Today

1) Metrics
2) Low diameter decompositions (LDDs)
3) The Multicut Problem (solved using (1)+(2))

Recall

$$\sum_{i=1}^{n} \frac{1}{i} \leq 2 \cdot \ln n$$

Metrics

A mathematical formalism of "distance"

A **metric space** consists of a set of "points" V and function $f : V \times V \to \mathbb{R}_{\geq 0}$ satisfying

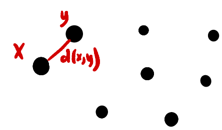
1) $f(x,x) = 0 \quad \forall x \in V$

2) $f(x,y) = f(y,x) \quad \forall x,y \in V$ (symmetric)

3) $f(x,y) \leq f(x,z) + f(z,y) \quad \forall x,y,z \in V$ (triangle inequality)

Examples

$(\mathbb{R}^n, d)$ where $d(x,y) := \|x-y\| = \sqrt{\sum_i (x_i - y_i)^2}$



↳ (1),(2) trivial, (3) from inequalities lecture

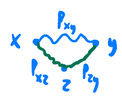
(V, d_G) where $G = (V, E, w)$ for $w: E \to \mathbb{R}$ is an edge-weighted graph

and $d_G(u,v)$ gives the shortest $u \rightsquigarrow v$ path length

↳ (1), (2) trivial, for (3) let P_{xy}, P_{xz}, P_{zy} be respective shortest paths

Then $P_{xz} \oplus P_{zy}$ is an $x \rightsquigarrow y$ path and since P_{xy} is shortest $x \rightsquigarrow y$ path have

$$d_G(x,y) = w(P_{xy}) \leq w(P_{xz}) + w(P_{zy}) \leq d_G(x,z) + d_G(z,y)$$



"distances" shouldn't be negative

Claim: If (V, f) is a metric space then $f(x,y) \geq 0 \quad \forall x,y \in V$

Have $\quad 0 = f(x,x) \underset{\text{triangle inequality}}{\leq} f(x,y) + f(y,x) \underset{\text{symmetry}}{=} f(x,y) + f(x,y) = 2 \cdot f(x,y)$

So $\quad 0 \leq f(x,y)$

The Metric Framework

i) Find a metric in your problem

ii) Find structure in your metric

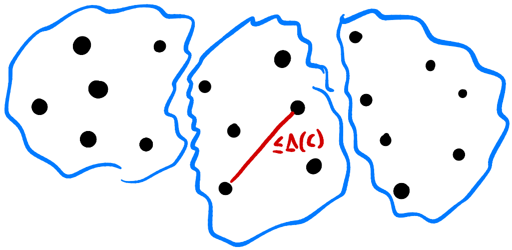
iii) Use metric structure to solve problem

Low Diameter Decompositions → (ii) today

A random clustering of metric points so nearby points probably in same cluster

A **clustering** of V is a partition of V into $C = \{V_i\}_i$

The **diameter** of clustering C is $\Delta(C) = \max_i \max_{u,v \in V_i} d(u,v)$

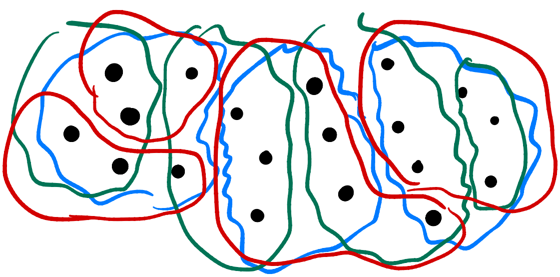


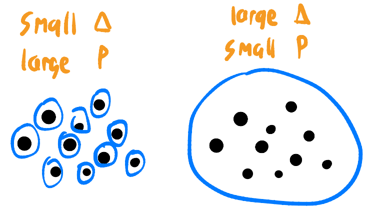
Say $u, v \in V$ **separated** by C if $u \in V_i$ and $v \in V_j$ for $i \neq j$

A **low diameter decomposition** (LDD) w/ diameter Δ and separation probability P

is a distribution $\mathcal{G}$ over Δ-diameter clustering s.t.

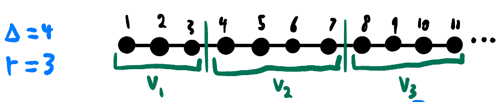
$$\Pr_{C \sim \mathcal{G}}(u,v \text{ separated by } C) \leq P \quad \forall u,v \in V$$





A tradeoff between P and Δ

Example: Suppose G a path w/ all length 1 edges; consider (V, d_G); given $\Delta \geq 0$

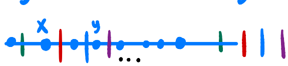


← shift Δ length intervals randomly

Let $r \sim U[0, \Delta + 1]$ $\quad V_1 = \{1, 2, \ldots, r\}$ and $V_i = [(i-2)\Delta + r + 1, (i-1)\Delta + r + 1) \cap \mathbb{Z}$ for $i \geq 2$

$\{V_i\}_i$ has diameter $\leq \Delta$ by construction

Consider $x, y \in V$ w/ $x \leq y$ → There are $\leq y - x = d_G(x,y)$ values of r s.t. x, y separated b/c



So $\forall \Delta$, $\exists$ LDD for path w/ $P \leq \frac{1}{\Delta}$

Theorem: Given any n-Point Metric Space (V,d), $\Delta \geq 0$ $\exists$ a (Poly-time computable) LDD s.t.

Separation Probability → $P \leq 4 \ln n \cdot d(u,v)$ $\forall u,v \in V$

Proof of Theorem

$\Delta/4$??

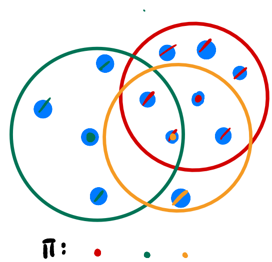
Alg. to Compute LDD

Let $r \sim U[0, \Delta/2]$

Let $\pi : [n] \to V$ be a Uniformly random Permutation on V

For $i = 1, 2, \ldots, n$

Let $V_i := B(\pi(i), r) \setminus \bigcup_{j<i} V_j$



Poly-time trivial

triangle inequality

$\{V_i\}_i$ has diameter $\leq \Delta$ b/c $x, y \in V_i \to d(x,y) \leq d(x, \pi(i)) + d(\pi(i), y)$

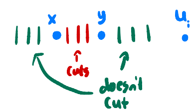
$\leq \frac{\Delta}{2} + \frac{\Delta}{2}$ $\left(x, y \in B(\pi(i), \frac{\Delta}{2})\right)$

Fix pair $t = \{u,v\}$ for $u,v \in V$

Let $d(u,t) := \min(d(u,x), d(u,y))$ and let $u_1, u_2, \ldots$ be V ordered by $d(\cdot, t)$

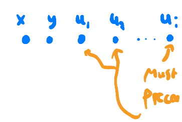
Say u_i <u>cuts</u> t iff $|t \cap B(u_i, r)| = 1$

$\to u_i$ only cuts t if $r \in [d(u_i,t), d(u_i,t) + d(x,y)]$



$\to$ so $\Pr(u_i \text{ cuts } t) \leq 2 \cdot \frac{d(x,y)}{\Delta}$

Say u_i <u>settles</u> t iff i is the min i s.t. $t \cap B(u_i, r) \neq \emptyset$



$\to u_i$ only settles t if u_i precedes $u_{i-1}, u_{i-2}, \ldots, u_1$ in π

$\to$ so $\Pr(u_i \text{ settles } t) \leq \frac{1}{i}$

$\to$ Also $\Pr(u_i \text{ settles } t) = \Pr(u_i \text{ settles } t \mid u_i \text{ cuts } t)$ ← Since cutting only about r

t is separated by $\{V_i\}_i$ iff $\exists\, i$ s.t. u_i settles + cuts t

So $\Pr(t \text{ separated}) = \sum_i \Pr(u_i \text{ settles + cuts } t)$

$= \sum_i \Pr(u_i \text{ cuts } t) \cdot \Pr(u_i \text{ settles } t \mid u_i \text{ cuts } t)$

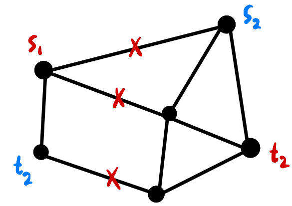
$\leq 2 \cdot \frac{d(x,y)}{\Delta} \cdot \sum_i \frac{1}{i}$

$\leq 4 \cdot \ln n \cdot \frac{d(x,y)}{\Delta}$ $\left(\sum_i \frac{1}{i} \leq 2 \cdot \ln n\right)$

Multicut Problem

Given $G=(V,E)$ and vertex pairs $(s_i,t_i)_i$, find minimum size $F \subseteq E$ s.t. s_i not connected to t_i in $(V, E \setminus F)$ $\forall i$

Generalizes s-t mincut



Fact: Multicut is NP-hard

Theorem: $\exists$ $\text{Poly}(n,m)$ -time $O(\log n)$-approximation for Multicut
($|V|$, $|E|$)

Alg.

LP Relaxation

Variable x_e $\forall e \in E$ so $x \in \mathbb{R}^m$

$$\min \sum_e x_e \quad \text{s.t.}$$

$$\sum_{e \in P} x_e \geq 1 \quad \forall i \; \forall \; s_i \rightsquigarrow t_i \text{ paths } P$$

$$x \geq 0$$

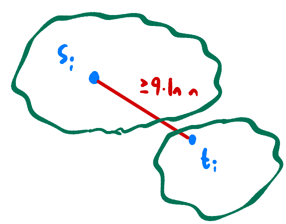
Can solve LP in $\text{Poly}(n,m)$ time via ellipsoid + separation oracle (see hw.)

Let x be an optimal LP solution

Let d_G be distances in G using $9 \cdot \ln n \cdot x$ as edge lengths ← (i)

Let $C \sim G$ be a $8 \cdot \ln n$ diameter LDD w/ $p \leq \frac{1}{2}$ (by previous LDD theorem) (ii)

Return $F := \{e \in E : e \text{ separated by } C\}$ ← (iii)



Runtime trivial

F is feasible; Consider a (s_i,t_i)

Have $d_G(s_i,t_i) \geq 9 \cdot \ln n$ ← Since LP forces every path have length ≥ 1 according to x^{th} to

But C has diameter $8 \cdot \ln n$ so s_i, t_i never in same cluster of C

So s_i, t_i not connected after deleting F

$\mathbb{E}[|F|] \leq O(\log n) \cdot OPT$ ← optimal multicut value

$$\mathbb{E}[|F|] = \sum_e \Pr(e \text{ separated by } C)$$

$$= \sum_e 9 \cdot \ln n \cdot x_e$$

$$= O(\log n) \cdot OPT_{LP} \leq O(\log n) \cdot OPT$$

(multicut LP OPT)