Today

1) Metrics
2) The Multicut Problem and approximation algorithms
3) Low diameter decompositions (LDDs)
4) Solving (2) w/ (1)+(3)

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

"metric" for short

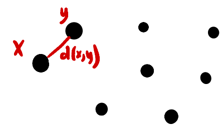
1) $f(x,x) = 0 \quad \forall x \in V$ → really $f(x,y) = 0$ iff $x = y$ (o/w a "pseudo metric", will ignore)

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



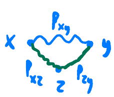
↳ (1),(2) trivial, (3) from inequalities lecture

(V, d) for any $V \subseteq \mathbb{R}^d$

↳ For same reason as above

(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, we have

$$d_G(x,y) = \underbrace{w(P_{xy})}_{w(P) := \sum_{e \in P} w(e)} \leq w(P_{xz}) + w(P_{zy}) \leq d_G(x,z) + d_G(z,y)$$

Claim: If (V, f) is a metric space then $f(x,y) \geq 0 \quad \forall x,y \in V$ ← "distances" shouldn't be negative

Have
$$0 \underset{(1)}{=} f(x,x) \underset{(3)}{\leq} f(x,y) + f(y,x) \underset{(2)}{=} f(x,y) + f(x,y) = 2 \cdot f(x,y)$$

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

The Metric Framework

→ will apply to multicut

i) Find a metric in your problem

ii) Find structure in your metric

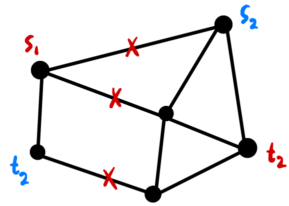
iii) Use metric structure to solve problem

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



Let $OPT := \min_{\text{feasible } F} |F|$

Fact: Multicut is NP-hard

Still possible: Find feasible $F \subseteq E$ s.t. $|F| \le \alpha \cdot OPT$ for small $\alpha > 1$ in P

↳ Given a family of minimization problems Π (e.g. all multicut instances) w/ objective w, randomized algorithm $\mathcal{A}: \Pi \to$ solutions is an α-approximation if $\mathbb{E}\left[w(\mathcal{A}(P))\right] \le \alpha \cdot OPT(P)$ $\forall P \in \Pi$

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

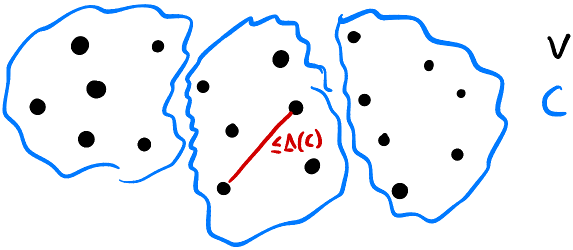
↳ Obvious metric is shortest path distances but we'll be more clever

Low Diameter Decompositions → (ii) today

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

A <u>clustering</u> of V is a partition of V into $C = \{V_1, V_2, V_3, \ldots\}$

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



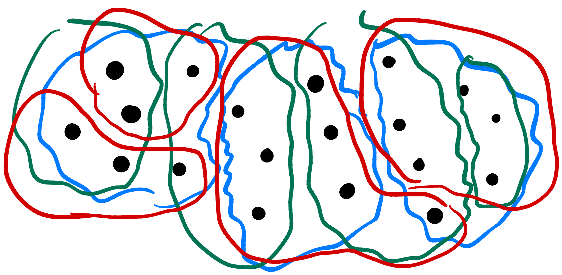
$u, v \in V$ are <u>separated</u> by C if $u \in V_i$ and $v \in V_j$ for $i \neq j$

A <u>low diameter decomposition</u> (LDD) w/ diameter Δ and quality q

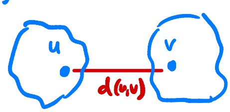
is a distribution $\mathcal{G}$ over Δ-diameter clustering s.t.

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

want small

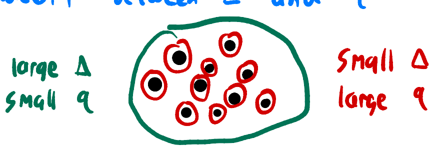


Intuition 1: why separation probability depends on distance



If $d(u,v) > \Delta$ then $\Pr(u,v \text{ separated}) = 1$

Intuition 2: tradeoff between Δ and q



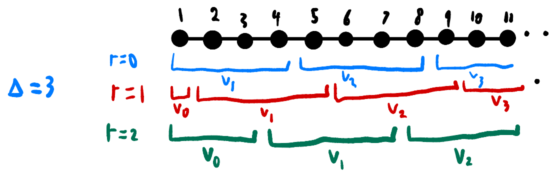
Claim: If $G=(V,E,w)$ and $\mathcal{C}$ is an LDD w/ quality q then $\forall \{u,v\}=e\in E$,

$$\Pr(u,v \text{ separated}) \le q\cdot w(e)$$

Have $d_G(u,v)\le w(e)$ so $\Pr(u,v \text{ separated}) \le q\cdot d_G(u,v) \le q\cdot w(e)$

Claim: If $G=(V,E)$ is a (unit-length) path then $\forall \Delta>0$, (V,d_G) has a (poly-time computable) LDD w/

$$q\le \frac{1}{\Delta}$$



shift ≈ Δ length intervals randomly

Let $r\sim U[0,\Delta-1]\cap\mathbb{Z}$ $V_0=\{1,2,\dots,r\}$ and $V_i=((i-1)\cdot\Delta+r,\ i\Delta+r]\cap\mathbb{Z}$ for $i\ge 1$

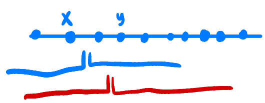
$\{V_0,V_1,\dots\}$ has diameter $\le \Delta-1\le\Delta$ by construction

For a fixed edge $\{u,v\}\in E$, there is exactly 1 value of r that separates u,v → See above picture

Consider $x,y\in V$ w/ $x\le y$

There are $\le y-x=d_G(x,y)$ values of r s.t. x,y separated b/c

E.g.



Since r chosen uniformly among Δ values

$$\Pr(x,y \text{ separated}) \le \frac{d(x,y)}{\Delta}$$

So $q\le \frac{1}{\Delta}$

Theorem: Given any n-point metric space (V,d), $\Delta \geq 0$ $\exists$ a (poly-time computable) LDD s.t. $\forall u,v \in V$

Separation Probability → $q \leq \frac{4 \ln n}{\Delta}$

Proof of Theorem

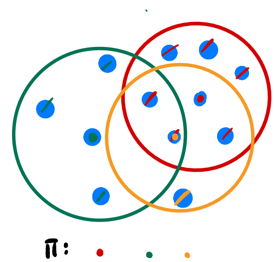
Alg. to Compute LDD

Let $r \sim U[0, \Delta/2]$ $\Delta/4$?!

Let $\pi:[n]\to V$ be a uniformly random permutation on V

For $i = 1, 2, \dots, n$ $B(u,r) := \{v : d(v,u) \leq r\}$

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



Poly-time trivial

$\{V_1, V_2, \dots\}$ has diameter $\leq \Delta$ b/c $x,y \in V_i \to d(x,y) \leq d(x,\pi(i)) + d(\pi(i),y)$ triangle inequality

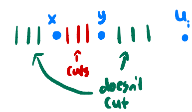
$\leq \frac{\Delta}{2} + \frac{\Delta}{2}$ $(x,y \in B(\pi(i), \frac{\Delta}{2}))$

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

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

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

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

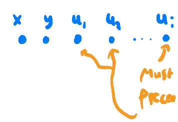


→ so $\Pr(u_i \text{ cuts } t) \leq 2 \cdot \frac{d(x,y)}{\Delta}$

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

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

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



i.e. endpoints separated

t is separated iff $\exists\, i$ s.t. u_i settles + cuts t

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

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

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

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

Theorem: $\exists$ $\mathrm{poly}(\overset{|V|}{n}, \overset{|E|}{m})$-time randomized $O(\log n)$-approximation for Multicut

Alg.

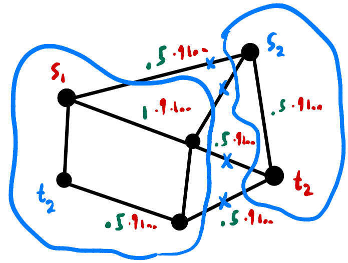
LP Relaxation → a "randomized rounding" alg.

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

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

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

$$x_e \in [0,1] \quad \forall e \in E$$



Let OPT_{LP} = optimal val. to above

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

Let x be an optimal LP solution

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

Let $C \sim \mathcal{G}$ be a $8 \cdot \ln n$ diameter LDD w/ $q = \frac{1}{2}$ so $\Pr(u, v \text{ separated}) \leq \frac{d(u,v)}{2}$ ← by prev. LDD thm. (ii)

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

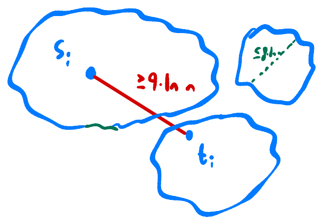
Runtime trivial

F is feasible

Consider a (s_i, t_i)

Have $d_G(s_i, t_i) \geq 9 \cdot \ln n$ b/c $\forall s_i \rightsquigarrow t_i$ paths P

$$\sum_{e \in P} x_e \geq 1$$
$$\Updownarrow$$
$$\sum_{e \in P} 9 \cdot \ln n \, x_e \geq 9 \cdot \ln n$$
$$\Updownarrow$$
$$\sum_{e \in P} w(e) \geq 9 \cdot \ln n$$
$$\Updownarrow$$
$$w(P) \geq 9 \cdot \ln n$$



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

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

$$\mathbb{E}[|F|] = \sum_e \Pr(e \text{ separated by } C)$$
$$\leq \sum_e w(e)/2 \quad \text{(by claim)}$$
$$= \sum_e 9 \cdot \ln n \cdot x_e / 2$$
$$= O(\log n) \cdot OPT_{LP} \leq O(\log n) \cdot OPT$$