Graph Sparsification

Note Title 4/20/2016

$G = (V, E)$ very big

↓

$\boxed{\text{smaller } G'}$ captures same info

$|E| = c \cdot |V|$ c: 20, 100

⟹ $|E| = |V|^{1+\beta}$ $\beta \in [0.1, 0.5]$

$|V| = 1$ million

$|V|^{1.1} = 4$ million $|V|^{1.5} = 1$ billion

"as they grow, on average, each node has more edges"

reduce V

fix V, reduce E

---

① general, mostly theory.

② planar (road network), practical.

$\left(\begin{array}{l}\text{MapReduce}\\ \text{Hadoop}\end{array}\right)$ → GraphLab → Giraph

In memory on one machine

$G = (V, E) \rightarrow G' = (V, E')$

$|E| = |V|^{1.5}$

$E' \subset E$

$|E'| = c \cdot |V| \cdot \log |V|$

$c \rightarrow \frac{1}{\varepsilon^2}$

$\log_2(1{,}000{,}000) \approx 20$

Edge Sampling

- sample each edge $e_{ij} = (v_i, v_j)$ into E' with prob. P_{ij}, independe

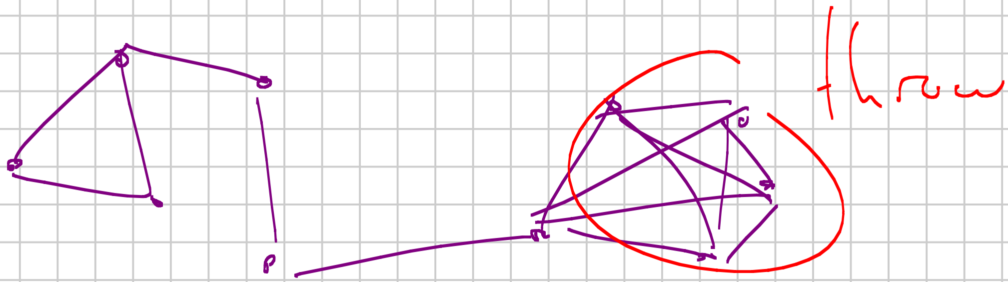
$$P_{ij} = \min\left\{1, \frac{t}{\min\{d_i, d_j\}}\right\}$$

degree V_j

$$\boxed{t = \frac{1}{\varepsilon^2}\log|V|}$$

$\text{Exp}\left[|E'|\right] \leq t \cdot |V|$

$d_i \cdot \frac{t}{d_i} = t$



$$w_{ij} = \begin{cases} \frac{1}{P_{ij}} & \text{if kept} \\ 0 & \text{otherwise} \end{cases}$$

Preserve cuts

G, $e \in E$ have weight $w_e = 1$
$e \in E'$ has weight w_e'

cut $C = (S, T)$
$S \cup T = V$
$S \cap T = \phi$

$\forall$ cut C

$$\left| \sum_{e \in C}' w_e - \sum_{e' \in C}' w_e' \right| < \varepsilon |V|$$

$$t = \overbrace{\frac{1}{\varepsilon^2}}^{10{,}000} \overbrace{\log|V|}^{20}$$

$\geq \frac{1}{\varepsilon^2}$

- spectral clustering
- communities

Laplacian

$L_G = D_G - A_G$

$\begin{bmatrix} \cancel{d_i} & \\ 0 & \end{bmatrix} \begin{bmatrix} & a_{ij} \\ & \end{bmatrix}$

$\|L_G - L_{G'}\|_2 \le \varepsilon |V|$

$(1-\varepsilon)\, x^T L_G x \le x^T L_{G'} x \le (1+\varepsilon)\, x^T L_G x$

$(0,1,1,0,0,1) = \text{cut}$

Sampling

$P_{ij} = c \cdot R_{eff}(e_i)$

$t = \frac{1}{\varepsilon^2} \log |V|$

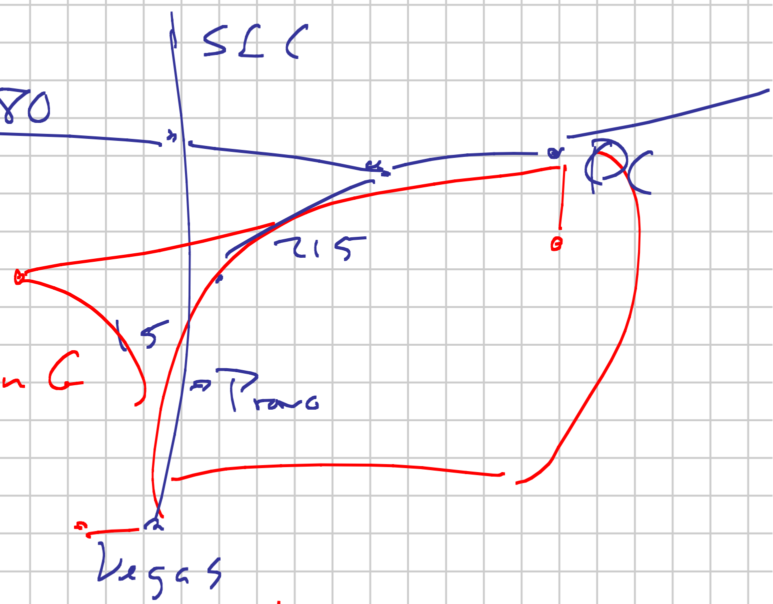
V live in a metric space

$e = (v_1, v_2)$ $\quad w_e \approx d(v_1, v_2)$

$V \subset \mathbb{R}^2$

$G \to G'$

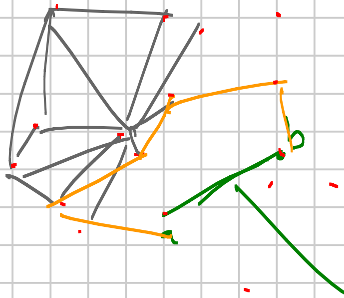
(V,E) $\quad (V,E')$



$d_G(v_1, v_2)$ = length shortest path $v_1 \to v_2$ on G

$$1 \le \frac{d_{G'}(a,b)}{d_G(a,b)} \le 1+t$$

t-spanner →
- small # edges
- small total weight
- small max degree



$V \subset \mathbb{R}^2$ $E = \{(v_1, v_2) \mid v_1, v_2 \in V\}$,
complete

$d(a,b) = d_G(a,b)$

t-spanner ?

• <u>Greedy</u> sort by edge weight = distance
small → large.

$d\,G'(a,b)$ if $d(a,b) < \frac{1}{(1+t)} d_{G'}(a,b)$
create then add $(a,b) \to G'$

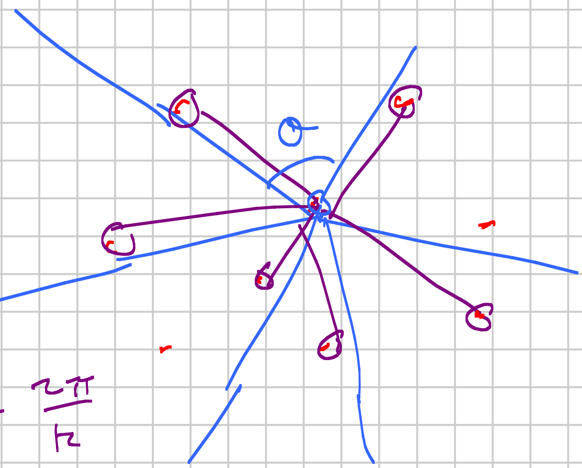
• <u>Cone-Based</u>

$\theta \le 60°$

max degree = 6
$= \frac{360}{\theta}$

t-spanner
$t = \frac{1}{1-\sin(\frac{\theta}{2})}$

$\theta = \frac{2\pi}{k}$



• <u>WSPD</u>

<u>Quadtree</u>

• Multipole