally, we see that $\lambda(G) = 2$.

Proof. Consider the n trees in C_n (corresponding to deleting each of the n edge) and assigning a fraction value of $1/(n-1)$ for each of them, with the corresponding tight partition consists of the n singleton vertices. ⊛

Note. Theorem 1.2.1 and its corollaries naturally extend to the capacitated case. For integer packing, we can assume c_e is an integer for each edge e, and the formula is changed to

$$\tau(G) = \left\lfloor \min_{P \in \mathcal{P}} \frac{c(E_P)}{|P| - 1} \right\rfloor.$$

Corollary 1.2.1 can also be proved in the same way when the edge capacity is rational.

Typically, one uses the connection between tree packing and min-cut to argue about the existence of many disjoint trees, since the global minimum cut is easier to understand than $\tau(G)$. However, we will see that one can use tree packing to compute $\lambda(G)$ exactly which may seem surprising at first due to the approximate relationship Corollary 1.2.2.

1.2.2 Proof of Corollary 1.2.1

Now, we give a different proof for Corollary 1.2.1 via LP duality without relying on Theorem 1.2.1.[3]

Proof of Corollary 1.2.1 [CQ17]. Consider $\mathcal{T}_G \coloneqq \{T \mid T \text{ is a spanning tree of } G\}$. Then, consider the following primal and the dual linear program:

$$\begin{aligned}
&\max \sum_{T \in \mathcal{T}_G} y_T && && \min \sum_{e \in E} c(e) x_e \\
&\sum_{T \ni e} y_T \leq c(e) && \forall e \in E; && \sum_{e \in T} x_e \geq 1 && \forall T \in \mathcal{T}_G; \\
&\text{(P)} \quad y_T \geq 0 && \forall T \in \mathcal{T}_G; && \text{(D)} \quad x_e \geq 0 && \forall e \in E.
\end{aligned}$$

Let y^* and x^* be the optimal solution to the primal and the dual. Then from the strong duality,

$$\sum_{T \in \mathcal{T}_G} y_T^* = \tau_{\text{frac}}(G) = \sum_{e \in E} c(e) x_e^*.$$

We see that if there exists e such that $x_e^* = 0$, then we can just contract all these edges, so without

[3] Indeed, this is a hard theorem to prove so we will not touch on this.

loss of generality, $x_e^* > 0$ for all $e \in E$.

Intuition. If $x_e^* = 0$, we can effectively increase $c(e)$ to ∞ without affecting the value of the dual solution, i.e., e is not a bottleneck in the primal tree packing, hence safe to contract.

Claim. If $x_e^* > 0$ for all $e \in E$, then $\tau_{\text{frac}}(G)$ is achieved via the singleton partition P. In particular,

$$\tau_{\text{frac}}(G) = \frac{\sum_{e \in E} c(e)}{n-1}.$$

Proof. From complementary slackness, we know that $\sum_{T \ni e} y_T^* = c(e)$ for all $e \in E$. Hence,

$$(n-1) \sum_{T \in \mathcal{T}_G} y_T^* = \sum_{T \in \mathcal{T}_G} \sum_{e \in T} y_T^* = \sum_{e \in E} \sum_{T \ni e} y_T^* = \sum_{e \in E} c(e),$$

implying that $\sum_{T \in \mathcal{T}_G} y_T^* = \sum_{e \in E} c(e)/(n-1)$. ⊛

Finally, we recall that $\tau_{\text{frac}}(G) \leq \min_P |E_P|/(|P|-1)$, hence, the above claim gives us the desired conclusion via induction: this is true if $x_e^* > 0$ for all $e \in E$; otherwise, we contract edges with $x_e^* = 0$ and reduce to this case. ■

Remark. In the above proof, the dual can be interpreted as a relaxation for the min-cut problem. In fact, if $x_e \in \{0, 1\}$, then this is exact.

1.2.3 Finding an Optimum Tree Packing and Approximating Tree Packing

If the linear program in the proof of Corollary 1.2.1 can be solved efficient to get $\tau_{\text{frac}}(G)$, then it will also yield an algorithm for the value of the integer packing $\tau(G)$ since it's just the floor of which. The problem is that while the primal has an exponentially many variables, the dual has an exponentially many constraints. We recall the following fact.

As previously seen. The Ellipsoid method needs a *separation oracle*. For example, applying it to the dual, we need to answer the following question efficiently:

- Given $x \in \mathbb{R}^E$, is it the case that $\sum_{e \in T} x_e \geq 1$ for all $T \in \mathcal{T}_G$?
- If not, find a tree T such that $\sum_{e \in T} x_e < 1$.

We see that this corresponds to solving MST, hence, the dual admits an efficient solution via the Ellipsoid method. One can convert an exact algorithm for the dual to an exact algorithm for the primal.

Remark. There are combinatorial algorithms for solving tree packing (both integer version and fraction versions) in strongly polynomial time [Sch+03].

On the other hand, we're also interested in whether we can find a faster algorithm for tree packing if one allows approximation. With an adaption of the *multiplicative weights update* (MWU) method and data structures for MST maintenance, there is a near-linear time algorithm:

Theorem 1.2.2 ([CQ17]). There is a deterministic algorithm to compute a $(1-\epsilon)$-approximate fractional tree packing in $O(m \log^3 n/\epsilon^2)$.

Lecture 3: Global Min-Cut with Tree Packing

1.3 Min-Cuts and Steiner Min-cuts via Tree Packing

3 Sep. 11:00

Consider the following famous problems about min-cuts.

Problem 1.3.1 (*s-t* min-cut)**.** Given a graph $G = (V, E)$ with edge capacity $c\colon E \to \mathbb{R}_+$, the *s-t min-cut* problem aims to find $\min_{S \subseteq V\colon s \in S, t \in V \setminus S} c(\delta(S))$.

Problem 1.3.2 (Global min-cut)**.** Given a graph $G = (V, E)$ with edge capacity $c\colon E \to \mathbb{R}_+$, the *global min-cut* problem aims to find $\min_{\emptyset \neq S \subsetneq V} c(\delta(S))$.

In what follows, we will simply use min-cut to refer to Problem 1.3.2 problem. A naive way to solve it is to first fix one end $s \in V$, and compute the *s-t* min-cut for all $t \in V - s$. Fairly recent work shows how one can do it with only poly-log max-flow computations.

Over the years, several very different algorithmic approaches have been developed for these problems. One of the surprising ones is based on MA-orderings [NI92], which is a combinatorial $O(mn + n^2 \log n)$ time algorithm that does not rely on flow at all.[4] Another approach is to combine several flow computations together via the push-relabel method [HO94], which also works for directed graphs. Karger developed elegant and powerful random contraction based algorithms for global min-cuts [Kar95], leading to many results. Two notable consequences are the following.

Theorem 1.3.1 ([KS96])**.** There is a randomized algorithm that runs in $O(n^2 \log n)$ time and outputs the min-cut with high probability.[a]

[a] This is a Monte-Carlo algorithm, so we cannot guarantee that the min-cut found is the correct one.

The following is a consequence of Karger's contraction algorithm [Kar95].

Theorem 1.3.2 (Approximate min-cut [Kar00])**.** The number of α-approximate min-cuts in a graph is at most $O(n^{2\alpha})$.

Karger then developed another approach via tree packing to obtain a randomized near-linear time algorithm for min-cut. He also was able to refine the bound on approximate min-cuts via this approach.

Theorem 1.3.3 ([Kar00])**.** There is a randomized algorithm that runs in time $O(m \log^3 n)$ and outputs the min-cut with high probability.

While the random contraction based algorithm is taught quite frequently due to its elegance and simplicity, the tree packing approach is more technical. More recently, the tree packing approach has led to several new results, which we now discuss.

1.3.1 Tree Packing-Based Algorithm for Min-Cut

Recall Corollary 1.2.2, which gives $\tau_{\text{frac}}(G) \in [\frac{\lambda(G)}{2} \frac{n}{n-1}, \lambda(G)]$. Intuitively, even if we can compute $\tau_{\text{frac}}(G)$ exactly, we have a 2-approximation to $\lambda(G)$. However, this already leads a crucial observation:

Intuition. On average, each tree can't cross the min-cut more than twice.

To formalize the above intuition, consider the following definition.

Definition 1.3.1 (Respecting)**.** Let $T = (V, E_T)$ be a spanning tree and $(S, V \setminus S)$ be a cut. The for an integer $h \geq 1$, we say T is *h-respecting* w.r.t. S if $|E_T \cap \delta(S)| \leq h$.

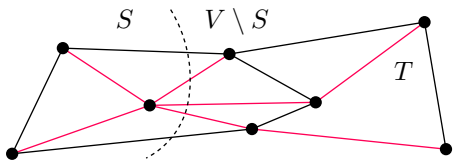


Figure 1.1: The spanning tree T is shown in red edges. T is 3-rspecting the cut $(S, V \setminus S)$.

[4] This approach generalizes to symmetric submodular functions.

We can now formalize the intuition in Lemma 1.3.1.

Lemma 1.3.1. Suppose $\{y_T\}_{T\in\mathcal{T}_G}$ is a $(1-\epsilon)$-approximate tree packing of G, and $\delta(S)$ is a min-cut of G. Let $\ell_T \coloneqq |E_T \cap \delta(S)|$ be the number of edges of T that cross the cut S. Furthermore, let $p_T = y_T / \sum_{T\in\mathcal{T}_G} y_T$ and $q \coloneqq \sum_{T:\,\ell(T)\le 2} p_T$. Then,

$$q \ge \frac{1}{2}\left(3 - \frac{2}{1-\epsilon}\left(1-\frac{1}{n}\right)\right).$$

In particular, if $\epsilon = 0$, then $1 \ge 1/2 + 1/n$, and if $\epsilon < 1/5$, then $q > 1/4$.

Proof. From the assumption, $\sum_{T\in\mathcal{T}_G} y_T \ge (1-\epsilon)\tau_{\text{frac}}(G)$. With Corollary 1.2.2, we have

$$\sum_{T\in\mathcal{T}_G} y_T \ge (1-\epsilon)\frac{n}{n-1}\frac{\lambda(G)}{2}.$$

Let $S \subseteq V$ be a min-cut, we have $1 = \sum_{T\in\mathcal{T}_G} p(T) = \sum_{T:\,\ell(T)\le 2} p_T + \sum_{T:\,\ell(T)\ge 3} p_T$. Observe that

- each tree T with $\ell(T) \ge 3$ uses up at least 3 edges from $\delta(S)$; while
- each tree T with $\ell(T) \le 2$ uses up at least 1 edge from $\delta(S)$.

Since the total capacity of $\delta(S)$ is $\lambda(G)$, and the tree packing solution is valid, we have

$$\sum_{T:\,\ell(T)\le 2} y_T + 3\sum_{T:\,\ell(T)\ge 3} y_T \le \lambda(G) \Rightarrow q + 3(1-q) \le \frac{\lambda(G)}{\sum_{T\in\mathcal{T}_G} y_T} \le \frac{2}{1-\epsilon}\left(1-\frac{1}{n}\right),$$

where the last inequality follows from the very first inequality we have derived. ■

Remark. Lemma 1.3.1 states that if the tree packing is sufficiently good, then a constant fraction of the trees in the packing will cross the min-cut at most twice.

Now, we're ready to see Karger's algorithm for min-cut [Kar00]. However, the original algorithm was more involved since at that time, there was no near-linear time approximation algorithm for tree packing, so he used a form of sparsification and then applied an approximation tree packing algorithm on the sparsified graph which is quite a feat. In our case, recall that following.

As previously seen. Theorem 1.2.2 states that we can compute a $(1-\epsilon)$-approximate tree packing of G, given by $\{y_T\}_{T\in\mathcal{T}_G}$, in $O(m\log^3 n/\epsilon^2)$ time.

By black-boxing this near-linear time tree packing algorithm, consider the following.

Algorithm 1.9: Tree Packing-Based Min-Cut Algorithm [Kar00; CQ17]

Data: A connected graph $G = (V, E)$ with edge capacity $c\colon E \to \mathbb{R}_+$, $\epsilon_0 \in < 1/5$
Result: A cut S

1 $\{y_T\}_{T\in\mathcal{T}_G} \leftarrow$ `Approximate-Tree-Packing`(G, c, ϵ_0) // $O(m\log^3 n)$
2 Sample a tree T with probability $p_T = y_T / \sum_{T\in\mathcal{T}_G} y_T$
3 Find the cheapest cut $(S, V\setminus S)$ in G such that T is 2-respecting w.r.t. S
4 **return** S

Firstly, we see that Algorithm 1.9 admits the following.

Lemma 1.3.2. Algorithm 1.9 outputs the min-cut of G with probability at least $1/4$.

Proof. It's immediate from Lemma 1.3.1. ■

To boost the success probability, we can simply repeat the last two steps (line 2, line 3) $\Theta(\log n)$ times, which results in a success probability to at least $1 - 1/n^c$ for any constant c. To analyze the running time, a key ingredient is line 3. Karger showed that one can implement line 3 via a clever dynamic programming coupled with link-cut tree data structure:

Theorem 1.3.4 ([Kar00]). Given a graph $G = (V, E)$ and a spanning tree $T = (V, E_T)$. There is a deterministic algorithm that computes a minimum cut $(S, V \setminus S)$ such that T is 2-respecting w.r.t. S in $O(m \log^2 n)$ time.

We can now prove Theorem 1.3.3.

Proof of Theorem 1.3.3. Since line 1 takes $O(m \log^3 n)$ for ϵ_0 being a constant, and observe that once the approximated tree packing $\{y_T\}_{T \in \mathcal{T}}$ is computed, we can reuse them and apply the repetition for line 2 and line 3 to boost the probability of success. With $\Theta(\log n)$ repetitions, we obtain an $O(m \log^3 n)$ time algorithm as desired with the running time guaranteed by Theorem 1.3.4. ■

1.3.2 Bounding the Number of Approximate Min-Cuts

As hinted in Theorem 1.3.2, we're now interested in how many distinct min-cuts can an undirected graph have. The following theorem was shown a long time ago:

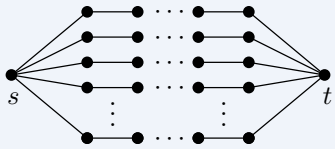
Theorem 1.3.5 ([DKL76]). The number of distinct min-cuts in an undirected graph is at most $\binom{n}{2}$.

Example (Cycle). The worst case example is an n-cycle C_n.

Remark. All the min-cuts of a graph can be represented in a nice and compact data structure called the cactus (cactus representation), which was also shown in [DKL76].

In contrast, for s-t min-cuts, it can be exponentially many in n.

Example. Consider the following multi-highway-like graph, which has exponentially many s-t min-cuts since if we choose one of the road section in each line of the road, it'll be a s-t min-cut.



Hence, we're interested in the number of α-approximation min-cut:

Definition 1.3.2 (Approximate min-cut). For $\alpha \geq 1$ an *α-approximate min-cut* is a cut $(S, V \setminus S)$ such that $c(\delta(S)) \leq \alpha \lambda(G)$.

Recall Theorem 1.3.2, where Karger used tree packing to prove that the number of α-approximation min-cuts is at most $O_\alpha(n^{\lfloor 2\alpha \rfloor})$. Before we prove Theorem 1.3.2, we recall some basic facts from linear programming.

As previously seen. A solution x^* to a linear program which has n non-trivial constraints means that the support size of x is at most n, i.e., $x_i > 0$ for at most n many i's.

We're now ready to prove Theorem 1.3.2, which is based on [CQX20].

Proof of Theorem 1.3.2. Consider an optimum fraction tree packing solution $\{(T, y_T^*)\}_{T \in \mathcal{T}_G}$. In the proof of Corollary 1.2.1, where we define the fractional tree packing linear program, we know that there are only m non-trivial constraints, hence there are only m many T's such that $y_T^* > 0$.

Consider an α-approximate min-cut $S \subseteq V$, and let $h = \lceil 2\alpha \rceil$. Now, let $q_{h,\alpha}$ be the fraction of tree packing that h-respects $S \subseteq V$, i.e.,

$$q_{h,\alpha} \coloneqq \sum_{T \colon \ell(T) \leq h} p_T.$$

Using a similar analysis as the one in Lemma 1.3.1, we can argue that

$$q_{h,\alpha} \geq \frac{1}{h}(1 - (2\alpha - \lfloor 2\alpha \rfloor)) \left(1 - \frac{1}{n}\right).$$

The main intuition is the following:

Intuition. Say at least one tree in the packing h-respects the cut (which is the case). Then, the total number of α-approximate min-cuts is at most $m \cdot n^h \leq m \cdot n^{\lfloor 2\alpha \rfloor}$.

But we can do better by noticing that $q_{h,\alpha} > 0$ is a fixed constant for any fixed α. Suppose N is the number of α-approximate min-cuts. For any fixed α-approximate min-cut, $q_{h,\alpha}$ fraction of the tree packing is h-respecting w.r.t. the cut. Consider the following question:

Problem. Fix a single tree T, how many distinct cuts are there such that T h-respects w.r.t.?

Answer. We can remove at most h edges from T to create at most $h+1$ components and combine these components into two sides of a cut, hence, each tree T correspond to at most $2^{h+1}\binom{n-1}{h} \leq 2^{h+1}n^h$ cuts. ⊛

Thus, the number of α-approximate min-cuts is at most $2^{h+1}n^h/q_{h,\alpha}$. ■

Lecture 4: Steiner Min-Cut with Isolating Cuts

1.3.3 Steiner Min-Cut

5 Sep. 11:00

Consider the following problem that generalizes the s-t min-cut and global min-cut.

Problem 1.3.3 (Steiner min-cut). Given a graph $G = (V, E)$ with edge capacity $c\colon E \to \mathbb{R}_+$ and a set $T \subseteq V$ of terminals, the *Steiner min-cut* problem aims to find the min-cut $(S, V \setminus S)$ which separates some pair of terminals, i.e., $S \cap T \neq \varnothing$ and $(V \setminus S) \cap T \neq \varnothing$.

Remark. Steiner min-cut generalizes both s-t min-cut and global min-cut.

Proof. s-t min-cut corresponds to $T = \{s, t\}$, while global min-cut corresponds to $T = V$. ⊛

A simple algorithm for the Steiner min-cut is the same as the global min-cut by solving s-t min-cut: for $T = \{t_1, \ldots, t_k\}$, fix a terminal, say t_1, then compute t_1-t_i min-cut for all $i \geq 2$. This requires $|T| - 1$ max-flow computations. In fact, this is the best known algorithm even for the global min-cut till [NI92].

Quite recently, a simple yet striking approach that computes the Steiner min-cut with high probability using only $O(\log^3 n)$ s-t cut computations is developed [LP20], which is based on isolating cut.

Submodular Function

The main interest here, i.e., solving isolating cut, will be essentially based on properties of symmetric submodular functions. Although we can prove various properties by appealing to only graph theoretic facts, it's useful to see the proofs via submodularity. Here, we give some background, and specifically, for the cut function of graphs.

Definition. Given a finite ground set V, consider a real-valued set function $f\colon 2^V \to \mathbb{R}$.

Definition 1.3.3 (Modular). The function f is *modular* if for all $A, B \subseteq V$,

$$f(A) + f(B) = f(A \cap B) + f(A \cup B).$$

Definition 1.3.4 (Submodular). The function f is *submodular* if for all $A, B \subseteq V$,

$$f(A \cap B) + f(A \cup B) \leq f(A) + f(B).$$

Definition 1.3.5 (Supermodular). The function f is *supermodular* if for all $A, B \subseteq V$,

$$f(A \cap B) + f(A \cup B) \geq f(A) + f(B).$$

Definition 1.3.6 (Posi-modular). The function f is *posi-modular* if for all $A, B \subseteq V$,

$$f(A - B) + f(B - A) \geq f(A) + f(B).$$

We note that perhaps a more common definition of submodularity is *diminishing marginal utility*, i.e., if $f(A + v) - f(A) \geq f(B + v) - f(B)$ for all $A \subseteq B$. Here, we see some examples.

Example (Modular function as weight function). f is modular if and only if there exists some $w \colon V \to \mathbb{R}$ such that $f(A) = \sum_{v \in A} w(v) + c$ for some shift c.

Example. If f and g are submodular, then so is $\alpha f + \beta g$ for some $\alpha, \beta \geq 0$.

One of the reason that submodularity is important for graphs is because of the following.

Example (Cut). Given a graph $G = (V, E)$, the cut size function $|\delta_G(\cdot)| \colon 2^V \to \mathbb{R}_+$ is submodular.

Proof. We simply note that for any $A, B \subseteq V$,

$$|\delta_G(A)| + |\delta_G(B)| = |\delta_G(A \cap B)| + |\delta_G(A \cup B)| + 2|E(A \setminus B, B \setminus A)| \geq |\delta_G(A \cap B)| + |\delta_G(A \cup B)|,$$

where $E(X, Y)$ is the set of edges crossing X and Y for some $X, Y \subseteq V$. ⊛

The above argument extends naturally to non-negative capacitied graph. Moreover, this is also true for directed graph.

Example. Let $G = (V, E)$ be a directed graph. $|\delta^+(\cdot)|$, and hence by symmetry $|\delta^-(\cdot)|$ are submodular.

We're also interested in the following property.

Definition 1.3.7 (Symmetric). A set function is *symmetric* if $f(A) = f(V \setminus A)$ for all $A \subseteq V$.

Clearly, $|\delta_G(\cdot)|$ is symmetric. However, for directed graph, this is not necessarily the case. Finally, we see that symmetric submodular function satisfies another important property.

Example. A symmetric submodular function is automatically posi-modular.

Now, we discuss uncrossing, a common and powerful technique that is frequently used in working with submodular functions. We illustrate this in the context of min-cuts.

Lemma 1.3.3. Let $G = (V, E)$ be a graph and $(A, V \setminus A)$, $(B, V \setminus B)$ be two s-t min-cuts. Then $(A \cap B, V \setminus (A \cap B))$ and $(A \cup B, V \setminus (A \cap B))$ are also s-t min-cuts.

Proof. From submodularity, we have $|\delta(A)| + |\delta(B)| \geq |\delta(A \cap B)| + |\delta(A \cup B)|$. However, as both $A \cup B$ and $A \cap B$ are themselves s-t cuts, all terms need to be equal. ■

Corollary 1.3.1. For any graph $G = (V, E)$, there is a unique (inclusion-wise) minimal s-t min-cut.[a]

[a] While maybe not that useful, from the same logic, there is a unique maximal s-t min-cut.

Proof. If there are two s-t min-cuts A, B that are both minimal and distinct, then $A \setminus B \neq \varnothing$ and $B \setminus A = \varnothing$ since otherwise one will be contained in the other, contradicting the minimality. From Lemma 1.3.3, $A \cap B$ is also a s-t min-cut and $A \cap B$ is a strict subset of A and B, again contradicting minimality of A and B. ■

The above proof applies to directed graph as well since we only used submodularity.

Remark (Graphic matroid). A second aspect of submodularity in graphs comes via *matroids*. We will not discuss it here but the rank function of a matroid is a special class of submodular functions; and in a formal sense, matroid rank functions are building blocks for all submodular functions. Given an undirected graph $G = (V, E)$ there is a fundamental matroid associated with the edge set of G called the *graphic matroid*.[a] Several properties of trees and forests can be better understood in the context of the graphic matroid including the Tutte-Nash-Williams theorem.

[a] There are other matroids that are also defined from graphs including the dual graphic matroid for instance.

Isolating Cuts via Poly-log Max-flow Computations

We can now formally introduce the isolating cut problem.

Problem 1.3.4 (Isolating cut). Given a graph $G = (V, E)$ with edge capacity $c\colon E \to \mathbb{R}_+$ and a set $T \subseteq V$ of terminals. The *t_i-isolating cut* problem aims to find a cut $(S_i, V \setminus S_i)$ such that $t_i \in S_i$ and $t_j \notin S_i$ (i.e., $t_j \in V \setminus S_i$) for all $j \neq i$.

Intuition. In other words, the cut isolates t_i from the rest of the terminals.

The minimum capacity t_i-isolating cut can be found by a single max-flow computation: by shrinking the terminals in $T - t_i$ into a single vertex s and computing the s-t_i min-cut. Thus, naively, computing all isolating cuts require k max-flow computations. The upshot is that this can be done in only $O(\log k)$ max-flow. Before we describe the algorithm, we first note that from submodularity, we also have a similar structural result, just like Corollary 1.3.1.

Lemma 1.3.4. There is a unique minimal t_i-isolating min-cut $(S_i^*, V \setminus S_i^*)$ such that if $(S_i, V \setminus S_i)$ is any t_i-isolating min-cut, then $S_i^* \subseteq S_i$.

We now describe the algorithm for computing the isolating cuts. Basically, we consider h bi-partitions $(A_1, B_1), \ldots, (A_h, B_h)$ with $h = \lceil \log k \rceil$, and compute a cut separating each bi-partition. Then, we take the intersection among the resulting cut sets, which will be isolating cuts as we will see. Finally, with the structural property Lemma 1.3.4, we can then find the minimum isolating cuts from them.

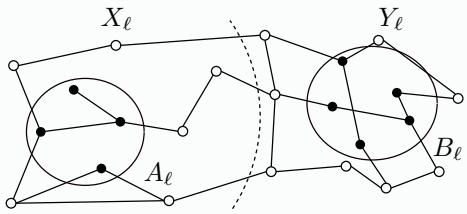
Algorithm 1.10: Isolating Cut [LP20] (also developed independently in [AKT21])

Data: A connected graph $G = (V, E)$ with edge capacity $c\colon E \to \mathbb{R}_+$, terminal $T = \{t_i\}_{i=1}^k$
Result: A set of isolating cuts $\{(S_i^*, V \setminus S_i^*)\}_{i=1}^k$ isolating t_i's

$h \leftarrow \lceil \log k \rceil$
for $\ell = 1, \ldots, h$ **do** // h bi-partitions (A_ℓ, B_ℓ) of T
  $A_\ell \leftarrow \{t_i \mid$ binary representation of i has 1 in ℓ^{th} bit$\}$
  $B_\ell \leftarrow T \setminus A_\ell$
  $(X_\ell, Y_\ell) \leftarrow$ `s-t-min-cut`$(G,\ A_\ell,\ B_\ell)$[a] // $Y_\ell = V \setminus X_\ell$
for $i = 1, \ldots, k$ **do**
  $S_i \leftarrow \bigcap_{\ell:\ t_i \in A_\ell} X_\ell \cap \bigcap_{\ell:\ t_i \in B_\ell} Y_\ell$
  $H_i = (V_i, E_i) \leftarrow$ shrinking $V \setminus S_i$ to a single vertex s_i
  $(S_i^*, V \setminus S_i^*) \leftarrow$ `s-t-min-cut`$(H_i,\ t_i,\ s_i)$
return $\{(S_i^*, V \setminus S_i^*)\}_{i=1}^k$

[a] This can be done via shrinking A_ℓ and B_ℓ in to two separate nodes, and compute the s-t min-cut.

Intuition. The following illustrates line 5, where terminals are black vertices. (A_ℓ, B_ℓ) is a bipartition of T, while (X_ℓ, Y_ℓ) is a min-cut that separates (A_ℓ, B_ℓ).



Additionally, line 7 is created by considering the intersections of all X_ℓ (or Y_ℓ) that includes t_i.

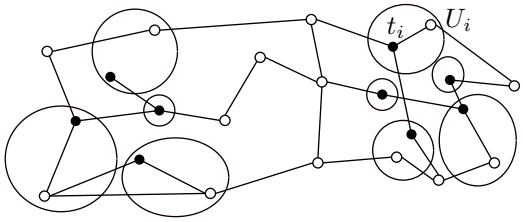
To see the correctness of the Algorithm 1.10, we want to say that s_i-t_i min-cut is exactly the minimum cost t_i-isolating cut in G. This is due to Lemma 1.3.5.

Lemma 1.3.5. For each i, $(S_i, V \setminus S_i)$ is a t_i-isolating cut. Furthermore, S_i's are pairwise disjoint.

Proof. Firstly, $t_i \in A_\ell \subseteq X_\ell$ or $t_i \in B_\ell \subseteq Y_\ell$, implying $t_i \in S_i$. Consider t_j with $j \neq i$. As $i \neq j$, there is some index ℓ in the binary representation of i and j differ in the bit position. Suppose i has 1 in the ℓ^{th} position and j has 0, then $t_i \in A_\ell$ and $t_j \in B_\ell$, implying $t_i \in X_\ell$ and $t_j \notin X_\ell$ as $t_j \in B_\ell \subseteq Y_\ell$ and $Y_\ell \cap X_\ell = \varnothing$. This means $t_j \notin S_i$.

We now prove that $S_i \cap S_j = \varnothing$ for all $i \neq j$. Firstly, there exists some ℓ such that $t_i \in A_\ell$ and $t_j \in B_\ell$ (or $t_i \in B_\ell$ and $t_j \in A_\ell$). Suppose $v \in X_\ell$, then v can't be in $S_j \subseteq Y_\ell$ and if $v \in Y_\ell$, then v can't be in $S_i \subseteq X_\ell$, hence v can't be in both S_i and S_j. ■

Lemma 1.3.5 gives the following picture, where each t_i lives in exactly one S_i.



Hence, for each i, we have a t_i-isolating cut $(S_i, V \setminus S_i)$. Now, Lemma 1.3.4 states that there is a t_i-isolating min-cut $(S_i^*, V \setminus S_i^*)$ where S_i^* is a subset of any t_i-isolating min-cut, it doesn't say it will be a subset of S_i in particular, as $(S_i, V \setminus S_i)$ is only a t_i-isolating cut. However, we do not lose anything:

Lemma 1.3.6. The minimal t_i-isolating min-cut $(S_i^*, V \setminus S_i^*)$ is in $(S_i, V \setminus S_i)$, i.e., $S_i^* \subseteq S_i$.

Proof. It suffices to prove that if $t_i \in A_\ell$ then $S_i^* \subseteq X_\ell$. Assume not, then $S_i^* \cap (V \setminus X_\ell) \neq \varnothing$. But since $S_i^* \cap X_\ell$ is a t_i-isolating cut, while S_i^* is the minimal t_i-isolating min-cut, $|\delta(S_i^* \cap X_\ell)| > |\delta(S_i^*)|$.

Moreover, it's trivial to see that $S_i^* \cup X_\ell$ is a A_ℓ-B_ℓ cut (not necessarily minimum, just a cut), hence we also have $|\delta(S_i^* \cup X_\ell)| \geq |\delta(X_\ell)|$. From submodularity of $|\delta(\cdot)|$, we have

$$|\delta(S_i^*)| + |\delta(X_\ell)| \geq |\delta(S_i^* \cap X_\ell)| + |\delta(S_i^* \cup X_\ell)|,$$

which is a contradiction. ■

With all the lemmas, it's now easy to see that Algorithm 1.10 is at least correct. Firstly, from Lemma 1.3.6, we know that there the optimal t_i-isolating min-cut $S_i^* \subseteq S_i$ (here, S_i^* is not necessary the one found by Algorithm 1.10: indeed, we're trying to argue this). As S_i's are disjoint, each terminal t_i lives in exactly one S_i, hence computing s_i-t_i min-cut will indeed recover S_i^*.

Intuition. When we contract $V \setminus S_i$, we do not lose the optimal isolating cut S_i^*.

Theorem 1.3.6 ([LP20]). Algorithm 1.10 is a deterministic algorithm that given $G = (V, E)$ and a terminal set $T \subseteq V$ with $|T| = k$, computes all the isolating cuts using $O(\log k)$ max-flow computations on graphs with $|V|$ vertices and $|E|$ edges each.

Proof. We analyze the runtime. It's easy to see that line 2 requires $O(\log k)$ max-flow computations on G. It's also easy to show that computing S_i's in line 7 can be done in $O((m+n)\log k)$ time given (X_ℓ, Y_ℓ) for $\ell \in [h]$. However, line 8 and line 9 seem to require k max-flow computations.

Claim. In total, line 9 only requires $O(\log k)$ max-flow computations.

Proof. Let us understand the size of H_i. It has $n_i + 1$ vertices where $n_i = |S_i|$, and it has m_i edges where $m_i = |E(S_i)| + |\delta(S_i)|$. Thus, the running time of max-flow on H_i is $T(n_i + 1, m_i)$ where $T(a, b)$ is the running time of max-flow on graph with a nodes and b edges. We observe that $\sum_i (n_i + 1) \leq 2n$ since S_i's are disjoint, while $\sum_i m_i \leq 2m$: consider any edge $uv \in E$. If $uv \in E(S_i)$ for some i, then it does not contribute to any other H_j. If $uv \in \delta(S_i)$ for some i, then it can be in $\delta(S_j)$ for only one more index $j \neq i$.

Thus, the total time to compute all k max-flows is $\sum_i T(n_i, m_i) \leq T(2n, 2m)$ under reasonable assumption, specifically, $T(a, b)$ is super-additive.[a] ⊛

[a] Formally, we first create a single H that includes each H_i as a copy in it, and we can run a single max-flow on H to recover all the max-flow values in each H_i. H will have $O(n)$ vertices and $O(m)$ edges.

With the correctness of Algorithm 1.10, the theorem is proved. ■

We see that this could have been discovered many years ago in terms of its simplicity. Algorithm 1.10 has been very influential in the last few years for a number of problems.

Note. Another perspective of the bi-partitions is that they are a way to *derandomize* a natural randomized algorithm that picks some $O(\log k)$ bi-partitions of T at random and computes the cuts between them. With high probability, every t_i, t_j with $i \neq j$ will be separated in at least on of the random bi-partitions.

Remark. The core idea of isolating cuts relies only on submodularity and symmetry, thus, this applies in much more generality and to several other problems. This is explicitly discussed in [CQ21], though the ideas are implicit in [LP20].

Randomized Algorithm for Steiner Min-Cut via Isolating Cuts

Isolating cut naturally lead to a simple randomized algorithm for Steiner min-cut. The basic idea is quite simple. Consider an optimum Steiner min-cut $(S, V \setminus S)$ and let $T_1 := S \cap T$ and $T_2 := (V \setminus S) \cap T$, with $k_1 = |T_1|$ and $k_2 = |T_2|$. We may assume that $1 \leq k_1 \leq k_2$.

Note. $(S, V \setminus S)$ is a t_i-t_j min-cut for any $i \neq j$ since otherwise, it induces a lower-cost cut.

The basic intuition is the following.

Intuition. If we can sample exactly one terminal in one side of the Steiner min-cut, then we can simply use the isolating cut to recover the Steiner min-cut.

Say we know k_1. We can sample each terminal in T independently with probability $1/k_1$ to obtain $T' \subseteq T$ such that with constant probability, $|T' \cap T_1| = 1$ and $|T' \cap T_2| \geq 1$ (recall $k_1 \leq k_2$). Suppose T' satisfies these properties and let $T' \cap T_1 = \{t_i\}$. Then, $(S, V \setminus S)$ is a minimum cost t_i-isolating cut w.r.t. T'. Hence, by computing t_i-isolating cuts for all $t_i \in T'$ and choosing the cheapest one identifies the Steiner min-cut for T.

The problem is that we don't know k_1, and trying all possible values for k_1 (from 1 to $k/2$) will be too expensive. The idea is that the above sampling procedure is robust: say if we sample with probability, say, $1/2k_1$, everything still happens with constant probability. Hence, we only need to try $k_i = 2^i$, i.e., $O(\log k)$ different sampling probabilities.

Algorithm 1.11: Steiner Min-Cut

Data: A connected graph $G = (V, E)$ with edge capacity $c\colon E \to \mathbb{R}_+$, terminal $T = \{t_i\}_{i=1}^k$
Result: A possible Steiner min-cut $(U^*, V \setminus U^*)$

$U^* \leftarrow \varnothing$ // Initialize Steiner min-cut
for $i = 0, \dots, \lceil \log k \rceil$ **do**
  $T' \leftarrow$ `Sample`$(T, 1/2^i)$ // Sample each terminal in T with probability $1/2^i$
  $\{(S_i^*, V \setminus S_i^*)\}_{i=1}^{|T'|} \leftarrow$ `Isolating-Cut`(G, c, T')
  $U^* \leftarrow$ `Min-Cost`$(\{(S_i^*, V \setminus S_i^*)\}_{i=1}^{|T'|} \cup \{(U^*, V \setminus U^*)\})$ // Update minimum cost cut
return $(U^*, V \setminus U^*)$

We now formally prove the robustness we have mentioned.

Lemma 1.3.7. Algorithm 1.11 finds the Steiner min-cut for T with a constant probability.

Proof. We see that for $k_1 = 1$, Algorithm 1.11 is correct (deterministically) since i can only be 0 and $T' = T$. Hence, let $k_1 > 1$. Consider the case that $1/2^{i+1} < 1/k_1 \le 1/2^i$, where i will be tried at some point during $i = 0$ to $\lceil \log k \rceil$ since $1 \le k_1 \le k/2$. Let $\ell = 2^i$, i.e., $\ell \le k_1 \le 2\ell$.

Now, let $\mathcal{E}_1$ be the event that $|T_1 \cap T'| = 1$, i.e., exactly one terminal from T_1 is chosen. Then

$$\Pr(\mathcal{E}_1) = k_1 \cdot \frac{1}{\ell} \cdot \left(1 - \frac{1}{\ell}\right)^{k_1 - 1} \ge \left(1 - \frac{1}{\ell}\right)^{2\ell} \ge \frac{1}{e^2}.$$

On the other hand, let $\mathcal{E}_2$ be the event that $T_2 \cap T' \neq \varnothing$. We see that

$$\Pr(\mathcal{E}_2) \ge 1 - \left(1 - \frac{1}{\ell}\right)^{k_2} \ge \left(1 - \frac{1}{\ell}\right)^{\ell} \ge 1 - \frac{1}{e}.$$

Since T_1 and T_2 are disjoint, $\mathcal{E}_1$ and $\mathcal{E}_2$ are independent, we have

$$\Pr(\mathcal{E}_1 \cap \mathcal{E}_2) \ge \left(1 - \frac{1}{e}\right) \cdot \frac{1}{e^2},$$

which is a constant. ∎

Theorem 1.3.7. There is a randomized algorithm that given $G = (V, E)$ and terminal set $T \subseteq V$ with $|T| = k$, outputs the Steiner min-cut with high probability using in $O(\log^2 k \log n)$ max-flow computations.[a]

[a] Again, potentially on graphs with $|V|$ vertices and $|E|$ edges each from Theorem 1.3.6.

Proof. From Lemma 1.3.7, Algorithm 1.11 successes with a constant probability. We further boost the overall success probability by rerunning Algorithm 1.11 $\Theta(\log n)$ times. With Theorem 1.3.6, this requires $O(\log^2 k \log n)$ max-flow computations. ∎

Remark (Deterministic algorithm). Li and Panigraphy [LP20] also developed deterministic min-cut and Steiner min-cut algorithms using additional ideas based on expander decomposition.

Lecture 5: Metric Embedding and Multi-Cut Problem

1.3.4 Max-Flow Min-Cut Theorem

10 Sep. 11:00

Finally, we conclude this section by proving the well-known max-flow min-cut theorem. Given a directed graph $G = (V, E)$ with edge capacity $c\colon E \to \mathbb{R}_+$, consider the following linear program relaxation of the

s-t min-cut where s, t are two distinct vertices:

$$\begin{aligned} \min \sum_{e\in E} c(e)x_e & & \max \sum_{P\in\mathcal{P}_{s,t}} y_P & \\ \sum_{e\in P} x_e \geq 1 \quad P\in\mathcal{P}_{s,t}; & & \sum_{P\ni e} y_P \leq c(e) \qquad \forall e\in E; & \\ \text{(P)} \quad x_e \geq 0 \qquad \forall e\in E; & & \text{(D)} \quad y_P \geq 0 \qquad \forall P\in\mathcal{P}_{s,t}, & \end{aligned} \tag{1.1}$$

where $\mathcal{P}_{s,t}$ is the set of all s-t paths. The integer version is with constraints $x_e \in \{0,1\}$.

Remark. An s-t cut is often also defined as $\delta^+(S)$ for some $S \subseteq V$ where $s \in S$ and $t \in V \setminus S$.

Proof. Suppose E' is an s-t cut and S is the set of nodes reachable from s in $G - E'$. Then, $\delta(S) \subseteq E'$ and $\delta(S)$ is an s-t cut. Hence, it suffices to focus on such limited type of cuts.[a] ⊛

[a] In some more general settings, it is useful to keep these notions separate.

It is well-known that s-t min-cut can be computed efficiently via s-t max-flow, establishing the max-flow min-cut theorem. This fundamental theorem in combinatorial optimization has many applications, and is typically established via the augmenting path algorithm. Here, we give another proof.

Theorem 1.3.8 (Max-flow min-cut). Let $G = (V, E)$ be a directed graph with rational edge capacities $c\colon E \to \mathbb{Q}_+$ and let $s, t \in V$ be two distinct vertices. The s-t max-flow value in G is equal to the s-t min-cut value, and both can be computed in strongly polynomial time. Furthermore, if c is integer valued, then there exists an integer-valued max-flow as well.

Proof. To start, we observe that following.

Intuition. The primal linear program assigns lengths to edges such that the s-t shortest path according to which is at least 1. This is a fractional relaxation of the cut.

We claim that it's possible to round the fractional solution of the primal to the exact s-t min-cut without any loss. Consider the following rounding algorithms for the primal linear program.

Algorithm 1.12: θ-Rounding Algorithm

Data: A directed graph $G = (V, E)$ with edge capacity $c\colon E \to \mathbb{R}_+$, $s, t \in V$
Result: A s-t min-cut F

```
{x_e}_{e∈E} ← LP-Solve(Min-Cut-LP(G, c, s, t))        // Solve the primal
θ ← Uniform((0, 1))
for v ∈ V do
    d_x(s, v) ← Shortest-Path-Dist(s, v, G, x)
return F = δ⁺(B_x(s, θ))        // B_x(s, θ) = {v ∈ V | d_x(s, v) ≤ θ}
```

Firstly, Algorithm 1.12 will output a valid s-t cut since $d_x(s,t) \geq 1$ by feasibility of the linear program solution x and hence $t \notin B_x(s,\theta)$ for any $\theta < 1$.

Claim. For any $e \in E$, $\Pr(e \text{ is cut by Algorithm 1.12}) \leq x_e$.

Proof. The edge $e = (u, v)$ is cut if and only if $d_x(s,u) \leq \theta < d_x(s,v)$. Hence, the edge is not cut if $d_x(s,v) \leq d_x(s,u)$. If $d_x(s,v) > d_x(s,u)$, we have $d_x(s,v) - d_x(s,u) \leq x_{(u,v)}$. Since θ is chosen uniformly at random from $(0,1)$, the probability that θ lies in the interval $[d_x(s,u), d_x(s,v)]$ is at most $x_{(u,v)}$. ⊛

With this claim, from linearity of expectation, we see that $\mathbb{E}[c(\delta^+(B_x(s,\theta)))] \leq \sum_{e\in E} c(e)x_e$. As $B_x(s,\theta)$ will always be a valid s-t cut, this implies that there is an integral cut whose cost is at most that of the linear program relaxation, implying that the linear program relaxation yields an optimum solution.

Finally, observe that the dual is the *path version* of the s-t max-flow. Hence, from strong duality, the optimal value of s-t min-cut is the same as the s-t max-flow. Moreover, we note that the primal

is strongly polynomial-time solvable if we have a separation oracle. In this case, given $\{x_e\}_{e\in E}$, we need to answer either this is a feasible solution, or outputs some path p such that $\sum_{e\in P} x_e < 1$, which is exactly the shortest s-t path algorithm and can be solved efficiently. ■

Intuition (Line embedding). The rounding can be thought as putting every vertex on a line from s to t, sorting by their distances given by x_e's. Then, observe that on that line, any two vertices $u, v \in V$ has distance at most $x_{(u,v)}$, and picking θ corresponds to picking a threshold on the line.

The above intuition not only helps the analysis, but also gives a way to derandomize Algorithm 1.12. Basically, we can try all possible θ's, and if we adapt this line embedding viewpoint, the only interesting θ's are given by the n values $d_x(s, v)$.

Chapter 2

Metric Methods

In this chapter, we will see a series of techniques that are based on the structure of the metric spaces underlying the graphs.

2.1 Multi-cut via Metric Decomposition

Recall that s-t min-cut problem we just saw. Now, consider a more general problem called multi-min-cut.

Problem 2.1.1 (Multi-min-cut). Given a graph $G = (V, E)$ with edge capacity $c\colon E \to \mathbb{R}_+$ and k pairs of vertices $\{(s_i, t_i)\}_{i=1}^k$, the *multi-min-cut problem* aims to find a minimum capacity cut that separates all pairs.

Multi-min-cut is NP-hard even on trees. In general, it is a NP-complete problem. Hence, we ask for an approximation algorithm instead. An $O(k)$-approximation algorithm is trivial by simply outputting the union of all s_i-t_i min-cuts. The goal is an $O(\log k)$-approximation algorithm, which also proves the multi-commodity flow-cut gap.

Note. It turns out that $O(\log k)$ is tight in general graphs. For planar graphs, one can get an $O(1)$-approximation and flow-cut gap. These results are only for undirected graphs since the situation is more complicated in directed graphs, and we will discuss that later.

Again, we can write the following linear program relaxation for the multi-min-cut problem:

$$\begin{aligned}
&\min \sum_{e\in E} c(e)x_e && && \max \sum_{i=1}^{k} \sum_{P\in\mathcal{P}_{s_i,t_i}} y_P && \\
&\sum_{e\in P} x_e \geq 1 && P \in \bigcup_{i=1}^{k} \mathcal{P}_{s_i,t_i}; && \sum_{i=1}^{k} \sum_{P\ni e} y_P \leq c(e) && \forall e \in E; \qquad (2.1)\\
&\text{(P)}\quad x_e \geq 0 && \forall e \in E; && \text{(D)}\quad y_P \geq 0 && \forall P \in \bigcup_{i=1}^{k} \mathcal{P}_{s_i,t_i}.
\end{aligned}$$

The primal assigns distance labels x_e to edges so that, on each path P between s_i and t_i, the distance labels of these edges on P sum up to at least one, just like the s-t min-cut.

Remark. The primal (with exponentially many constraints) is efficiently solvable.

Proof. With the ellipsoid method, we just need a separation oracle. Consider setting the length of each edge to x_e and for each pair (s_i, t_i), compute the length of the shortest path between s_i and t_i and check whether it is at least 1. This only takes k many times compared to the previous separation oracle for the s-t min-cut linear program relaxation, hence it's still polynomial time. ⊛

On the other hand, the dual variable can be interpreted as the amount of flow between s_i and t_i that is routed along the path P. This is called the *maximum throughput multi-commodity flow* problem,

where we don't care about individual demands, but only the overall flow. The dual tries to assign an amount of flow y_P to each path P so that the total flow on each edge is at most the capacity of the edge.

Note. The flow conservation constraints are automatically satisfied and the endpoints of the path P determine which kind of commodity is routed along the path.

2.1.1 Approximation via Randomized Decomposition

The first algorithm [GVY93] (see [Vaz01; WS11]) that achieved an $O(\log k)$-approximation for multi-min-cut is based on the region growing technique [LR99]. Here, we present the randomized rounding algorithm due to its future application for metric embedding [CKR05], which in particular also shows the integrality gap of the primal linear program to be $O(\log k)$. The goal is to find a procedure to cut (i.e., decompose) the graph into components that satisfies the requirement of being a multi-cut, i.e.,

- each component has diameter at most 1 when the edge capacity is induced by the primal solution;
- the probability that edge e is cut is at most αx_e,

then we will get an α-approximation algorithm. To do this, we consider the following algorithm.

Algorithm 2.1: Random Partition [CKR05]

Data: A connected graph $G = (V, E)$ with edge capacity $c\colon E \to \mathbb{R}_+$, $\{(s_i, t_i) \mid s_i, t_i \in V\}_{i=1}^k$
Result: A multi-min-cut F

$\{x_e\}_{e\in E} \leftarrow$ `LP-Solve(Multi-Min-Cut-LP(`G, c, $\{(s_i,t_i)\}_{i=1}^k$`))` `// Solve the primal`
$\theta \leftarrow$ `Uniform(`$[0, 1/2)$`)`
$\sigma \leftarrow$ `Uniform(`$S([k])$`)` `// Random permutation from permutation group` $S([k])$
for $i = 1, \dots, k$ **do**
  $V_{\sigma(i)} \leftarrow B_x(s_{\sigma(i)}, \theta) \setminus \bigcup_{j<i} V_{\sigma(j)}$
$F \leftarrow \bigcup_{i=1}^k \delta(V_i)$
return F

Intuition. It essentially reduces to *low-diameter decomposition* in a metric space.

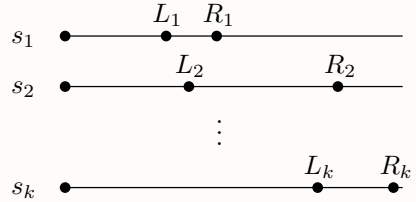
Lemma 2.1.1. Algorithm 2.1 outputs a feasible multi-cut for the given instance.

Proof. Suppose not, then there exists a pair (s_i, t_i) are still connected in $G - F$. This can only happen if s_i is "grabbed" by some terminals s_j which is proceeded before s_i, i.e., there exists some $V_j \subseteq B_x(s_j, \theta)$ that contains both s_i and t_i. However, if s_j grabs both s_i and t_i, it means the distance between s_i and t_i is at most $2\theta < 1$, a contradiction to the feasibility of $\{x_e\}_{e\in E}$. ■

Lemma 2.1.2. For any $e \in E$, $\Pr(e = uv \text{ is cut by Algorithm 2.1}) \le 2H_k x_e \le O(\log k) x_e$.[a]

[a] H_k is the k^{th} harmonic number.

Proof. Let $L_i := \min(d_x(s_i, u), d_x(d_i, v))$ and $R_i := \max(d_x(s_i, u), d_x(s_i, v))$. Without loss of generality, we can renumber s_i's such that $L_1 \le L_2 \le \cdots \le L_k$.

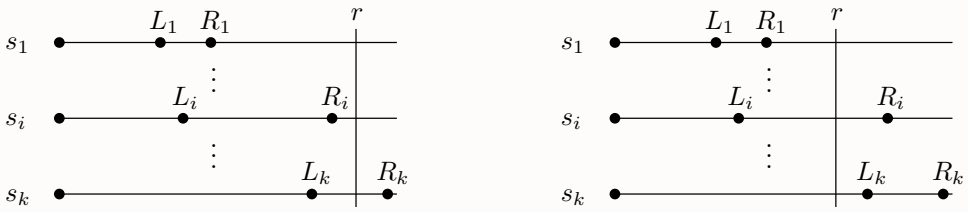


Let A_i be the event that s_i cut $e = uv$ *first*, i.e., A_i is the event that $|V_i \cap \{u, v\}| = 1$ and $|V_j \cap \{u, v\}| = 0$ for all j such that $\sigma(j) < \sigma(i)$, where $|V_i \cap \{u, v\}| = 1$ simply says that s_i *cuts* the edge e. If A_i happens, then for all j that come before i in σ, neither u nor v can be in V_j since:

- if only one of u and v is in V_j, then s_j cuts e;
- if both u and v are in V_j, s_i can't cut e as the cut set only grabs the leftover vertices (line 5).

Let A be the event that e is cut, which is the union of the disjoint events A_i's, hence $\Pr(A) = \sum_{i=1}^{k} \Pr(A_i)$. Now, for any fixed $r \in [0, 1/2)$, we see that

- $r \notin [L_i, R_i)$: This is easy to understand as if $r \notin [L_i, R_i)$, s_i is impossible to cut e;
- $r \in [L_i, R_i)$: Consider some $j < i$ and suppose j comes before i in the permutation (i.e., $\sigma(j) < \sigma(i)$). Since $j < i$, $L_j \le L_i \le r$. Hence, at least one of u and v is inside the ball of radius r centered at s_j. Consequently, s_i can't be the first to cut e, resulting in the fact that s_i is the first to cut the edge e if $\sigma(i) < \sigma(j)$ for all $j < i$.



Since σ is a random permutation, i appears before j for all $j < i$ with probability $1/i$. Hence,

$$\begin{cases} \Pr(A_i \mid \theta = r) = 0, & \text{if } \theta \notin [L_i, R_i); \\ \Pr(A_i \mid \theta = r) \le 1/i, & \text{if } \theta \in [L_i, R_i). \end{cases}$$

As θ is independent of σ, we have

$$\Pr(A_i) \le \frac{1}{i} \cdot \mathbb{P}(\theta \in [L_i, R_i)) = \frac{2}{i}(R_i - L_i) \le \frac{2x_e}{i}$$

from the triangle inequality $R_i \le L_i + x_e$. This finally leads to

$$\Pr(A) = \sum_{i=1}^{k} \Pr(A_i) \le \sum_{i=1}^{k} \frac{2x_e}{i} \le 2H_k x_e,$$

and we conclude the theorem by noting that $H_k = O(\log k)$. ■

Theorem 2.1.1. Algorithm 2.1 is an $O(\log k)$-approximation (in expectation) algorithm for the multi-min-cut problem. Furthermore, the integrality gap of the multi-min-cut linear program is $O(\log k)$.

Proof. Let F be the set of edges outputted by Algorithm 2.1. For each edge e, let $\xi_e = \mathbb{1}_{e \in F}$ in an indicator random variable. Hence, we have $\mathbb{E}[\xi_e] = \mathbb{P}(\xi_e = 1) \le 2H_k x_e$ from Lemma 2.1.2. This leads to

$$\mathbb{E}[c(F)] := \mathbb{E}\left[\sum_{e \in F} c(e)\right] = \mathbb{E}\left[\sum_{e \in E} c(e)\xi_e\right] = \sum_{e \in E} c(e) \Pr(\xi_e) \le 2H_k \sum_{e \in E} c(e)x_e = 2H_k \, \mathsf{OPT}_{\mathrm{LP}}$$

where $\mathsf{OPT}_{\mathrm{LP}}$ is the optimal value of the linear program. Since $\mathsf{OPT}_{\mathrm{LP}} \le \mathsf{OPT}$ where OPT is the optimum value of the multi-min-cut problem, we have

$$\mathbb{E}[c(F)] \le 2H_k \, \mathsf{OPT}_{\mathrm{LP}} \le 2H_k \, \mathsf{OPT} = O(\log k) \, \mathsf{OPT}\,.$$

This also implies that there exists a set of edges F such that the total capacity of edges in F is at most $2H_k \, \mathsf{OPT}_{\mathrm{LP}}$, i.e., $\mathsf{OPT}_{\mathrm{LP}} \le \mathsf{OPT} \le 2H_k \, \mathsf{OPT}_{\mathrm{LP}}$, which proves the integrality gap result. ■

The expected cost analysis can be used to obtain a randomized algorithm via repetition that outputs an $O(\log k)$-approximation with high probability. The algorithm can also be derandomized, but it's not straight forward.

Remark (Flow-cut gap). Recall that when $k = 1$, we have the max-flow min-cut theorem. The integrality gap of the standard linear program for multi-min-cut is the same as the relative gap between flow and cut when k is arbitrary. The upper bound on the integrality gap gives an upper bound on the gap.

Lecture 6: Low-Diameter Decomposition and Tree Embeddings

2.1.2 Low-Diameter Decomposition

12 Sep. 11:00

Given a graph $G = (V, E)$ and edge length $\ell \colon E \to \mathbb{R}_+$, which define a metric space (V, d) where $d(u, v)$ is the shortest path distances between u and v in G as before. As we have seen, a useful notion is to decompose or partition the graph into subgraphs (or clusters) of small diameter. More precisely, given a graph $G = (V, E)$, we would like to partition V into clusters with vertex sets $\{V_i\}_{i=1}^h$ such that each V_i has diameter at most some given parameter δ.

Example (Singleton). It's trivial to consider the singleton partition, where $V_i = \{v\}$ for all $v \in V$.

However, the goal in partitioning is to ensure that two vertices $u, v \in V$ that are close to each other, say $d(u, v) < \delta$, should ideally not be split apart into different clusters. However, as the graph (or metric space) is connected, it's impossible to do this deterministically.

Example (Line graph). Considered a line graph L_n. The most natural randomized algorithm is to shift the line by $\theta \in [0, \delta)$, and then we separate the line graph by δ-length clusters. In this way, the probability that any pair $u, v \in V$ is cut is at most $d(u, v)/\delta$.

This is the best we can hope for, i.e., u, v are separated only with probability proportional to $d(u, v)/\delta$.

Definition 2.1.1 (Low-diameter decomposition). Let $G = (V, E)$ be a graph with edge lengths $\ell \colon E \to \mathbb{R}_+$, which induces a distance metric $d \colon V \times V \to \mathbb{R}_+$. Let Δ be the diameter of the metric space (V, d). For a given $\delta \in [0, \Delta]$, a *low-diameter decomposition* with cutting probability parameter α is a probability distribution $\mathcal{D}$ over the set $\mathcal{P}$ of all partitions of V such that

- for any partition $P = (V_1, \ldots, V_h) \in \mathcal{P}$ in $\operatorname{supp}(\mathcal{D})$, $\operatorname{diam}((V_i, d)) \le \delta$;
- for all $u, v \in V$, $\Pr(u, v \text{ are separated}) \le \alpha d(u, v)/\delta$.

Given a partition $P \in \mathcal{P}$, we write E_P be the set of edges (pairs) that are separated by P. Hence, in Definition 2.1.1, the second point is equivalent to $\Pr((u, v) \in E_P) \le \alpha d(u, v)/\delta$.

Definition 2.1.2 (Low-diameter decomposition scheme). A *low-diameter decomposition scheme* with parameter α is a family of algorithms that given any $\delta \in [0, \Delta)$, generates a low-diameter decomposition with cutting (separation) probability parameter at most α.

Notation (Strong v.s. weak diameter guarantee). A low-diameter decomposition is said to have the *strong diameter* guarantee if the diameter of each cluster V_i in the induced graphs $G[V_i]$ is at most δ. Note that Definition 2.1.2 does not require that because it is based on the metric closure (V, d) of the given graph. The standard definition is called the *weak diameter* guarantee. Some applications require the strong diameter guarantee. We will, by default, work with weak diameter guarantee and mention strong diameter guarantee when needed.

Note (Padded decomposition). Definition 2.1.1 is often strengthened to require more. Given a point u, let $B_d(u, r) = \{v \in V \mid d(u, v) \le r\}$. In *padded decomposition*, we require that for each u, and for each $r \le \delta$, the probability of $B_d(u, r)$ being contained in the same part is at least $e^{-\beta r}$.

Remark (Sparse cover). A *sparse cover* consists of several clusters $\{V_i\}_{i=1}^{h}$. Each cluster should have weak/strong diameter at most δ. For each u, v with distance at most δ, there must be some cluster V_i that contains both u and v, and no vertex u must be in more than some number s of clusters. The techniques underlying sparse covers and low-diameter decomposition are related though we will mostly work only with the latter.

The main question here for the low-diameter decomposition is the smallest α that one can obtain. It turns out that for general metric spaces, $\alpha = O(\log n)$ is a tight bound for both strong and weak diameter guarantee [Bar96]. For planar graph metrics, $\alpha = O(1)$ is achievable, where weak diameter guarantee was shown in [KPR93], and the strong diameter guarantee was more difficult and was shown later. See [Fil24] for some recent work and pointers to literature on these topics.

Now, we present the algorithm for weak diameter guarantee. Let (V, d) be a metric space with $|V| = n$. Borrowing ideas from Algorithm 2.1, we see that implicitly this is a metric partitioning scheme. We simply modify the algorithm to ensure that the weak diameter of each cluster is at most a given parameter δ.

Algorithm 2.2: Random Partition [CKR05]

Data: A metric space (V, d), δ
Result: E_P for the partition P

$\theta \leftarrow$ `Uniform`$([0, \delta/2))$
$\sigma \leftarrow$ `Uniform`$(S(V))$ `// Random permutation from permutation group` $S(V)$
for $i = 1, \dots, n$ **do**
  $V_{\sigma(i)} \leftarrow B_d(v_{\sigma(i)}, \theta) \setminus \bigcup_{j<i} V_{\sigma(j)}$
return $\bigcup_{i=1}^{n} \delta(V_i)$

From the same analysis as in Lemma 2.1.2, claim the following.

Claim. Algorithm 2.2 correctly outputs a partition of V into clusters, each of which has weak diameter at most δ.

Furthermore, the probability guarantee can also be stated.

Theorem 2.1.2. The probability that u and v are in different clusters outputted by Algorithm 2.2 is at most $2H_n d(u, v)/\delta$, i.e., $\alpha = H_k = O(\log n)$.

Proof. We see that $\Pr(A_j) \le 2d(u, v)/\delta \cdot 1/j$, hence $\Pr(A) \le 2H_n d(u, v)/\delta$. ■

Finally, we consider a not so apparent modification of Algorithm 2.2: we sample θ from $[\delta/4, \delta/2)$ instead of $[0, \delta/2)$.

Algorithm 2.3: Refined Random Partition [CKR05]

Data: A metric space (V, d), δ
Result: E_P for the partition P

$\theta \leftarrow$ `Uniform`$([\delta/4, \delta/2))$
$\sigma \leftarrow$ `Uniform`$(S(V))$ `// Random permutation from permutation group` $S(V)$
for $i = 1, \dots, n$ **do**
  $V_{\sigma(i)} \leftarrow B_d(v_{\sigma(i)}, \theta) \setminus \bigcup_{j<i} V_{\sigma(j)}$
return $\bigcup_{i=1}^{n} \delta(V_i)$

Intuition. Intuitively, this will preserve closer points.

It's clear that the guarantee about the diameter remains the same.

Claim. Algorithm 2.3 correctly outputs a partition of V into clusters, each of which has weak diameter at most δ.

The main difference is in the probability guarantee which is refinement of the previous bound.

Theorem 2.1.3. The probability that u and v are in different clusters outputted by Algorithm 2.3 is at most $\frac{4d(u,v)}{\delta}\log\frac{|B(u,\delta/2)|}{|B(u,\delta/8)|}$, i.e., $\alpha = \alpha(\delta) = 4\log\frac{|B(u,\delta/2)|}{|B(u,\delta/8)|}$.

Proof. We sketch the proof based on the proof of Lemma 2.1.2. Assuming the exact same notation, and we fix $u, v \in V$ and think of V as $v_1, \ldots, v_n$. If $d(u,v) \geq \delta/8$, then the edge is going to get cut with constant probability, and the bound is not giving anything interesting, so we are primarily interested in the case when $d(u,v) < \delta/8$.

We consider the event A_i which is that v_i is the first vertex to separate the pair u, v. We can argue as before that $\Pr(A_i) \leq 1/i \cdot \Pr(\theta \in [L_i, R_i])$, and this is at most $4d(u,v)/\delta i$ since we're choosing the radius from $[\delta/4, \delta/2)$. The new twist is that since we choose $\theta \in [\delta/4, \delta/2)$ and $d(u,v) < \delta/8$, no vertex $v_j \in B(u, \delta/8)$ can separate u, v because $L_j \leq d(v_j, u) < \delta/8$ and $R_j \leq L_j + d(u, v_j) \leq \delta/8 + \delta/8 = \delta/4$. Any such vertex will capture both u, v if they are not already separated. Similarly, any vertex $v_j \notin B(u,\delta)$ can cut the pair because $L_j \geq \delta - d(u,v) \geq \delta - \delta/8 \geq \delta/2$. Therefore, if A is the event of u, v being cut, then

$$\Pr(A) \leq \sum_{j \in B(u,\delta)\setminus B(u,\delta/8)} \Pr(A_j) \leq \frac{4d(u,v)}{\delta} \sum_{|B(u,\delta/8)| < j \leq |B(u,\delta)|} \frac{1}{j} \leq \frac{4d(u,v)}{\delta}\log\frac{|B(u,\delta)|}{|B(u,\delta/8)|},$$

proving the result. ■

2.2 Dominating Tree Metrics Embedding

Using tree representations of graphs is a powerful tool in algorithm design. Here, we are interested in representing the distances in an undirected graph via distances in a spanning tree. Let $G = (V, E)$ be a graph with edge length $\ell \colon E \to \mathbb{R}_+$, which induces a metric space (V, d) via shortest path distances. The main question is the following:

Problem 2.2.1 (Tree embedding). Given a graph $G = (V, E)$ with edge length $\ell \colon E \to \mathbb{R}_+$, the *tree embedding* problem aims to find a spanning tree $T = (V, E_T)$ of G such that for any $u, v \in V$, $d_T(u,v) \leq \alpha d_G(u,v)$ where α is called the *distortion* or *stretch*.[a]

[a] Clearly, $d_T(u,v) \geq d_G(u,v)$.

Example (Cycle). Consider a cycle C_n. We see that for a fixed edge (u,v), there exists one spanning tree T such that $d_T(u,v) = n - 1$.

Motivated by applications of spanning tree based metric approximations, we observe that if we are allowed to pick a probability distribution over spanning trees, then the expected distance for any pair of vertices can be much better than the above worst-case example.

Example (Cycle). Again, consider a cycle C_n. If we allow randomization (picking trees randomly),

$$\mathbb{E}[d_T(u,v)] = \frac{n-1}{n} \cdot 1 + \frac{1}{n} \cdot (n-1) \leq 2.$$

It's showed that [Alo+95] for any weighted graph $G = (V, E)$, there is a distribution $\mathcal{D}$ over spanning trees of G such that for any $u, v \in V$,

$$\mathbb{E}_{T \sim \mathcal{D}}[d_T(u,v)] \leq \exp\left(\sqrt{\log n \log\log n}\right) \cdot d_G(u,v) < n^{o(1)} d_G(u,v).$$

Intuition. This is a probabilistic approximation of a graph metric by spanning tree metrics.

This can also be viewed as a metric *embedding* result. In keeping with metric embedding terminology, we're interested in the worst-case guarantee of how much the expected distance for any pair increases, i.e., minimizing α. In the above example $\alpha \leq \exp\left(\sqrt{\log n \log\log n}\right)$.

Note (Lower bound). A lower bound of $\alpha = \Omega(\log n)$ is required for probabilistic tree approximation, and it's conjectured that this is tight [Alo+95].

2.2.1 Dominating Tree Metric

It turned out that this is quite difficult to obtain even a poly-logarithmic bound [Elk+05], and currently the best known bound is $O(\log n \log\log n)$ [ABN08]. To make the problem easier, [Bar96] proposed to forget about the graph topology and just focus on (V, d), i.e., consider the *metric embeddings* instead of spanning tree embeddings. More generally, we work with the metric completion (V, d) and view it as a complete graph on V, where any spanning tree of the complete graph is now allowed. Moreover, we allow additional vertices. This is formalized as follows.

Definition 2.2.1 (Dominating tree metric). A tree $T = (V_T, E_T)$ with edge length $\ell_T \colon E_T \to \mathbb{R}_+$ is a *dominating tree metric* for a finite metric space (V, d) if $V \subseteq V_T$ and for all $u, v \in V$, $d_T(u, v) \geq d_G(u, v)$.

Then, we're interested in approximating metrics probabilistically by dominating tree metrics:

Definition 2.2.2 (Probabilistic approximation). A *probabilistic approximation* of a metric space (V, d) by dominating tree metrics is a probability distribution $\mathcal{D}$ over a collection of trees $\{T_i\}_{i=1}^h$ if each T_i is a dominating tree metric for (V, d).

Furthermore, we say that the probabilistic approximation $\mathcal{D}$ has *stretch* α if for all $u, v \in V$,

$$\mathbb{E}_{T\sim\mathcal{D}}[d_T(u, v)] \leq \alpha d(u, v).$$

2.2.2 Tree Embedding with Low-Diameter Decomposition

One can use low-diameter decomposition to efficiently sample from a distribution that has stretch $\alpha = O(\log^2 n)$ [Bar96], and this was subsequently improved to $O(\log n \log\log n)$ [Bar98], and finally improved to the optimal $O(\log n)$ [FRT03] using Algorithm 2.3.

Intuition. Recursively decompose (V, d) using low-diameter decomposition.

Specifically, let (V, d) be a metric space with diameter Δ. We use low-diameter decomposition with parameter $\delta = \Delta/2$ to randomly partition V into clusters $\{V_i\}_{i=1}^h$ of diameter at most $\Delta/2$, then we recursively find a tree for each of the V_i, with root r_i. We create a new dummy root r and connect each r_i to r with an edge of length Δ.

Algorithm 2.4: Tree Embedding

Data: A metric space (V, d), diameter D
Result: A rooted tree (T, r)

if $|V| = 1$ **then**
  $T \leftarrow (V, \varnothing)$
  return (T, v) // $V = \{v\}$

Create a tree T with root r
$\{V_i\}_{i=1}^h \leftarrow$ `Low-Diameter-decomposition`$((V, d),\ D/2)$
for $j = 1, \ldots, h$ **do**
  $(T_j, r_j) \leftarrow$ `Tree-Embedding`$((V_j, d),\ D/2)$
  Connect T_j to T by adding edge (r, r_j) of length D
return (T, r)

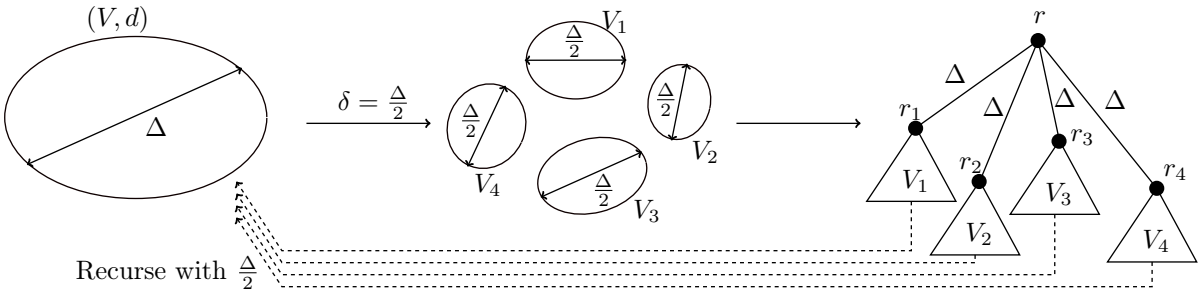


Figure 2.1: Illustration of Algorithm 2.4.

Notation. We say $\delta = \Delta/2^i$ at level i to make the analysis cleaner.

We will now assume that the minimum distance is at least 1 by scaling, and let Δ be the diameter of the metric space with this assumption.

Remark. If the minimum distance is at least 1, Algorithm 2.4 yields a stretch of $O(\log n \log \Delta)$.

Theorem 2.2.1. Let Δ be the diameter of (V, d). Algorithm 2.4 outputs a random dominating tree metric $T = (V_T, E_T)$ with length ℓ_T such that for each $u, v \in V$, $\mathbb{E}_{T\sim\mathcal{D}}[d_T(u, v)] \leq O(\alpha \log \Delta) d(u, v)$ where α is the cutting probability of the low-diameter decomposition algorithm used.

Proof. We prove this by induction. It's clear that the base case is trivial. If we start with $D = \Delta$, then at depth i of the recursion, the parameter is $\Delta/2^{i-1}$, and it is the upper-bound on the diameter of the metric space in that recursive call. We note the following claims.

Claim. The length of the root to leaf path of a tree created at level i of the recursion is at most $\sum_{j\geq i} \Delta/2^{j-1} \leq 2\Delta/2^{i-1}$.

Suppose u and v are first separated at level i of the recursion. Then, $d_T(u, v) \leq 4\Delta/2^{i-1}$ from the above claim. We see that if u and v are separated in the first level of the recursion due to the low-diameter decomposition algorithm, its probability is at most $\alpha d(u, v)/(\Delta/2) \leq 2\alpha d(u, v)/\Delta$, in which case their distance in the tree is at most 4Δ. Otherwise, they are in the same cluster, and we can apply induction. Note that u and v are definitely separated by level t where t is the smallest integer such that $\Delta/2^{t+1} < d(u, v)$. Hence, the depth of the recursion is at most $1 + \lceil \log \Delta \rceil \leq 2 \log \Delta$. It's easy to unroll the induction and use the preceding claim to obtain

$$\mathbb{E}_{T\sim\mathcal{D}}[d_T(u, v)] \leq \sum_{i=0}^{t+1} 2\alpha \frac{d(u, v)}{(\Delta/2^{i-1})} \cdot \left(4\frac{\Delta}{2^{i-1}}\right) \leq O(\alpha \log \Delta) d(u, v) \tag{2.2}$$

since the depth of the recursion is $O(\log \Delta)$. ■

If we use Algorithm 2.2 as the low-diameter decomposition algorithm, then we have $\alpha = O(\log n)$. From Theorem 2.2.1, we see that the tree may require depth $\log \Delta$ to provide a good approximation, and in general $\log \Delta$ can be as large as n, so we get an $O(n \log n)$ approximation.

Example. Consider the metric induced by a path with n edges and edge lengths are 2^i for all $i = 1, \ldots, n$. In such cases, the dependence of the stretch on $\log \Delta$ is undesirable.

Remark. One can alter Algorithm 2.4 to make the stretch bound $O(\log^2 n)$.

Proof. In applying the low-diameter decomposition algorithm with parameter δ, we ensure that any pair u, v such that $d(u, v) \leq \delta/n^2$ is not cut during the procedure. We can do this by contracting all

such pairs without changing the diameter of the resulting metric space too much. This will ensure that in the tree construction process, a pair u, v participates in only $O(\log n)$ levels and hence the expected stretch can be bounded by $O(\alpha \log n)$. ⊛

Remark (Hierarchically well-separated tree)**.** The trees constructed by Algorithm 2.4 have an additional strong property: the edge lengths at each level are the same and the length from the root to the leaf go down by a factor of 2 at each level. A tree metric with such a property is called *hierarchically well-separated* tree metric and this additional property can be exploited in algorithms and comes for free in the construction.

Finally, we note that if we choose Algorithm 2.3 as the low-diameter decomposition algorithm specifically, the expected stretch is actually $O(\log n)$, which is optimal [FRT03].

As previously seen (Theorem 2.1.3)**.** The probability that u and v are in different clusters outputted by Algorithm 2.3 is at most $\frac{4d(u,v)}{\delta} \log \frac{|B(u,\delta/2)|}{|B(u,\delta/8)|}$.

Thus, the guarantee $\alpha(\delta)$ from the low-diameter decomposition algorithm is no longer uniform but depends on the diameter. Plugging this to Equation 2.2, we have

$$\begin{aligned}\mathbb{E}_{T\sim\mathcal{D}}[d_T(u,v)] &\le \sum_{i=0}^{t+1} 2\log\frac{|B(u,\Delta/2^i)|}{|B(u,\Delta/2^{i+3})|}\frac{d(u,v)}{\Delta/2^{i-1}}\cdot\left(4\frac{\Delta}{2^{i-1}}\right)\\ &\le 8d(u,v)\left(\log|B(u,\Delta/2)| + \log|B(u,\Delta/4)| + \log|B(u,\Delta/8)|\right) \le O(\log n)d(u,v).\end{aligned}$$

Remark (Lower bound)**.** $O(\log n)$ bound for low-diameter decomposition algorithm and tree embeddings are near optimal (modulo precise constant factors).

Proof. We can use the existence of low-girth graphs which are closely tied to expanders in a direct fashion. Another way is via indirection. Previously, in the lecture note, we saw that the integrality gap of the linear program relaxation for multi-min-cut is $\Omega(\log k)$ (the upper-bound is proved in Theorem 2.1.1 via a low-diameter decomposition algorithm). In fact, if α is the factor for the low-diameter decomposition algorithm, then we get an $O(\alpha)$-approximation for multi-min-cut via the linear program. Thus, one see that $\alpha = \Omega(\log n)$ for general metrics (in the integrality gap example for multi-min-cut $k = \Omega(n^2)$). One can use similar approaches to prove for tree embedding. ⊛

Note (Efficient algorithms)**.** While we focus on the quality of low-diameter decomposition algorithms and tree embedding but not so much on the running times, it is easy to see that the algorithms themselves can be implemented in polynomial time. The main computation is about the shortest paths. If one computes all pairs shortest paths (APSP), then the algorithms are pretty simple. However, APSP is slow, which takes $O(mn)$ times. It is possible to compute the low-diameter decomposition algorithm and metric tree embeddings in close to linear time on a weighted graph with m edges. This involves computing approximate shortest paths and several tricks, and sometimes we give up on the quality of the approximation by logarithmic factors.

Lecture 7: Linear Programming for Sparsest Cut

2.3 Sparsest Cut

17 Sep. 11:00

Consider building a network of n vertices. The best network might be the complete graph, which can do everything and is robust. However, the problem is that the degree is too high $(n-1)$.

Example. For degree equal to 2, the best we can hope for is a cycle.

The magic happens whenever the degree goes up to 3.

Intuition. If we can down-weight edges in a complete graph K_n by $3/(n-1)$, then any cut S has

$$\delta(S) = |S| \cdot |V \setminus S| \cdot \frac{3}{n-1} \approx c|S|$$

for $|S| \ll |V \setminus S|$ and some $c \geq 0$.

This notion can be formalized as the so-called expander [HLW06]. We postpone the formal introduction of expander, and first focus on a closely related problem, the sparsest cut problem, specifically, the non-uniform version. It turns out that solving this helps us answer various questions for expanders.

2.3.1 Uniform and Non-Uniform Sparsest Cut Problem

To introduce the problem, we first define the sparsity for a cut.

Definition 2.3.1 (Sparsity). Given a graph $G = (V, E)$ with edge capacity $c\colon E \to \mathbb{R}_+$ and a demand graph $H = (V, F)$ with demand capacity $D\colon F \to \mathbb{R}_+$, for any cut $S \subseteq V$, its *sparsity* is defined as

$$\frac{c(\delta(S))}{\sum_{i\colon |S \cap \{s_i, t_i\}|=1} D_i}.$$

That is, the sparsity of a cut is the ratio of the capacity of the cut and the total demand of the pairs separated by S. Now, we can introduce the non-uniform sparsest cut problem.

Problem 2.3.1 (Non-uniform sparsest cut). Given a supply graph $G = (V, E)$ with edge capacity $c\colon E \to \mathbb{R}_+$ and k pairs of vertices $\{(s_i, t_i)\}_{i=1}^k$ along with non-negative demand values $D_1, \ldots, D_k$.[a] The *non-uniform sparsest cut problem* aims to find a cut S with minimum sparsity.

[a] If G is undirected, then the demand pairs are unordered, i.e., we do not distinguish (s_i, t_i) from (t_i, s_i).

Here, demands forms a demand graph $H = (V, F)$ with edges (s_i, t_i) and demand capacity $D\colon F \to \mathbb{R}_+$ such that $D((s_i, t_i)) = D_i$. With this representation, the sparsity of cut S is simply $c(\delta_G(S))/D(\delta_H(S))$ where $\delta_G(S)$ (respectively, $\delta_H(S)$) represents the supply (respectively, demand) edges crossing S.

Intuition. We're trying to find the best "bang per buck" cut, i.e., how much capacity do we need to remove per amount of demand separated to satisfy the demand?

Remark. We can define a cut as removing a set of edges, leading to more than two components. In the case of sparsest cut in undirected graphs, it suffices to restrict attention to cuts of the form $\delta(S)$ for some $S \subseteq V$. This is not necessarily true for directed graphs or even in undirected graphs with node-weights or in hypergraphs.

Problem 2.3.2 ((Uniform) sparsest cut). The *sparsest cut* problem is the same as Problem 2.3.1 with all $D(u, v) = 1$ for each unordered pair of vertices (u, v).[a]

[a] That is, $\{(s_i, t_i)\}_{i=1}^k$ is the set of all unordered pairs of vertices.

We see that in the uniform sparsest cut problem, the sparsity of a cut S is given by $c(\delta_G(S))/|S||V \setminus S|$. In this case, the demand graph H is a complete graph with unit demand values on each edge.

Finally, to further motivate the problem, we see that the uniform sparsest cut helps us directly and indirectly solve the balanced separator problem, a central problem is graph algorithm:

Problem 2.3.3 (Balanced separator). The *balanced separator problem* aims to partition a graph $G = (V, E)$ into two pieces $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ such that $|V_1| \approx |V_2|$ while minimizing edges between V_1 and V_2.

2.3.2 Linear Program Relaxation and Maximum Concurrent Flow

It's not clear how to start solving the uniform sparsest cut problem as writing a linear program relaxation for which is not obvious, compared to multi-min-cut and other cut problems where we have explicit terminal pairs that we wish to separate. Hence, to write an integer program for which, we let $y_i \in \{0,1\}$ being the indicator variable of whether we want to separate (s_i, t_i). Moreover, let $x_e \in \{0,1\}$ for all $e \in E$ to be the cut indicator variables.

> **Intuition.** If we decide to separate (s_i, t_i), then for every path between s_i and t_i we should cut at least one edge on the path.

Hence, a natural integer program of non-uniform sparsest cut and its relaxation is given by:

$$\begin{aligned}
&\min \frac{\sum_{e\in E} c(e)x_e}{\sum_{i=1}^{k} D_i y_i} && && \min \sum_{e\in E} c(e)x_e \\
&\sum_{e\in P} x_e \geq y_i && \forall P \in \bigcup_{i=1}^{k} \mathcal{P}_{s_i,t_i}; \quad \to && \sum_{i=1}^{k} D_i y_i = 1 \\
&x_e \in \{0,1\} && \forall e \in E; && \sum_{e\in P} x_e \geq y_i && \forall P \in \bigcup_{i=1}^{k} \mathcal{P}_{s_i,t_i}; \\
&y_i \in \{0,1\} && \forall i = 1,\dots,k; && x_e \geq 0 && \forall e \in E; \\
& && && \text{(P)} \quad y_i \geq 0 && \forall i = 1,\dots,k,
\end{aligned} \tag{2.3}$$

where we use the standard trick, i.e., linearization to normalize the denominator to be 1, to make the ratio of the integer program into a linear program. With the dual variable z_P for each path such that it indicates the amount of "flow" sent on the path P, the dual of the above is

$$\begin{aligned}
&\max \lambda \\
&\sum_{P\in\mathcal{P}_{s_i,t_i}} z_P \geq \lambda D_i && \forall i = 1,\dots,k; \\
&\sum_{i=1}^{k} \sum_{\substack{P\in\mathcal{P}_{s_i,t_i} \\ P\ni e}} z_P \leq c(e) && \forall e \in E; \\
&z_P \geq 0 && \forall P \in \bigcup_{i=1}^{k} \mathcal{P}_{s_i,t_i}; \\
&\text{(D)} \quad \lambda \geq 0,
\end{aligned} \tag{2.4}$$

which is a multi-commodity flow. In particular, it solves the *maximum concurrent multi-commodity flow* problem for the given instance, i.e., it finds the largest value of λ such that there is a feasible multi-commodity flow for the given pairs in which the flow routed for pair (s_i, t_i) is at least λD_i.

> **Notation** (Concurrent flow). It is called *concurrent flow* since we need to route all demand pairs to the same factor which is in contrast to the dual of multi-min-cut, which corresponds to the maximum throughput multi-commodity flow.[a]
>
> [a] Recall that in this case, some pairs may have zero flow while others have a lot of flow.

We note that this dual can be solved efficiently via ellipsoid method since the separation oracle is just the shortest path problem. One can also write a compact linear program via distance variables as

$$\begin{aligned}
&\min \sum_{uv\in E} c(uv)d(uv) \\
&\sum_{i=1}^{k} D_i d(s_i,t_i) = 1 \\
&d \text{ is a metric on } V.
\end{aligned}$$

In this context, we can understand the *flow-cut gap* of non-uniform sparsest cut by the following equivalent way of thinking about the problem:

Intuition (Cut-condition). Given a multi-commodify flow instance on G. A necessary condition to route all the demand pairs is that if G satisfies the *cut-condition*, i.e., for every $S \subseteq V$, the capacity $c(\delta(S))$ is at least the demand separated by S. However, the converse it not necessarily true, i.e., the cut-condition is not sufficient.

The cut-condition is sufficient when $k = 1$ but is not true in general even for $k = 3$ in undirected graphs. The question is the maximum value of λ such that we can route λD_i for each pair i. The worst-case integrality gap of the proceeding linear program is precisely the flow-cut gap.

2.3.3 Rounding Linear Program via ℓ_1 Embeddings

It's known that there is an $O(\log n)$-approximation algorithm and the flow-cut gap for uniform sparsest cut, together with a lower bound of $\Omega(\log n)$ on the flow-cut gap for uniform sparsest cut via expanders [LR99]. This leads to an $O(\log^2 n)$-approximation for non-uniform sparsest cut, and it was an open problem to obtain a tight conjectured bound of $O(\log n)$. It turns out to be possible via the optimal rounding algorithm for the linear program relaxation. This goes via metric embedding theory [LLR95; AR98], hence we need some basics in metric embeddings to point out the connection and rounding.

Note. Even though the metric embedding machinery is powerful, it can seem like magic. The more basic ideas for uniform sparsest cut based on region growing is useful to know [WS11].

Remark (Rounding via multi-min-cut). There are also close connections between sparsest cut and multi-min-cut. In particular, suppose there is an $\alpha(k, n)$-approximation for non-uniform sparsest cut, then we have an $O(\alpha(k, n) \ln k)$-approximation for multi-min-cut. The converse is also true.[a]

[a] See the note for a reference.

Before we start, consider the following simple setting when G is a tree $T = (V, E)$.

Example (Tree). Given a tree $G = T = (V, E)$, for each edge $e \in T$, we can associate a cut S_e which is one side of the two components in $T - e$. The capacity of the cut $\delta(S_e)$ is $c(e)$. Let $D(e) = \sum_{i:\ |S_e \cup \{s_i, t_i\}|=1} D_i$ be the demand separated by e. The sparsity of the cut S_e is simply $c(e)/D_e$, hence finding the sparsest cut in this case is easy. Interestingly, the linear program relaxation give an optimum solution on a tree.

Proof. Let (x, y) be a feasible solution to the dual of the linear program relaxation with objective value λ. We want to prove that if G is a tree T, then there is an edge $e \in T$ such that $c_e/D_e \leq \lambda$. We note that by considering the compact linear program with distance variables, we have

$$\lambda = \frac{\sum_{e \in E} c(e) x_e}{\sum_{i=1}^{k} D_i d_x(s_i, t_i)}$$

where $d_x(s_i, t_i)$ is the shortest path distance between s_i and t_i induced by x. Since there is a unique path P_{s_i,t_i} from s_i to t_i in T such that $d_x(s_i, t_i) = \sum_{e \in P_{s_i,t_i}} x_e$, we have

$$\lambda = \frac{\sum_{e \in E} c(e) x_e}{\sum_{i=1}^{k} D_i d_x(s_i, t_i)} = \frac{\sum_{e \in E} c(e) x_e}{\sum_{i=1}^{k} D_i \sum_{e \in P_{s_i,t_i}} x_e} = \frac{\sum_{e \in E} c(e) x_e}{\sum_{e \in E} x_e \sum_{i:\ e \in P_{s_i,t_i}} D_i} = \frac{\sum_{e \in E} c(e) x_e}{\sum_{e \in E} D(e) x_e}.$$

Finally, the result follows from $\sum_i a_i / \sum_i b_i \geq \min_i a_i/b_i$ for positive a_i and b_i's. ⊛

Example (Ring). For a ring graph, the same technique works where we need to remove two edges.

The reason why the above proof works for trees is because of a more general phenomenon: the shortest path distances are ℓ_1 metrics, or equivalently, cut metrics.

Cut, Line, and ℓ_1 Metrics, and Metric Embedding

To explain the above phenomenon, consider a finite metric space (V, d) with following metrics:

Definition 2.3.2 (Cut metric)**.** Let (V, d) be a finite metric space. The metric d is a *cut metric* if there is a set $S \subseteq V$ such that $d = d_S$, where d_S associated with the cut S is defined as

$$d_S(u, v) = \begin{cases} 1, & \text{if } |S \cap \{u, v\}| = 1; \\ 0, & \text{otherwise.} \end{cases}$$

The cut-cone consists of non-negative combination of cut metrics:

Definition 2.3.3 (Cut cone)**.** Let (V, d) be a finite metric space. The metric d is in the *cut cone* if there exist non-negative scalars y_S where $S \subseteq V$ such that for all $u, v \in V$,

$$d(u, v) = \sum_{S \subseteq V} y_S d_S(u, v).$$

Beside the cut metric, another useful metric is the line metric and the well-known ℓ_1 metric:

Definition 2.3.4 (Line metric)**.** Let (V, d) be a finite metric space. The metric d is a *line metric* if there is a mapping $f \colon V \to \mathbb{R}$ such that for all $u, v \in V$,

$$d(u, v) = |f(u) - f(v)|.$$

Definition 2.3.5 (ℓ_1 metric)**.** Let (V, d) be a finite metric space. The metric d is an ℓ_1 *metric* if there is some integer d and a mapping $f \colon V \to \mathbb{R}^d$ such that for all $u, v \in V$,

$$d(u, v) = \|f(u) - f(v)\|_1$$

It might not be too surprising that the following holds.

Lemma 2.3.1. A metric d of a metric space (V, d) is an ℓ_1 metric if and only if it is a non-negative combination of line metrics (in the cone of the line metrics).

Proof. If d is an ℓ_1 metric then each dimension corresponds to a line metric. Conversely, any non-negative combination of line metrics can be made into an ℓ_1 metric where each line metric becomes a separate dimension (scalar multiplication for a line metric is also a line metric). ■

A more interesting observation is that any cut metric d_S is a simple line metric: map all vertices in S to 0 and all vertices in $V \setminus S$ to 1. This leads to the following.

Lemma 2.3.2. A metric d is an ℓ_1 metric if and only if d is in the cut cone.

Proof. If d is in the cut cone, then it is a non-negative combination of the cut metrics, hence it is a non-negative combination of line metrics by the above observation, hence an ℓ_1 metric.

For the converse, it suffices to argue that any line metric is in the cut cone. Let $V = \{v_i\}_{i=1}^n$ and let d be a line metric on V. Without loss of generality, assume that the coordinates x_i of the points for each v_i corresponding to the line metric d are $x_1 \leq x_2 \leq \cdots \leq x_n$ on the real line. For $1 \leq i < n$, let $S_i = \{v_1, v_2, \ldots, v_i\}$. It is not hard to verify that $\sum_{i=1}^{n-1} |x_{i+1} - x_i| d_{S_i} = d$. ■

Now we have introduced all the necessary metrics we will use. Consider a finite metric space (V, d).

Claim. Any finite metric space can be viewed as one that is derived from the shortest path metric induced on a graph with some non-negative edge lengths.

If $G = (V, E)$ is a simple graph and $\ell \colon E \to \mathbb{R}_+$ are some edge-lengths, the metric induced on V depends both on the *topology* of G and the lengths as well, i.e., finite metrics can encode graph structure, hence it can be diverse. When trying to round we may want to work with simpler metric spaces.

Intuition (Embedding)**.** Embed a given metric space (V, d) into a simpler host metric space (V', d') via an embedding $f \colon V \to V'$.

Note. Even though we may be interested in finite metric spaces, the host metric space can be continuous or infinite such as the $\mathbb{R}^h$ for dimension h.

As embedding typically distorts the distances, thus, we want to find embeddings with small distortion.

Definition 2.3.6 (Distortion). Let (V, d) and (V', d') be two metric spaces and let $f\colon V \to V'$ be an embedding. The *distortion* of f is given by[a]

$$\max_{\substack{u,v\in V\\ u\neq v}} \left(\frac{d'(f(u), f(v))}{d(u,v)}, \frac{d(u,v)}{d'(f(u), f(v))} \right).$$

[a] Additive version are also explored, although they are very restrictive due to lack of scale invariance.

Additionally, we're interested in the following kind of embeddings:

Definition. Let (V, d) and (V', d') be two metric spaces and let $f\colon V \to V'$ be an embedding.

Definition 2.3.7 (Isometric embedding). f is an *isometric embedding* if $d(u,v) = d'(f(u), f(v))$ for all $u, v \in V$.

Definition 2.3.8 (Contraction). f is a *contraction* if $d(u,v) \geq d'(f(u), f(v))$ for all $u, v \in V$.

Definition 2.3.9 (Non-contracting). f is *non-contracting* if $d(u,v) \leq d'(f(u), f(v))$ for all $u, v \in V$.

Of particular importance are embeddings of finite metric spaces into $\mathbb{R}^h$, where the distance in the host space is measured under a norm such as ℓ_p norm. The dimension h is also important in various applications but in some settings like with non-uniform sparsest cut, it is not. We assume the following:

Theorem 2.3.1 (Bourgain). Any n-point finite metric space can be embedded into ℓ_2 (hence also ℓ_1) with distortion $O(\log n)$. Moreover, the embedding is a contraction and can be constructed in randomized polynomial time and embeds points into $\mathbb{R}^h$ where $h = O(\log^2 n)$.

In fact, one can obtain a refined version of Theorem 2.3.1 that is useful for non-uniform sparsest cut.

Theorem 2.3.2 (Bourgain). Let (V, d) be an n-point finite metric space and let $S \subseteq V$ with $|S| = k$. Then there is a randomized polynomial time algorithm to compute an embedding $f\colon V \to \mathbb{R}^{O(\log^2 n)}$ such that the embedding is a contraction[a] and for every $u, v \in S$, $\|f(u) - f(v)\|_1 \geq cd(u,v)/\log k$ for some universal constant c.

[a] I.e., $\|f(u) - f(v)\|_1 \leq d(u,v)$ for all $u, v \in V$

We now see how can one utilize Theorem 2.3.2 with the previous insight we have on tree to provide a general guarantee.

As previously seen. The integrality gap of the linear program is 1 on trees since the shortest path metric on trees is in the cut cone, i.e., ℓ_1-embeddable.

More generally, one can prove that if the shortest path metric on a graph G embeds into ℓ_1 with distortion α, then the integrality gap of the linear program is at most α. This will imply an $O(\log n)$-integrality gap via Theorem 2.3.2 since any n-point finite metric space embeds into ℓ_1 with distortion $O(\log n)$.

Lecture 8: Randomized Rounding for Sparsest Cut and Expanders

19 Sep. 11:00

Randomized Rounding Algorithm with Metric Embeddings

Now, we see how to utilize Theorem 2.3.2 to design a randomized rounding algorithm for the non-uniform sparsest cut problem. The main theorem is the following.

Theorem 2.3.3. Let $G = (V, E)$ be a graph. Suppose any finite metric induced by edge lengths on E can be embedded into ℓ_1 with distortion α, then the integrality gap of the linear program for the non-uniform sparsest cut is at most α for any instance on G.

Proof. Let (x, y) be a feasible fraction solution of the linear program relaxation, and let d be the metric induced by edge lengths given by x. Let λ be the value of the solution, i.e.,

$$\lambda = \frac{\sum_{uv \in E} c(uv) d(u, v)}{\sum_{i=1}^{k} D_i d(s_i, t_i)}.$$

Since d can be embedded into ℓ_1 with distortion at most α, and any ℓ_1 metric is in the cut cone from Lemma 2.3.2, it implies that there are scalars z_S, $S \subseteq V$, such that for all $u, v \in V$,

$$\frac{1}{\alpha} \sum_{S \subseteq V} z_S d_S(u, v) \leq d(u, v) \leq \sum_{S \subseteq V} z_S d_S(u, v).$$

Without loss of generality, we assume that the embedding is a contraction. Then, we have

$$\begin{aligned} \lambda = \frac{\sum_{uv \in E} c(uv) d(u, v)}{\sum_{i=1}^{k} D_i d(s_i, t_i)} &\geq \frac{1}{\alpha} \frac{\sum_{uv \in E} c(uv) \sum_{S \subseteq V} z_S d_S(u, v)}{\sum_{i=1}^{k} D_i \sum_{S \subseteq V} d_S(s_i, t_i)} \\ &= \frac{1}{\alpha} \frac{\sum_{S \subseteq V} z_S c(\delta_G(S))}{\sum_{S \subseteq V} z_S D(\delta_H(S))} \geq \frac{1}{\alpha} \min_{S \subseteq V} \frac{c(\delta_G(S))}{D(\delta_H(S))}. \end{aligned}$$

Hence, there is a cut whose sparsity is at most $\alpha\lambda$. ■

Theorem 2.3.3 shows that one of the cuts with $z_S > 0$ has sparsity at most $\alpha\lambda$. Now, suppose we have an ℓ_1 embedding into h-dimensions, i.e., $\mathbb{R}^h$. Observe the following.

Intuition. First, each dimension in $\mathbb{R}^h$ corresponds to a line embedding, and each line embedding is in the cut cone with only $n - 1$ cuts used to express it (recall Lemma 2.3.2). Thus, given an ℓ_1 embedding into $\mathbb{R}^h$ with distortion α, we only need to try $d(n - 1)$ cuts and one of them will be guaranteed to have sparsity at most $\alpha\lambda$.

Algorithm 2.5 exploits this intuition.

Algorithm 2.5: Non-Uniform Sparsest Cut via Embedding

Data: A supply graph $G = (V, E)$ with edge capacity $c\colon E \to \mathbb{R}_+$, demand graph $H = (V, F)$ with demand capacity $D\colon F \to \mathbb{R}_+$
Result: The sparsest cut $S \subseteq V$

$(\{x_e\}_{e \in E}, \{y_i\}_{i=1}^{k}) \leftarrow$ `LP-Solve(Non-Uniform-Sparsest-Cut-LP(`G, c, H, D`))`
$f \leftarrow$ `Bourgain-Embedding(`(V, d_x)`)` // $f\colon V \to \mathbb{R}^h$, $h = O(\log^2 n)$
for $\ell = 1, \ldots, h$ **do**
  Sort ℓ^{th} coordinate of $\{f(v)\}_{v \in V}$ as $x_1^{(\ell)} \leq x_2^{(\ell)} \leq \cdots \leq x_n^{(\ell)}$ // $f(v)_\ell = x_i^{(\ell)}$ `for some` i
  for $j = 1, \ldots, n - 1$ **do**
    $S_j^{(\ell)} \leftarrow \{v \in V \mid f(v)_\ell \leq x_j^{(\ell)}\}$
$S \leftarrow \arg\min_{\ell \in [h], j \in [n-1]} c(\delta_G(S_j^{(\ell)}))/D(\delta_H(S_j^{(\ell)}))$
return S

The guarantee of Algorithm 2.5 can be derived from Theorem 2.3.3.

Theorem 2.3.4. Algorithm 2.5 outputs cuts of sparsity at most $\alpha\lambda^*$,[a] in particular, it's a randomized

$O(\log k)$-approximation algorithm for non-uniform sparsest cut.

[a] λ^* is the optimal solution of the dual linear program.

Proof. From Theorem 2.3.2, f is a contraction and with distortion $\alpha = O(\log k)$. From a similar argument as in Theorem 2.3.3, we see that for some z_S, $S \subseteq V$, we have

$$\begin{aligned}
\lambda^* &= \frac{\sum_{e\in E} c(e)x_e}{\sum_{i=1}^{k} D_i d_x(s_i,t_i)} \\
&\geq \frac{\sum_{uv\in E} c(uv)\|f(u)-f(v)\|_1}{\alpha \sum_{i=1}^{k} D_i \|f(s_i)-f(t_i)\|_1} \\
&= \frac{1}{\alpha}\frac{\sum_{uv\in E} c(uv)\sum_{\ell=1}^{h}|f(u)_\ell - f(v)_\ell|}{\sum_{i=1}^{k} D_i \sum_{\ell=1}^{h}|f(s_i)_\ell - f(t_i)_\ell|} \\
&= \frac{1}{\alpha}\frac{\sum_{\ell=1}^{h}\sum_{j=1}^{n-1}|x_{j+1}^{(\ell)} - x_j^{(\ell)}|c(\delta_G(S_j^{(\ell)}))}{\sum_{\ell=1}^{h}\sum_{j=1}^{n-1}|x_{j+1}^{(\ell)} - x_j^{(\ell)}|D(\delta_H(S_j^{(\ell)}))} = \frac{1}{\alpha}\frac{\sum_{S\subseteq V} z_S c(\delta_G(S))}{\sum_{S\subseteq V} z_S D(\delta_H(S))} \geq \frac{1}{\alpha}\min_{S\subseteq V}\frac{c(\delta_G(S))}{D(\delta_H(S))},
\end{aligned}$$

where the second last equality follows from the fact that $s_i, t_i \in V$ as well. ■

2.3.4 Line, ℓ_1, and Tree Embeddings, and State-of-the-Art

Theorem 2.3.2 shows that any finite metric space on n points embeds into ℓ_1 with distortion $O(\log n)$. Here, we hint on the underlying algorithm of the construction: from Lemma 2.3.1, ℓ_1 embeddings are a non-negative combination of line embeddings. A particular type of line embedding is the following.

Definition 2.3.10 (Fréchet embedding)**.** Let (V,d) be a metric space and let $S \subseteq V$. The *Fréchet embedding* is a contraction $f\colon V \to \mathbb{R}$ such that $f(v) = d(S,v)$.

Many results in embeddings into ℓ_p spaces are based on using Fréchet embeddings in various clever and often highly non-trivial ways. In particular, Theorem 2.3.2 is based on picking many random sets and combining the resulting Fréchet embeddings.

Now, we note that Theorem 2.3.2 can also be derived via probabilistic tree embeddings because every tree metric embeds into ℓ_1 isometrically. For general metrics, tree embeddings provide a more constrained space while yielding the same worst-case distortion. However, one can ask if ℓ_1 embeddings provide better distortion for concrete graph classes. This is indeed the case.

Example (Ring)**.** Consider a ring graph (a cycle with capacities). One can prove that tree embeddings require a distortion 2, while the ring metric can be isometrically embedded into ℓ_1. Thus, the flow-cut gap on ring is 1 which is not obvious.

Rather than looking at distortion, one can ask about the flow-cut gap obtained via different embeddings for a particular graph class. First, recall the followings for the non-uniform sparsest cut.

As previously seen (Theorem 2.3.4)**.** The flow-cut gap in general undirected graphs is $O(\log k)$.

Additionally, for general graph, a lower bound is also known.

Remark (Lower bound)**.** Expanders give a lower bound on the flow-cut gap to be $\Omega(\log k)$ even for uniform sparsest cut [LR99].

With these general bounds in mind, we now consider the flow-cut gap for planar graphs in particular.

Example (Planar graph)**.** There is a famous conjecture that the flow-cut gap in planar graphs is $O(1)$ [Gup+04]. Interestingly, for tree embeddings, there is a lower bound of $\Omega(\log n)$ even on the special case of planar graphs called series parallel graphs. Hence, tree embeddings are not powerful enough to prove the conjecture.

The best flow-cut gap so far is $O(\sqrt{\log n})$ via ℓ_1 embeddings, thus separating the general graph

case from the planar graph case. For series parallel graphs, we know that the flow-cut gap is a tight bound of 2 and establishing this tight bound took a fair amount of work.

For uniform sparsest cut, the flow-cut gap in planar graphs is $O(1)$ [KPR93]. One can show a tight connection between embeddability into ℓ_1 and flow-cut gap [Gup+04].

On the other hand, we can ask for a better approximation guarantee. First, note the following.

Note. Approximating the non-uniform sparsest cut problem is not the same as establishing the flow-cut gap as the flow-cut gap relies on the linear program relaxation.

Problem. Can we obtain a better approximation than $O(\log k)$ for non-uniform sparsest cut?

Answer. Yes! By using semi-definite programming based relaxation, one can obtain an $O(\sqrt{\log n})$-approximation for uniform sparsest cut [ARV09] and also the product instances. Based on this, an $O(\sqrt{\log n}\log\log n)$-approximation for non-uniform sparsest cut is achieved [ALN05; ALN07].[a] ⊛

[a] There was a conjecture that the SDP based relaxation would yield an $O(1)$-approximation, but it was shown that the integrality gap is essentially close to $\Omega(\sqrt{\log n})$.

2.3.5 Node Capacities[1]

A slight generalization of the uniform sparsest cut problem is obtained by considering demands induced by weights on vertices, which we refer to product instance.

Problem 2.3.4 (Product instance of sparsest cut)**.** Given a graph $G = (V, E)$ with edge capacity $c\colon E \to \mathbb{R}_+$ and vertex weight $\pi\colon V \to \mathbb{R}_+$. The *product instance of sparsest cut problem* is the non-uniform sparsest cut problem with demand $D(u, v)$ for edge $uv \in E$ setting to be $\pi(u)\pi(v)$.

Notation. In this case, the dual flow instances are called *product multi-commodity flow*.

We see that the product instances indeed generalizes uniform sparsest cut: If $\pi(u) = 1$ for all u, then this reduces to the uniform sparsest cut problem.

Remark (Subset sparsity)**.** If $\pi(u) \in \{0, 1\}$ for all u, then we are focusing our attention on the sparsity w.r.t. the set $V' = \{v \in V \mid \pi(v) = 1\}$. Since vertices with $\pi(u) = 0$ play no role.

2.4 Expander and Well-Linked Set

We now introduce expanders [HLW06], which relate to the non-uniform sparsest cut in an intricate way.

Definition 2.4.1 (Expander)**.** An *expander* with parameter α is a graph $G = (V, E)$ such that for all $S \subseteq V$ with $|S| \le |V|/2$ such that $|\delta(S)| \ge \alpha|S|$.

Definition 2.4.2 (Expansion)**.** The *expansion* of a graph $G = (V, E)$ is $\min_{S\colon |S|\le|V|/2}|\delta(S)|/|S|$.

Another related notion called conductance also has a nice connection to uniform sparsest cut.

Definition 2.4.3 (Conductance)**.** Given a graph $G = (V, E)$ and a cut $S \subseteq V$, the *conductance* $\phi(G)$ of G is defined as $|\delta(S)|/\operatorname{vol}(S)$ where $\operatorname{vol}(S) = \sum_{v\in S}\deg(v)$.

It's clear that G is an α-expander if the expansion of G is at least α. Initially, expansion arose from the graph bisection problem.

[1] We refer to the note for further information on further generalizations to node capacitated graph and directed graph.

Problem 2.4.1 (Graph bisection). Given a graph $G = (V, E)$, the *graph bisection problem* aims to find a partition of G into $(S, V \setminus S)$ such that $|S| = |V \setminus S|$ while minimizing $|\delta(S)|$.

However, as we will soon see, expander itself is quite interesting. Firstly, we note that expanders do exist, and they are quite common.

Lemma 2.4.1. There exists expanders with degree 3 with expansion $\alpha = \Omega(1)$. More specifically, a random 3-regular graph is an expander with high probability.

Hence, with Lemma 2.4.1, a natural way to generate an expander is to first sample a random regular graph, then check its expansion. However, computing the expansion is coNP-hard.

2.4.1 Expansion and Conductance via Sparsest Cut

The first connection of expander to the uniform sparsest cut is that the latter can be used to find the expansion of a graph. In particular, we see that when $|S| \le |V|/2$, we have

$$\frac{1}{|V|}\frac{|\delta(S)|}{|S|} \le \frac{|\delta(S)|}{|S||V \setminus S|} \le \frac{2}{|V|}\frac{|\delta(S)|}{|S|}.$$

Remark. The expansion and the uniform sparsest cut's sparsity are within a factor of 2 of each other. Hence, while determining the expansion exactly is coNP-hard, we can use uniform sparsest cut to certify the expansion of a graph within a factor of 2.

Sometimes it is useful to consider expansion with vertex weights $w \colon V \to \mathbb{R}_+$ as well. In this case, the expansion is defined as $\min_{S \colon w(S) \le w(V)/2} |\delta(S)|/w(S)$.

Note. It's dual corresponds to the *product multi-commodity flow instances* where $\pi(v) = w(v)$.

As for conductance, it's easy to see that one can capture it by expansion via setting weights on vertices with $w(v) = \deg(v)$, which further reduces to the product instance of sparsest cut.

Claim. For regular graphs, expansion and conductance are the same.

2.4.2 Spectral Relaxation for Conductance

In several applications it is important to obtain constant-degree expanders with constant expansion.

Intuition. The $O(\sqrt{\log k})$-approximation algorithms we saw are not useful in this regime.

It turns out that there is a very different method based on spectral graph theory that helps in this regime. For an undirected graph on n vertices, consider the Laplacian $\mathcal{L}_G := D - A$, where D is the diagonal degree matrix and A is the adjacent matrix. In particular, we have

$$(\mathcal{L}_G)_{ij} := \begin{cases} \deg(v_i), & \text{if } i = j; \\ -1, & \text{if } i \neq j \text{ and } A_{ij} = 1; \\ 0, & \text{otherwise.} \end{cases}$$

Since $\mathcal{L}_G$ is symmetric, all its eigenvalues are real. Moreover, this matrix is also positive semi-definite, hence all its eigenvalues are actually non-negative. Let $0 = \lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$ be its eigenvalues, then a well-known and famous result in spectral graph theory is the following.

Theorem 2.4.1 (Chegger's inequality). Given a graph G with conductance $\phi(G)$,

$$\frac{\lambda_2}{2} \le \phi(G) \le \sqrt{2\lambda_2}.$$

Remark. λ_2 provides a constant factor approximation for the conductance when it is a constant!

Since the expansion and conductance are related by the maximum degree, when the degree is a small constant, one can use λ_2 to certify expansion. Due to its importance for certifying expansion/conductance, some use λ_2 as the definition of expansion since it is computable and also helps in construction of expanders.

2.4.3 Expander Decomposition

Even if a graph is not an expander at first, it is possible to decompose a graph into smaller subgraphs such that each of them has good expansion/conductance. More explicitly, the goal is to remove as few edges as possible such that the graph decomposes into expanders. This is useful since expander leads to good algorithms. The question is the trade-off between the number of edges that we remove and the expansion that we can guarantee for the pieces.

Notation. Technically the process works with conductance, but we use the terminology of expander decomposition for historical reasons.

Since one is often interested in finding fast algorithms for expander decomposition, it is common to explore trade-offs between the quality of the conductance of the pieces and the number of edges.

Definition 2.4.4 (Expander decomposition). A (ϕ, ϵ)-*expander decomposition* for some $\epsilon \in (0,1)$ is a partition of the (connected) graph $G = (V, E)$ into vertex induced subgraphs $\{G_i = G[V_i]\}_{i=1}^h$ such that each G_i has conductance at least ϕ, and the number of inter-cluster edges is at most ϵm, i.e.,

$$\frac{1}{2}\sum_{i=1}^{h} |\delta(V_i)| \leq \epsilon m.$$

Lecture 9: Expander Decomposition and Well-Linked Sets

24 Sep. 11:00

Theorem 2.4.2. Let $G = (V, E)$ be a graph and $\epsilon \in (0,1)$. Suppose there is an α-approximation for the uniform sparsest cut, then there is an efficient algorithm that outputs an $(\Omega(\frac{\epsilon}{\alpha \log m}), \epsilon)$-expander decomposition of G.

Proof. Consider the following algorithm.

Algorithm 2.6: Expander decomposition

Data: A connected graph $G = (V, E)$, base graph edge set size M, parameter $\epsilon \in (0,1)$
Result: An $(\Omega(\frac{\epsilon}{\alpha \log m}), \epsilon)$-expander decomposition $\{G_i\}_{i=1}^h$

```
if m ≤ 10 log M/ε then                                   // Base case
    return G

(S, V \ S) ←α-Uniform-Sparsest-Cut(G)[a]                 // vol(S) ≤ vol(V \ S)

if |δ(S)|/vol(S) > c/10 log M then                       // Check sparsity
    return G
else
    {G_i^(1)}_{i=1}^{h1} ←Expander-Decomposition(G[S], M, ε)
    {G_i^(2)}_{i=1}^{h2} ←Expander-Decomposition(G[V \ S], M, ε)
    return {G_i^(1)}_{i=1}^{h1} ∪ {G_i^(2)}_{i=1}^{h2}
```

[a] Recall that this is for conductance, which can be formalized as a project instance.

Claim. The conductance of each subgraph output by Algorithm 2.6 is at least $\epsilon/10\alpha \log M$.

Proof. For the base case, if G is connected and has at most $10 \log M/\epsilon$ edges, then the conductance of G is at least $c/10 \log M$ since at least one edge crosses any cut, and the volume of the smaller side is at most $10 \log M$.

On the other hand, if the α-approximation algorithm of the uniform sparsest cut for conductance outputs a cut $(S, V \setminus S)$ with sparsity at least $c/10 \log M$, we know that the actual sparsity of G is at least $c/10\alpha \log M$ as desired. ⊛

Next, we analyze the total number of edges cut, which needs to be at most ϵm.

Intuition. If G is of constant size (and connected) or it does not have a sparse cut, Algorithm 2.6 does not cut any edges.

Let $T(m)$ be the total number of edges cut by Algorithm 2.6 on a graph with m edges. Algorithm 2.6 removes edges between S and $V \setminus S$ only when $|\delta(S)| \leq \epsilon \operatorname{vol}(S)/10 \log M$ where $\operatorname{vol}(S) \leq \operatorname{vol}(V \setminus S)$. Let $m' \coloneqq |\delta(S)|$, $m_1 \coloneqq |E(G[S])|$, and $m_2 \coloneqq |E(G[V \setminus S])|$, then

$$m' \leq \frac{\epsilon}{10 \log m}(2m_1 + m') \Rightarrow (1 - o(1))m' \leq \frac{\epsilon}{5 \log m} m_1 \leq \frac{\epsilon}{4 \log M} m_1.$$

With $m_1 \leq m_2$, the recurrence can be written as

$$T(m) \leq T(m_1) + T(m_2) + \frac{\epsilon}{4 \log M} \min(m_1, m_2) = T(m_1) + T(m_2) + \frac{\epsilon}{4 \log M} m_1$$

where $m_1 + m_2 \leq m$, which gives $T(m) \leq \epsilon m$. ■

If we don't care about efficiency, we can set $\alpha = 1$ and solve the uniform sparsest cut exactly. In particular, Theorem 2.4.2 guarantees that the decomposed pieces have conductance $\Omega(1/\log m)$ while cutting only a constant fraction of the edges.

Note. The bound $\Omega(1/\log m)$ is tight as shown by the hypercube [Ale+17].

We can rephrase Theorem 2.4.2 in a different form where we want a lower bound on the conductance of the pieces and express the number of edges cut as a function that parameter:

Corollary 2.4.1. Let $G = (V, E)$ be a graph and ϕ be a parameter. Suppose there is an α-approximation for the uniform sparsest cut, then there is an efficient algorithm that computes a $(\phi, O(\alpha \cdot \phi \cdot \log m))$-expander decomposition.

Note. Number of edges cut is less than m only if $\alpha\phi \log m < 1$, so one should think of $\phi \leq 1/\alpha \log m$.

Remark. Theorem 2.4.2 is phrased in terms of m, the number of edges. Capacitated graphs can be handled by scaling since we do not assume that G is simple. However, the dependence on $\log m$ means that when capacities are large, we are not guaranteed a strongly polynomial bound. One can handle this issue in various ways depending on the application. In most applications of expander decomposition, it is the case that the total capacity of the edges can be assumed to be polynomially bounded in n and in this case, the $\log m$ factor is typically replaced with $\log n$.

We remark that Algorithm 2.6 is based on sparsest cut algorithms. Traditionally, these algorithms were quite slow. There have been several developments in the last few years which enabled sparsest cut to be reduced to a poly-logarithmic number of s-t flows via the so-called *cut-matching game* [KRV09; Ore+08], which in turn enabled faster flow algorithms. There are now near-linear time randomized algorithms for expander decomposition (with slightly weaker parameters than the ideal one) for the regimes of interest [SW19]. In some applications the randomized algorithm is not adequate and there has been considerable effort to obtain deterministic algorithms. There are now almost-linear time deterministic algorithm [Chu+20; SL21].

2.4.4 Well-Linked Set

Consider the following generalization of Definition 2.4.1, where we only care about expansion of a subset.

Definition 2.4.5 (Well-linked). A set $X \subseteq V$ is *well-linked* in a graph $G = (V < E)$ if for all $S \subseteq V$,

$$|\delta(S)| \geq \min(|S \cap X|, |S \cap (V \setminus X)|).$$

On the other hand, recall that α-expansion means that for all sets $S \subseteq V$ with $|S| \leq |V|/2$, $|\delta(S)| \geq \alpha|S|$. This is a cut condition. Suppose A, B are two disjoint sets of vertices of equal size $|A| = |B|$, clearly we have $|A|, |B| \leq |V|/2$. We can ask for a similar guarantee as the same cut condition. This turns out to be another generalization of expander:

Definition 2.4.6 (Linkage). Let $A, B \subseteq V$, $A \cap B = \varnothing$, and $|A| = |B|$. An A-B *linkage* is a set of edge-disjoint paths connecting A to B with each vertex in $A \cup B$ in exactly one path. In this case, we say A and B are *linked* in G.

Note. We do not have to insist on $A \cap B = \varnothing$. If not and we allow each vertex in $A \cap B$ to connect to itself via an empty path, then it is the same as asking $A \setminus B$ and $B \setminus A$ to be linked. Thus, requiring $|A| = |B|$ suffice.

We can view linkage as sending flows:

Definition 2.4.7 (Fractional linkage). An A-B linkage is *fractional* if there is a flow in G with that satisfies demand of 1 on each vertex in A and a demand of -1 on each vertex of B.[a] In particular, we say that A, B are α-*linked* for some parameter α if there is a flow in G that satisfies the demand of α on each vertex in A and a demand of $-\alpha$ on each vertex of B.

[a] Note that this corresponds to a single-commodity flow.

Lemma 2.4.2. Suppose G is an α-expander with $\alpha \geq 1$. Then there is an A-B linkage in G for every pair of disjoint equal sized sets A, B.

Proof. ■

One can scale capacities or directly prove the following.

Corollary 2.4.2. Suppose G is an α-expander. Then if A, B are disjoint vertex sets with $|A| = |B|$, then A, B are α-linked.

Interestingly, the converse is also true.

Lemma 2.4.3. Suppose G is a graph and for any two disjoint set A, B of equal size, A, B are α-linked. Then G is an α-expander.

The following shows the connection of linkage and well-linked.

Claim. A set X is well-linked if for all $A, B \subseteq X$ and $|A| = |B|$, A and B are linked.

Definition 2.4.8 (Fractinoal well-linked). A set X is α-*well-linked* in a graph G is for any two $A, B \subseteq X$ with $|A| = |B|$, the sets A, B are α-linked.

More generally, we have the following.

Lemma 2.4.4. A set $X \subseteq V$ is α-well-linked in G if and only if for any set $S \subseteq V$, $|\delta(S)| \geq \alpha \min(|S \cap X|, |S \cap (V \setminus X)|)$.

Corollary 2.4.3. A graph $G = (V, E)$ is an α-expander if and only if V is α-well-linked in G.

Thus, the notion of well-linked sets extends the definition of expansion to subsets of the graph. This is very useful in a number of settings.

Example (Star). A start on n vertices is an expander and has a well-linked set of size n. This strange artifact is because of the large degree of the center vertex.

This artifact disappears if we ask for constant degree graphs or if we insist on node linkage, i.e., we now want *node-disjoint paths*.

Definition 2.4.9 (Node linkage). Let $A, B \subseteq V$, $A \cap B = \varnothing$, and $|A| = |B|$. An A-B *node linkage* is a set of node-disjoint paths connecting A to B with each vertex in $A \cup B$ in exactly one path.[a]

[a] We also skip the definition of α-linkage for now. The definition is basically the same where we want flow with node capacities rather than edge capacities.

Definition 2.4.10 (Node well-linked). A set X is *α-node-well-linked* in G if for any two $A, B \subseteq X$ with $|A| = |B|$, A, B are α-linked.

Intuition. If X is node-well-linked in a graph, then X cannot have a sparse node separator, i.e., if S separates $G \setminus S$ into components and S does not have any vertices of X then no component of $G \setminus S$ can have more than $|S|$ vertices of X.

In graph theory literature on treewidth, the notion of linkages is defined primarily via node-disjoint paths. We will not use treewidth very often in this course hence we overload edge and node well-linked notations. In particular, its connection to node-well-linkedness is the following.

Theorem 2.4.3. Let k be the cardinality of the largest node-well-linked set in a graph G. Then $k \leq \mathrm{tw}(G) \leq 4k$.

Remark. In fact, most algorithmic approaches to computing treewidth are based on algorithms for sparse node separator computations.

Node-well-linkedness is connected to vertex-expanders. Sometimes people do not distinguish between these two notions too much in the expansion literature because of the following.

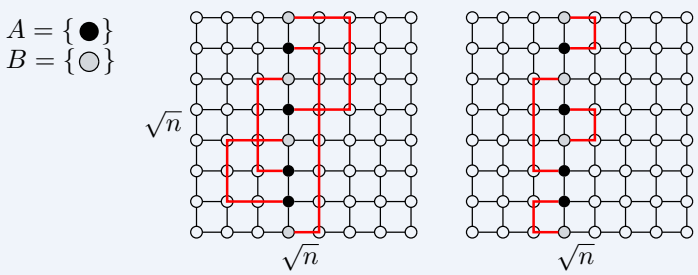
Claim. If G is an α-edge-expander with maximum degree d, then G is an $\Omega(\alpha/d)$-vertex-expander.

Thus, if one is working with constant degree graphs, the two notions are not very far.

Example (Star). Consider the start graph again. One can see that it is an edge-expander, but it is very far from being a vertex expander. In fact, the largest node-well-linked set in a star is of size 2.

On the other hand, the grid is not only edge-well-linked, but also node-well-linked.

Example (Grid). A $\sqrt{n} \times \sqrt{n}$ grid is a planar graph with n vertices. It has a bisection with $O(\sqrt{n})$ edges hence it is at best a $1/\sqrt{n}$-expander (which in fact it is). It has a well-linked set of size $\Omega(\sqrt{n})$, i.e., rows or columns X of $\sqrt{n}$ vertices. We see that although G is not a good expander, but it has good expansion w.r.t. X. Actually, it's even node well-linked (right).



In some sense, grid is the best planar graph in terms of node-well-linkedness.

Theorem 2.4.4. Every planar graph has a balanced separator of size $O(\sqrt{n})$. Hence, no planar graph on n vertices have a node-well-linked set of size more than $c\sqrt{n}$ for some fixed constant c.

Theorem 2.4.5 (Hierarchical expander decomposition)**.** Given a graph $G = (V, E)$ with edge capacity $c\colon E \to \mathbb{R}_+$, there is a rooted tree $T = (V_T, E_T)$ such that

(a) leaves of T are V,

(b) every edge $e \in E_T$ is associated with capacity of the cut induced by T_u, and

(c) any demand matrix D that is reachable in T is reachable in G with congestion $O(\log^3 n)$.

Note. Hierarchical expander decomposition has found many applications in fast graph algorithms recently [Gor+21].

Lecture 10: Tree-Based Oblivious Routing

2.5 Oblivious Routing

26 Sep. 11:00

Consider routing demands between source-sink pairs in a network $G = (V, E)$ with capacity $c\colon E \to \mathbb{R}_+$. In particular, the actual demands are not known in advance and come and go in an online fashion. Hence, the problem is, which routes should the demand for some specific pair (s, t) be routed on?

Intuition. To feasibly route a given set of demands, we need to essentially solve a multi-commodity flow, which helps to balance the network's capacity among many competing pairs.

This is a non-trivial problems and leads to lots of the breakthroughs. Oblivious routing is one of which that we will be interested in. The goal in oblivious routing, in some sense, is to bridge the static setting where the demands are all fully known to the fully online setting where we specify a route to a new demand when it arrives.

Problem 2.5.1 (Oblivious routing)**.** Given a directed graph $G = (V, E)$ with edge capacity $c\colon E \to \mathbb{R}_+$, the goal of *oblivious routing* is to output a distribution of paths $\mathcal{P}_{u,v}$ between every pair of $u, v \in V$ such that when the demand $D\colon V \times V \to \mathbb{R}_+$ come, specifically, $D(u, v)$ for a pair $u, v \in V$, a flow is routed on a path $P \in \mathcal{P}_{u,v}$ according to the distribution.

Alternatively, we can also think of sending $D(u, v)$ demand fractionally along the paths in proportion to the probability. Clearly, Problem 2.5.1 is oblivious since the probability distribution over $\mathcal{P}_{u,v}$ is specified before knowing any of the actual demands. In particular, if we know the entire set of demands in advance, we can compute an optimal routing via some sort of multi-commodity flow computation. The question now is that, how well can an oblivious routing do when compared to a static routing, and how to measure the quality. More fundamentally, do good oblivious routing even exists?

The initial work that motivated the problem of oblivious routing is that good deterministic oblivious routing exist in special classes of graphs such as hypercubes [VB81]; later the need for randomization in the sense of picking paths randomly as we described was realized.

2.5.1 Routable Demands and Congestion

We first digress towards defining routable demands and some properties.

Definition 2.5.1 (Routable)**.** A demand matrix $D \in \mathbb{R}_+^{n \times n}$ is *routable* in a directed graph $G = (V, E)$ with capacity $c\colon E \to \mathbb{R}_+$ if there exists a feasible multi-commodity flow for D in G.

It's easy to see that $\mathcal{D}_G := \{D \in \mathbb{R}_+^{n \times n} \mid D \text{ is routable}\}$ is a convex set. Moreover, given a demand matrix D, we can efficiently check if D is routable in G via solving a linear program for multi-commodity flow. Thus, we have a membership oracle for the convex set $\mathcal{D}_G$.

Intuition. The flow linear program is nice because it is also linear in $D(u,v)$ values, hence we can claim something stronger.

Suppose we have $w(u,v)$ for all $u,v \in V$. We ask the following question.

Problem 2.5.2 (Optimiznig demand polytope)**.** Can we solve $\max_{D \in \mathcal{D}_G} \sum_{u,v} w(u,v)D(u,v)$?

Answer. By the equivalence of optimization and separation, we have an efficient separation oracle over $\mathcal{D}_g$! That is, there is an efficient algorithm that given $D \in \mathbb{R}_+^{n \times n}$, either outputs that $D \in \mathcal{D}_G$ or outputs a separating hyperplane that separates D from the convex set $\mathcal{D}_G$. ⊛

The above is a convoluted but easy way of deriving a useful fact. A more direct way is often referred to as the Japanese theorem on multi-commodity flow, which can be derived from linear program duality, as we saw in non-uniform sparsest-cut.

Lemma 2.5.1 (Japanese theorem)**.** The demand D is routable if and only if for all $\ell \colon E \to \mathbb{R}_+$,

$$\sum_{e \in E} c(e)\ell(e) \geq \sum_{u,v \in V} D(u,v) d_\ell(u,v)$$

where d_ℓ is the shortest path distance induced by ℓ.

Hence, to prove that D is not routable in G, it suffices to produce one length function $\ell \colon E \to \mathbb{R}_+$ that violates the inequality in Lemma 2.5.1.

Intuition. One can view $\mathcal{D}_G$ as being defined by an infinite set of linear inequalities, one for each non-negative length function.

Next, to measure the quality of oblivious routing, the central notion is the congestion.

Definition 2.5.2 (Congestion)**.** A demand matrix $D \in \mathbb{R}_+^{n \times n}$ is *routable with congestion* $\rho > 0$ in a directed graph $G = (V,E)$ with capacity $c \colon E \to \mathbb{R}_+$ if D is routable in G where edge capacities are multiplied by ρ.

When $\rho = 1$, it corresponds to the routable case. Moreover, the set of all routable demand metrics in G with a fixed congestion ρ is also convex.

2.5.2 Oblivious Routing to Minimize Congestion

In oblivious routing, we want to specify a probability distribution over $\mathcal{P}_{u,v}$ for each pair (u,v). Naively, since there are exponential many paths, specifying a probability on each path can be tricky.

Intuition. One can view such a probability distribution as a flow of one unit from u to v where the flow on a path $P \in \mathcal{P}_{u,v}$ is equal to the probability of P.

Thus, one compact way to represent a probability distribution over paths is via a unit flow via flow values on edges, which can be specified using only m numbers. Formally, consider the following:

Definition 2.5.3 (Oblivious routing scheme)**.** Given a graph $G = (V,E)$, consider the following.

> **Definition 2.5.4** (Edge-based oblivious routing)**.** An *edge-based oblivious routing* is a collection of unit flows in G, one for each $(u,v) \in V \times V$, specified via edge-based flows $f_e^{(u,v)}$.
>
> **Definition 2.5.5** (Path-based oblivious routing)**.** A *path-based oblivious routing* is similar to edge-based where the unit flow is specified via a path based flow.

However, edge-based flow does not uniquely specify a path-based flow, so we are losing information by using the compact representation, but it is helpful in some computational situations.

Note. $f_e^{(u,v)}$ is a number in $[0,1]$, and whether the oblivious routing is specified via edge-based or path-based, this quantity is well-defined.

Now, for oblivious routing, we can define its congestion as follows.

Definition 2.5.6 (Congestion of oblivious routing)**.** Let D be a demand matrix. The *congestion* incurred for D via the oblivious routing specified by $f_e^{(u,v)}$ for $e \in E$ and $(u,v) \in V \times V$ is

$$\max_{e \in E} \frac{\sum_{u,v \in V} D(u,v) f_e^{(u,v)}}{c(e)}.$$

Moreover, the *congestion* of an oblivious routing is the smallest $\rho \geq 1$ such that every routable demand $D \in \mathcal{D}_G$ is routed with congestion at most ρ by the oblivious routing.

We say that an optimal oblivious routing for G is the one which achieves the smallest ρ among all oblivious routings. In other words, we are asking how much should we scale up the capacities of G so that any demand matrix D that is routable in G can be routed in G via the oblivious routing, i.e.,

$$\max_{D \in \mathcal{D}_G} \max_{e \in E} \frac{\sum_{u,v \in V} D(u,v) f_e^{(u,v)}}{c(e)}.$$

Claim. Every graph on n vertices has an oblivious routing scheme with congestion n^2.

Proof. Consider computing the max-flow for every pair and then scale it down to a unit flow. ⊛

We give some pointers to the fundamental results on oblivious routing in a chronological order:

- Harald Räcke proved that every undirected graph admits an oblivious routing with congestion $O(\log^3 n)$ [Räc02]. The initial proof is based on an optimal algorithm for non-unifrom sparsest cut, hence does not lead to an efficient algorithm to construct the oblivious routing. Soon after, efficient algorithms are known with the congestion being $O(\log^2 n \log\log n)$ [BKR03; HHR03]. These results are based on the hierarchical expander decomposition of the graph which results in a compact cut representation of every graph.
- Via a simple idea in retrospect, that the optimal oblivious routing can be computed efficiently via linear program techniques even for directed graphs [Aza+03]. Note that being able to compute an optimum oblivious routing for any given graph does not tell us an easy way to understand a universal bound. We will see this next.
- Oblivious routing in directed graphs requires congestion $\Omega(\sqrt{n})$ [Aza+03]; this lower bound holds even for the restricted case of single-source demands. A similar lower bound was shown to hold also for undirected graphs with node capacities [Haj+07]. Recently, the lower bound for directed graphs has been improved to $\Omega(n)$ [Ene+16].
- In another breakthrough, Harald Räcke made a beautiful and surprising connection between oblivious routing and probabilistic tree embeddings for metric distortion [Räc08] and obtain an optimal $O(\log n)$-congestion tree-based oblivious routing scheme. Via this algorithm, an $O(\log n)$-approximation for the minimum bisection problem in graphs is developed.

2.5.3 Efficient Algorithm for Optimal Oblivious Routing

In this section, we will prove that one can find an optimal edge-flow based oblivious routing in polynomial time [Aza+03]. In particular, we will consider directed graphs since it is easier to define edge-based flows. To find an edge-flow based oblivious routing with minimum congestion, we write the following

linear program which essentially follows from the definition:

$$\begin{aligned} \min\ & \rho \\ & f_e^{(u,v)} \text{ defines a unit flow from } u \text{ to } v && \forall e \in E, \forall (u,v) \in V \times V; \\ & \sum_{(u,v)\in V\times V} D(u,v) f_e^{(u,v)} \le \rho c(e) && \forall e \in E, \forall D \in \mathcal{D}_G; \\ & f_e^{(u,v)} \ge 0 && \forall e \in E, \forall (u,v) \in V \times V. \end{aligned} \tag{2.5}$$

Note. We omit writing down the linear constraints for the unit flow from u to v since it's easy.

The only catch in solve Equation 2.5 is that it has an infinite number of constraints, with polynomially many (mn^2) variables. Hence, one can use the ellipsoid method if we have an efficient separation oracle.

Claim. There is an efficient separation oracle for Equation 2.5.

Proof. Given $f_e^{(u,v)}$ and ρ, we see that checking unit flow is easy. For the second constraint, given any fix edge e, we simply maximize $\sum_{u,v\in V} D(u,v) f_e^{(u,v)}$ via Problem 2.5.2 (efficiently!) and compare it with $\rho c(e)$. Hence, by iterating through all $e \in E$, we're done. ⊛

While the resulting algorithm is not very efficient in practice, but it shows the power of the ellipsoid method and working with large implicit linear programs fearlessly. One can use multiplicative weight updates and other techniques to obtain fast approximation algorithms for some of these problems.

Remark. This technique to design optimum oblivious routing is fairly general. We do not need to consider the full set of demand matrices $\mathcal{D}_G$, as long as we have a nice convex set of demand matrices that we can optimize over, we can find an optimum oblivious routing when restricted to those demand matrices.

2.5.4 Tree-Based Oblivious Routing via Duality and Low Stretch Trees

We now prove that in undirected graphs, there is always an oblivious routing with congestion $O(\log n)$, which is an optimal bound. Räcke prove this result via an elegant connection to tree embeddings for distance preservation [Räc08]. The bound, the connection, and the tree-based aspect are all important.

As previously seen. Probabilistic approximation of a graph by spanning trees for distances incurs an $O(\log n \log\log n)$ distortion [ABN08] while we can obtain an optimal $O(\log n)$ distortion if we allow dominating tree metrics (recall the differences).

It is easier to understand the ideas, and in particular the notation, by working with spanning trees than with general hierarchical tree representations (Theorem 2.4.5) of graphs (in some cases, the spanning tree result is useful and needed). The difference in the distortion is insignificant for our purposes here, and we will use these results in a black-box fashion.

Now, let $\alpha(n)$ denote the best bound that we can obtain for approximating the distance in an n-node graph $G = (V, E)$ with edge length $\ell\colon E \to \mathbb{R}_+$ by a probability distribution over spanning tree $p\colon \mathcal{T}_G \to [0,1]$ such that $\sum_{T\in\mathcal{T}} p_T = 1$. A simple implication of $\alpha(n) = O(\log n \log\log n)$ is the following.

Lemma 2.5.2. Let $G = (V, E)$ be a graph with non-negative edge lengths $\ell\colon E \to \mathbb{R}_+$. Let $w\colon E \to \mathbb{R}_+$ be any set of no-negative edge weights. Then there is a spanning tree $T \in \mathcal{T}_G$ such that

$$\frac{\sum_{uv\in E} w(uv) d_T(u,v)}{\sum_{e\in E} w(e)\ell(e)} \le \alpha(n),$$

where $d_T(u,v)$ is the length of the unique path from u to v in T.

Proof. Recall that there is a probability distribution $p\colon \mathcal{T}_G \to [0,1]$ such that for every pair of vertices $u, v \in V$, $\mathbb{E}[d_T(u,v)] \leq \alpha(n) d_G(u,v)$ where distances are induced by the edge lengths $\ell\colon E \to \mathbb{R}_+$. Now, fix any weight function w over pairs of vertices. Then by linearity of expectation,

$$\mathbb{E}\left[\sum_{u,v\in V} w((u,v)) d_T(u,v)\right] \leq \alpha(n) \sum_{u,v\in V} w((u,v)) d_G(u,v).$$

Thus, there exists a tree T such that $\sum_{u,v\in V} w((u,v)) d_T(u,v) \leq \alpha(n) \sum_{u,v\in V} w((u,v)) d_G(u,v)$. Now, consider the case when the support of the weights w is only on the edges of G, i.e., $w((u,v)) = 0$ when $uv \notin E$. In this case, we write $w(uv)$ for $uv \in E$. Furthermore, we let $d_T(u,v)$ is the shortest path length along the path in T. Since $d_G(u,v) \leq \ell(u,v)$,

$$\sum_{uv\in E} w(uv) d_T(u,v) \leq \alpha(n) \sum_{uv\in E} w(uv) d_G(u,v) \leq \alpha(n) \sum_{uv\in E} w(uv)\ell(uv),$$

which gives the desired result. ■

We now introduce a specific type of oblivious routing. Consider any probability distribution p over the spanning trees of G. Observe that this distribution induces an oblivious routing since each tree $T \in \mathcal{T}_G$ gives a unique path between any pair $(u,v) \in V \times V$. We sample a tree from the distribution and whenever a demand arrives, we simply route it along the unique path in the tree.

Notation (Tree-based oblivious routing)**.** The above scheme is what we called *tree-based oblivious routing*, which is indeed path-based.

Note. The distribution over $\mathcal{T}_G$ induces a distribution for every pair of vertices simultaneously.

While this is a restricted class of oblivious routing, but it's particularly nice, at least form a theoretical point of view. Two natural questions arise.

Problem. How good is the best tree-distributions-based oblivious routing?

Problem. Can the best tree-distributions-based oblivious routing be efficiently computed?

Räcke showed that tree-based oblivious routing have a nice structural property that allows one to characterize the congestion in a simple way [Räc08]. Let T be a spanning tree and consider an edge $e \in E_T$. $T - e$ induces a partition of the vertex set of $G = (V, E)$ into two sets $(S_e, V \setminus S_e)$. We define the load $L(T,e)$ on edge e to be $c(\delta(S_e))$. If e is not in T, then we let $L(T,e) = 0$.

Intuition. Think of all the edges crossing the cut $(S_e, V \setminus S_e)$. For each of those edges $e' = (s,t) \in \delta(S_e)$, the path from s to t in T has to go via e.

Hence, if we want to route demands corresponding to edges, then the congestion on e will be $L(T,e)/c(e)$. Formally, since we now have a probability distribution over $\mathcal{T}_G$, hence we consider the expected load and expected congestion:

Definition. Let $p\colon \mathcal{T}_G \to [0,1]$ be a probability distribution that induces a tree-based oblivious routing. Consider a given edge $e \in E$.

Definition 2.5.7 (Expected load)**.** The *expected load* on e is $L(e) = \sum_{T\in\mathcal{T}_G} p(T) L(T,e)$.

Definition 2.5.8 (Expected congestion)**.** The *expected congestion* on e is $\rho(e) = L(e)/c(e)$.

A simple yet important observation is the following, which characterizes the quality of the tree-based oblivious routing based on the expected congestion of the edges.

Lemma 2.5.3. Given $p\colon \mathcal{T} \to [0,1]$, the maximum congestion of the tree-based oblivious routing induced by p is at most $\max_{e\in E} \rho(e)$.

Proof. Fix a demand matrix $D \in \mathcal{D}_G$. Congestion of e is less than $\rho(e)$ since

$$\frac{\sum_{T\in\mathcal{T}_G, T\ni e} p_T \sum_{|S_e\cap\{u,v\}|=1} D(u,v)}{c(e)} \le \frac{\sum_{T\in\mathcal{T}_G, T\ni e} p_T L(T,e)}{c(e)}.$$

Taking maximum over $e \in E$ gives the result. ■

Lemma 2.5.3 allows us to write an linear program to find the best tree-based oblivious routing. Let x_T to denote the probability that $T \in \mathcal{T}_G$ is chosen in the probability distribution, and ρ to denote the congestion that we wish to minimize. We write down constraints that express x as a probability distribution and to express the load on each edge being bounded by $\rho c(e)$:

$$\begin{aligned}
&\min \rho && && \max \beta \\
&\sum_{T\in\mathcal{T}_G} x_T = 1 && && \sum_{e\in E} c(e) z_e = 1 \\
&\sum_{T\in\mathcal{T}_G} x_T L(T,e) \le \rho c(e) && \forall e \in E; && \sum_{e\in T} L(T,e) z_e \ge \beta && \forall T \in \mathcal{T}_G; \\
\text{(P)}\quad & x_T \ge 0 && \forall T \in \mathcal{T}_T; && \text{(D)}\quad z_e \ge 0 && \forall e \in E.
\end{aligned}$$

We see that the dual is equivalent to

$$\max_{z\colon E\to\mathbb{R}_+} \min_{T\in\mathcal{T}_G} \frac{\sum_{e\in T} L(T,e) z_e}{\sum_{e\in E} c(e) z_e}. \tag{2.6}$$

The observation is that Lemma 2.5.2 implies that the optimal dual value is $\alpha(n)$, which corresponds to the bound for tree-based distance approximation! Suppose this was true then we have shown that there exists a tree-based oblivious routing with congestion $O(\log n \log\log n)$. We now prove this formally.

Theorem 2.5.1. The optimal dual value β is at most $\alpha(n)$.

Proof. Observe that by the definition of $L(T,e)$, after interchanging the order of summation,

$$\sum_{e\in T} L(T,e) z_e = \sum_{uv\in E} c(uv) \sum_{e\in P_T(u,v)} z_e,$$

where $P_T(u,v)$ is the unique path from u to v in T. Note that $\sum_{e\in P_T(u,v)} z_e$ can be thought of as the length of the path from u to v in T according to lengths given by z, which we denote as $d_T(u,v)$. Now, think of $c(e)$ as a weight $w(e)$ and think of z_e as length $\ell(e)$, Equation 2.6 can be written as

$$\max_{\ell\in\mathbb{R}_+^m} \min_{T\in\mathcal{T}_G} \frac{\sum_{uv\in E} w(uv) d_T(u,v)}{\sum_{uv\in E} w(uv)\ell(uv)},$$

and by Lemma 2.5.2, this is at most $\alpha(n)$. ■

Theorem 2.5.1 implies that we have $\alpha(n) = O(\log n \log\log n)$ expected congestion for tree-based oblivious routing.[2] Since Theorem 2.5.1 is based on duality, it is not immediately clear that it leads to an efficient algorithm. As one would expect, we need an efficient algorithm for the dual separation oracle. This is the problem of approximating distances in the graph by spanning trees, and we have seen efficient algorithms for it. However, it is still not obvious that we can use such approximation algorithms in the dual, but there are standard techniques via the multiplicative weight update method and related ideas [Räc08].

[2] See note for how to obtain the optimum bound $O(\log n)$.

Appendix

Bibliography

[ABN08] Ittai Abraham, Yair Bartal, and Ofer Neiman. "Nearly tight low stretch spanning trees". In: *2008 49th Annual IEEE Symposium on Foundations of Computer Science*. IEEE. 2008, pp. 781–790.

[AKT21] Amir Abboud, Robert Krauthgamer, and Ohad Trabelsi. "Subcubic algorithms for Gomory–Hu tree in unweighted graphs". In: *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*. 2021, pp. 1725–1737.

[Ale+17] Vedat Levi Alev et al. *Graph Clustering using Effective Resistance*. 2017.

[ALN05] Sanjeev Arora, James R Lee, and Assaf Naor. "Euclidean distortion and the sparsest cut". In: *Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*. 2005, pp. 553–562.

[ALN07] Sanjeev Arora, James R Lee, and Assaf Naor. "Fréchet embeddings of negative type metrics". In: *Discrete & Computational Geometry* 38.4 (2007), pp. 726–739.

[Alo+95] Noga Alon et al. "A graph-theoretic game and its application to the k-server problem". In: *SIAM Journal on Computing* 24.1 (1995), pp. 78–100.

[AR98] Yonatan Aumann and Yuval Rabani. "An $O(logk)$ approximate min-cut max-flow theorem and approximation algorithm". In: *SIAM Journal on Computing* 27.1 (1998), pp. 291–301.

[ARV09] Sanjeev Arora, Satish Rao, and Umesh Vazirani. "Expander flows, geometric embeddings and graph partitioning". In: *Journal of the ACM (JACM)* 56.2 (2009), pp. 1–37.

[Aza+03] Yossi Azar et al. "Optimal oblivious routing in polynomial time". In: *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*. 2003, pp. 383–388.

[Bar96] Yair Bartal. "Probabilistic approximation of metric spaces and its algorithmic applications". In: *Proceedings of 37th Conference on Foundations of Computer Science*. IEEE. 1996, pp. 184–193.

[Bar98] Yair Bartal. "On approximating arbitrary metrices by tree metrics". In: *Proceedings of the thirtieth annual ACM symposium on Theory of computing*. 1998, pp. 161–168.

[BKR03] Marcin Bienkowski, Miroslaw Korzeniowski, and Harald Räcke. "A practical algorithm for constructing oblivious routing schemes". In: *Proceedings of the fifteenth annual ACM symposium on Parallel algorithms and architectures*. 2003, pp. 24–33.

[Cha00] Bernard Chazelle. "A minimum spanning tree algorithm with inverse-Ackermann type complexity". In: *Journal of the ACM (JACM)* 47.6 (2000), pp. 1028–1047.

[Chu+20] Julia Chuzhoy et al. "A deterministic algorithm for balanced cut with applications to dynamic connectivity, flows, and beyond". In: *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE. 2020, pp. 1158–1167.

[CKR05] Gruia Calinescu, Howard Karloff, and Yuval Rabani. "Approximation algorithms for the 0-extension problem". In: *SIAM Journal on Computing* 34.2 (2005), pp. 358–372.

[CQ17] Chandra Chekuri and Kent Quanrud. "Near-linear time approximation schemes for some implicit fractional packing problems". In: *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM. 2017, pp. 801–820.

[CQ21] Chandra Chekuri and Kent Quanrud. "Isolating Cuts, (Bi-)Submodularity, and Faster Algorithms for Global Connectivity Problems". In: *CoRR* abs/2103.12908 (2021). arXiv: 2103.12908. URL: https://arxiv.org/abs/2103.12908.

[CQX20] Chandra Chekuri, Kent Quanrud, and Chao Xu. "LP relaxation and tree packing for minimum k-cut". In: *SIAM Journal on Discrete Mathematics* 34.2 (2020), pp. 1334–1353.

[DKL76] Efim A Dinitz, Alexander V Karzanov, and Michael V Lomonosov. "On the structure of the system of minimum edge cuts of a graph". In: *Issledovaniya po Diskretnoi Optimizatsii* (1976), pp. 290–306.

[DRT92] Brandon Dixon, Monika Rauch, and Robert E Tarjan. "Verification and sensitivity analysis of minimum spanning trees in linear time". In: *SIAM Journal on Computing* 21.6 (1992), pp. 1184–1192.

[Eis97] Jason Eisner. "State-of-the-art algorithms for minimum spanning trees". In: *Unpublished survey* (1997).

[Elk+05] Michael Elkin et al. "Lower-stretch spanning trees". In: *Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*. 2005, pp. 494–503.

[Ene+16] Alina Ene et al. "Routing under balance". In: *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*. 2016, pp. 598–611.

[Fil24] Arnold Filtser. "On sparse covers of minor free graphs, low dimensional metric embeddings, and other applications". In: *arXiv preprint arXiv:2401.14060* (2024).

[FRT03] Jittat Fakcharoenphol, Satish Rao, and Kunal Talwar. "A tight bound on approximating arbitrary metrics by tree metrics". In: *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*. 2003, pp. 448–455.

[FT87] Michael L Fredman and Robert Endre Tarjan. "Fibonacci heaps and their uses in improved network optimization algorithms". In: *Journal of the ACM (JACM)* 34.3 (1987), pp. 596–615.

[Gab+86] Harold N Gabow et al. "Efficient algorithms for finding minimum spanning trees in undirected and directed graphs". In: *Combinatorica* 6.2 (1986), pp. 109–122.

[Gor+21] Gramoz Goranci et al. "The expander hierarchy and its applications to dynamic graph algorithms". In: *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*. SIAM. 2021, pp. 2212–2228.

[Gup+04] Anupam Gupta et al. "Cuts, trees and ℓ_1-embeddings of graphs". In: *Combinatorica* 24.2 (2004), pp. 233–269.

[GVY93] Naveen Garg, Vijay V Vazirani, and Mihalis Yannakakis. "Approximate max-flow min-(multi) cut theorems and their applications". In: *Proceedings of the twenty-fifth annual ACM symposium on Theory of computing*. 1993, pp. 698–707.

[Haj+07] Mohammad Taghi Hajiaghayi et al. "Oblivious routing on node-capacitated and directed graphs". In: *ACM Transactions on Algorithms (TALG)* 3.4 (2007), 51–es.

[HHR03] Chris Harrelson, Kirsten Hildrum, and Satish Rao. "A polynomial-time tree decomposition to minimize congestion". In: *Proceedings of the fifteenth annual ACM symposium on Parallel algorithms and architectures*. 2003, pp. 34–43.

[HLW06] Shlomo Hoory, Nathan Linial, and Avi Wigderson. "Expander graphs and their applications". In: *Bulletin of the American Mathematical Society* 43.4 (2006), pp. 439–561.

[HO94] JX Hao and James B Orlin. "A faster algorithm for finding the minimum cut in a directed graph". In: *Journal of Algorithms* 17.3 (1994), pp. 424–446.

[Kar00] David R Karger. "Minimum cuts in near-linear time". In: *Journal of the ACM (JACM)* 47.1 (2000), pp. 46–76.

[Kar95] David Ron Karger. *Random sampling in graph optimization problems*. stanford university, 1995.

[Kar98] David R Karger. "Random sampling and greedy sparsification for matroid optimization problems". In: *Mathematical Programming* 82.1 (1998), pp. 41–81.

[Kin97] Valerie King. "A simpler minimum spanning tree verification algorithm". In: *Algorithmica* 18 (1997), pp. 263–270.

[KKT95] David R Karger, Philip N Klein, and Robert E Tarjan. "A randomized linear-time algorithm to find minimum spanning trees". In: *Journal of the ACM (JACM)* 42.2 (1995), pp. 321–328.

[Kom85] János Komlós. "Linear verification for spanning trees". In: *Combinatorica* 5.1 (1985), pp. 57–65.

[KPR93] Philip Klein, Serge A Plotkin, and Satish Rao. "Excluded minors, network decomposition, and multicommodity flow". In: *Proceedings of the twenty-fifth annual ACM symposium on Theory of computing*. 1993, pp. 682–690.

[KRV09] Rohit Khandekar, Satish Rao, and Umesh Vazirani. "Graph partitioning using single commodity flows". In: *Journal of the ACM (JACM)* 56.4 (2009), pp. 1–15.

[KS96] David R Karger and Clifford Stein. "A new approach to the minimum cut problem". In: *Journal of the ACM (JACM)* 43.4 (1996), pp. 601–640.

[LLR95] Nathan Linial, Eran London, and Yuri Rabinovich. "The geometry of graphs and some of its algorithmic applications". In: *Combinatorica* 15 (1995), pp. 215–245.

[LP20] Jason Li and Debmalya Panigrahi. "Deterministic min-cut in poly-logarithmic max-flows". In: *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE. 2020, pp. 85–92.

[LR99] Tom Leighton and Satish Rao. "Multicommodity max-flow min-cut theorems and their use in designing approximation algorithms". In: *Journal of the ACM (JACM)* 46.6 (1999), pp. 787–832.

[Mar08] Martin Mareš. "The saga of minimum spanning trees". In: *Computer Science Review* 2.3 (2008), pp. 165–221.

[NI92] Hiroshi Nagamochi and Toshihide Ibaraki. "Computing edge-connectivity in multigraphs and capacitated graphs". In: *SIAM Journal on Discrete Mathematics* 5.1 (1992), pp. 54–66.

[Ore+08] Lorenzo Orecchia et al. "On partitioning graphs via single commodity flows". In: *Proceedings of the fortieth annual ACM symposium on Theory of computing*. 2008, pp. 461–470.

[PR02] Seth Pettie and Vijaya Ramachandran. "An optimal minimum spanning tree algorithm". In: *Journal of the ACM (JACM)* 49.1 (2002), pp. 16–34.

[Räc02] Harald Räcke. "Minimizing congestion in general networks". In: *The 43rd Annual IEEE Symposium on Foundations of Computer Science, 2002. Proceedings.* IEEE. 2002, pp. 43–52.

[Räc08] Harald Räcke. "Optimal hierarchical decompositions for congestion minimization in networks". In: *Proceedings of the fortieth annual ACM symposium on Theory of computing*. 2008, pp. 255–264.

[Sch+03] Alexander Schrijver et al. *Combinatorial optimization: polyhedra and efficiency*. Vol. 24. 2. Springer, 2003.

[SL21] Thatchaphol Saranurak and Jason Li. "Deterministic Weighted Expander Decomposition in Almost-linear Time". In: *CoRR* abs/2106.01567 (2021).

[SW19] Thatchaphol Saranurak and Di Wang. "Expander decomposition and pruning: Faster, stronger, and simpler". In: *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM. 2019, pp. 2616–2635.

[Vaz01] Vijay V Vazirani. *Approximation Algorithms*. 2001.

[VB81] Leslie G Valiant and Gordon J Brebner. "Universal schemes for parallel communication". In: *Proceedings of the thirteenth annual ACM symposium on Theory of computing*. 1981, pp. 263–277.

[WS11] David P Williamson and David B Shmoys. *The design of approximation algorithms*. Cambridge university press, 2011.

[Yao75] Andrew Chi-Chih Yao. "An $O(|E| \log \log |V|)$ algorithm for finding minimum spanning trees". In: *Information Processing Letters* 4 (1975), pp. 21–23.