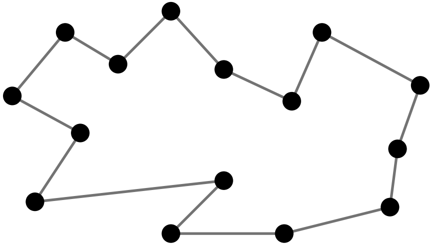
Random Metrics in the Analysis of Algorithms

Bodo Manthey

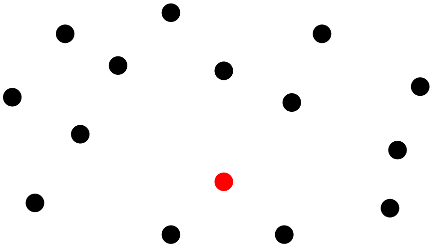
UNIVERSITY OF TWENTE.

February 7, 2023

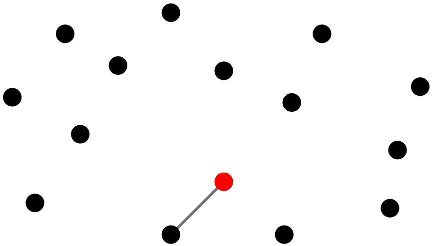
Heuristics



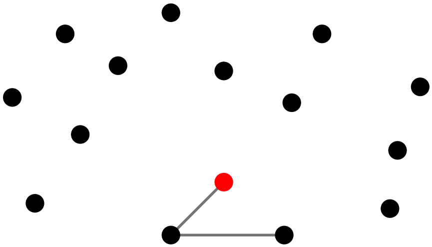
Heuristics



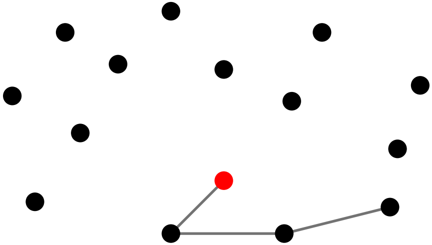
Heuristics



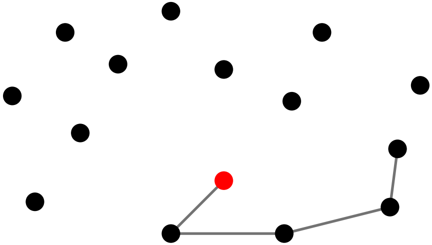
Heuristics



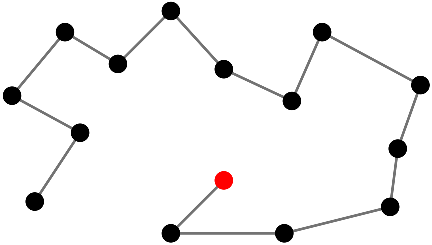
Heuristics



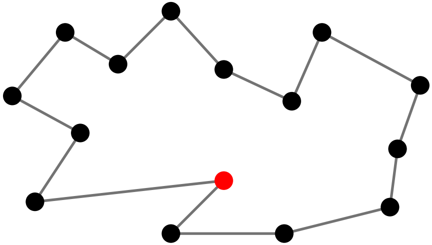
Heuristics



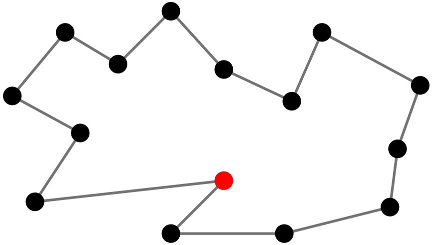
Heuristics



Heuristics



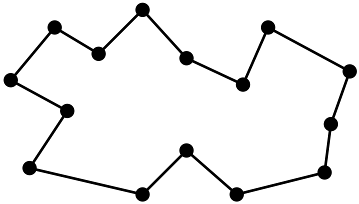
Heuristics

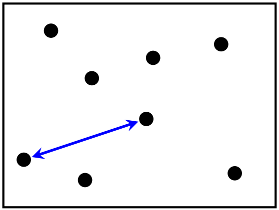


nearest neighbor for TSP:

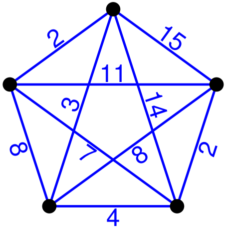
- simple construction heuristic
- worst-case approximation ratio (metric): $O(\log n)$
- experimental: ≈ 1.25
- average-case performance?

Why random metric spaces?





random in $[0, 1]^2$

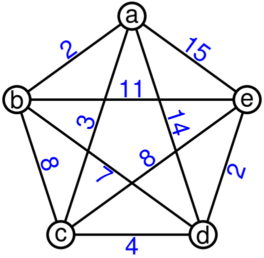


independent edge lengths

Random shortest paths = First-passage percolation

1. edge weights:
   exponentially distributed,
   independent

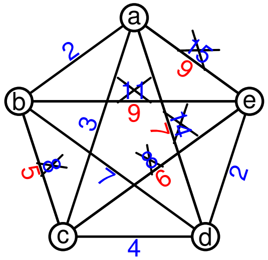
2. shortest paths w.r.t. weights



Random shortest paths = First-passage percolation

1. edge weights: exponentially distributed, independent

2. shortest paths w.r.t. weights



Outline

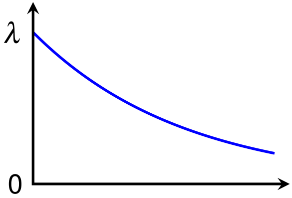
1. Properties of RSP
2. Heuristics for TSP
   - Nearest neighbor
   - Insertion heuristics
3. Facility location problem
4. General probability distributions
5. RSP with non-complete graphs
   - Random graphs
   - 2-opt on sparse graphs
6. Conclusions

Outline

Exponential distribution – Properties

Exp(λ)

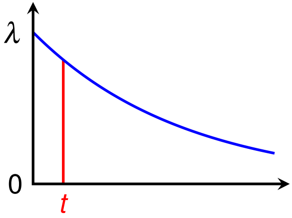
- density: $\lambda e^{-\lambda x}$ for $x \geq 0$
- CDF: $1 - e^{-\lambda x}$ for $x \geq 0$
- expected value: $1/\lambda$



Exponential distribution – Properties

Exp(λ)

- density: $\lambda e^{-\lambda x}$ for $x \geq 0$
- CDF: $1 - e^{-\lambda x}$ for $x \geq 0$
- expected value: $1/\lambda$



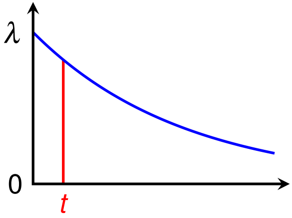
memorylessness ($R \sim \text{Exp}(\lambda)$)

$$\mathbb{P}(R \geq t + x \mid R \geq t) = \frac{e^{-\lambda(x+t)}}{e^{-\lambda t}} = \mathbb{P}(R \geq x)$$

Exponential distribution – Properties

$\text{Exp}(\lambda)$

- density: $\lambda e^{-\lambda x}$ for $x \geq 0$
- CDF: $1 - e^{-\lambda x}$ for $x \geq 0$
- expected value: $1/\lambda$



memorylessness ($R \sim \text{Exp}(\lambda)$)

$$\mathbb{P}(R \geq t + x \mid R \geq t) = \frac{e^{-\lambda(x+t)}}{e^{-\lambda t}} = \mathbb{P}(R \geq x)$$

minimum ($R_1, \ldots, R_k \sim \text{Exp}(\lambda)$; $M = \min\{R_1, \ldots, R_k\}$)

$$\mathbb{P}(M \geq x) = \prod_{i=1}^{k} \mathbb{P}(R_i \geq x) = (e^{-\lambda x})^k = e^{-(\lambda k)x}$$

$$\Rightarrow M \sim \text{Exp}(\lambda k)$$

Question

What is the typical distance in RSP?

What is $\mathbb{E}\big(d(u,v)\big)$?

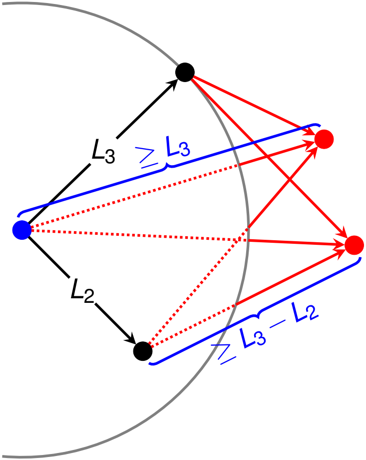
Question

What is the typical distance in RSP?

What is $\mathbb{E}\big(d(u,v)\big)$?

1. $\log n$
2. 1
3. $1/\sqrt{n}$
4. $\log n/n$
5. $1/n$

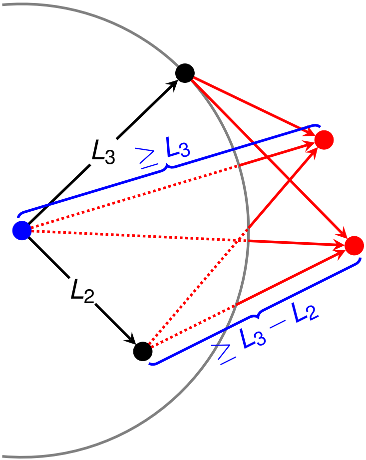
Distribution of RSP



$(H_n = \sum_{i=1}^{n} \frac{1}{i})$

- L_k = distance to k-th closest vertex
- $L_1 = 0$
- $L_2 = \text{Exp}(n-1)$

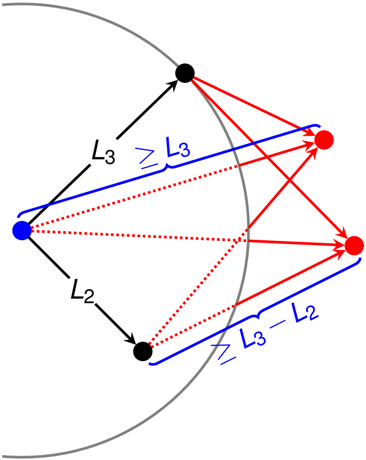
Distribution of RSP



$(H_n = \sum_{i=1}^{n} \frac{1}{i})$

- $L_k =$ distance to k-th closest vertex
- $L_1 = 0$
- $L_2 = \text{Exp}(n-1)$
- $L_{k+1} = L_k + \text{Exp}\big((n-k)\cdot k\big)$
- $L_k = \sum_{i=1}^{k-1} \text{Exp}\big((n-i)\cdot i\big)$

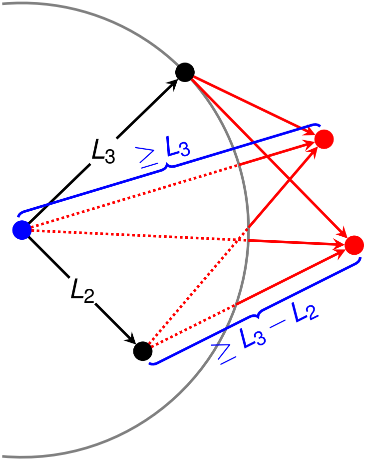
Distribution of RSP



$(H_n = \sum_{i=1}^{n} \frac{1}{i})$

- L_k = distance to k-th closest vertex
- $L_1 = 0$
- $L_2 = \mathrm{Exp}(n-1)$
- $L_{k+1} = L_k + \mathrm{Exp}\big((n-k)\cdot k\big)$
- $L_k = \sum_{i=1}^{k-1} \mathrm{Exp}\big((n-i)\cdot i\big)$
- $\mathbb{E}(L_k) = \sum_{i=1}^{k-1} \frac{1}{(n-i)\cdot i} = \sum_{i=1}^{k-1} \frac{1}{n}\cdot(\frac{1}{i} + \frac{1}{n-i})$

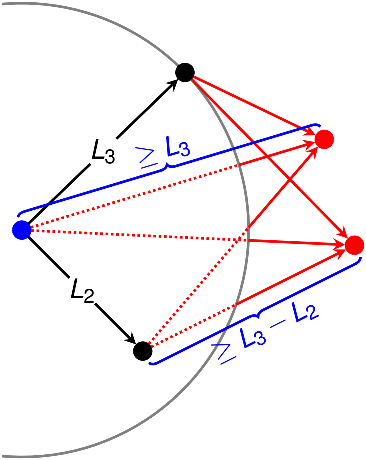
Distribution of RSP



$(H_n = \sum_{i=1}^{n} \frac{1}{i})$

- $L_k =$ distance to k-th closest vertex
- $L_1 = 0$
- $L_2 = \text{Exp}(n-1)$
- $L_{k+1} = L_k + \text{Exp}\big((n-k)\cdot k\big)$
- $L_k = \sum_{i=1}^{k-1} \text{Exp}\big((n-i)\cdot i\big)$
- $\mathbb{E}(L_k) = \sum_{i=1}^{k-1} \frac{1}{(n-i)\cdot i} = \sum_{i=1}^{k-1} \frac{1}{n}\cdot(\frac{1}{i} + \frac{1}{n-i})$
  $= \frac{1}{n}\cdot\big(H_{k-1} + H_{n-1} - H_{n-k}\big)$

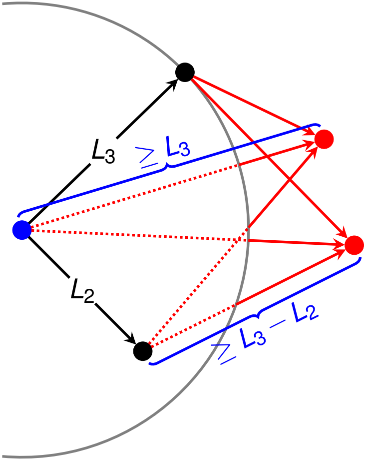
Distribution of RSP



$(H_n = \sum_{i=1}^{n} \frac{1}{i})$

- $L_k =$ distance to k-th closest vertex
- $L_1 = 0$
- $L_2 = \text{Exp}(n-1)$
- $L_{k+1} = L_k + \text{Exp}\big((n-k)\cdot k\big)$
- $L_k = \sum_{i=1}^{k-1} \text{Exp}\big((n-i)\cdot i\big)$
- $\mathbb{E}(L_k) = \sum_{i=1}^{k-1} \frac{1}{(n-i)\cdot i} = \sum_{i=1}^{k-1} \frac{1}{n}\cdot(\frac{1}{i} + \frac{1}{n-i})$
  $= \frac{1}{n}\cdot\big(H_{k-1} + H_{n-1} - H_{n-k}\big)$

Distribution of RSP



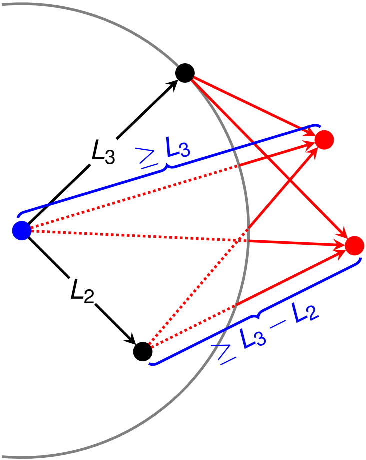
$(H_n = \sum_{i=1}^{n} \frac{1}{i})$

- $L_k =$ distance to k-th closest vertex
- $L_1 = 0$
- $L_2 = \mathrm{Exp}(n-1)$
- $L_{k+1} = L_k + \mathrm{Exp}\big((n-k)\cdot k\big)$
- $L_k = \sum_{i=1}^{k-1} \mathrm{Exp}\big((n-i)\cdot i\big)$
- $\mathbb{E}(L_k) = \sum_{i=1}^{k-1} \frac{1}{(n-i)\cdot i} = \sum_{i=1}^{k-1} \frac{1}{n}\cdot(\frac{1}{i} + \frac{1}{n-i})$
  $= \frac{1}{n}\cdot\big(H_{k-1} + H_{n-1} - H_{n-k}\big)$

Theorem (Janson 1999)

$$\mathbb{E}\big(\max_v d(u,v)\big) = \mathbb{E}\big(L_n\big) = 2\cdot\frac{H_{n-1}}{n} \approx 2\cdot\frac{\ln n}{n}$$

Distribution of RSP



$(H_n = \sum_{i=1}^{n} \frac{1}{i})$

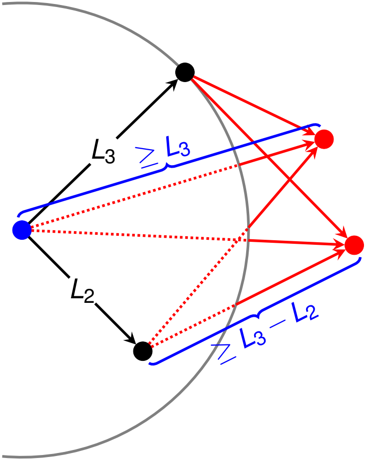
- $L_k =$ distance to k-th closest vertex
- $L_1 = 0$
- $L_2 = \text{Exp}(n-1)$
- $L_{k+1} = L_k + \text{Exp}\big((n-k)\cdot k\big)$
- $L_k = \sum_{i=1}^{k-1} \text{Exp}\big((n-i)\cdot i\big)$
- $\mathbb{E}(L_k) = \sum_{i=1}^{k-1} \frac{1}{(n-i)\cdot i} = \sum_{i=1}^{k-1} \frac{1}{n}\cdot(\frac{1}{i} + \frac{1}{n-i})$
  $= \frac{1}{n}\cdot\big(H_{k-1} + H_{n-1} - H_{n-k}\big)$

Theorem (Janson 1999)

$$\mathbb{E}\big(d(u,v)\big) = 1\cdot\frac{H_{n-1}}{n-1} \approx 1\cdot\frac{\ln n}{n}$$

$$\mathbb{E}\big(\max_v d(u,v)\big) = \mathbb{E}\big(L_n\big) = 2\cdot\frac{H_{n-1}}{n} \approx 2\cdot\frac{\ln n}{n}$$

Distribution of RSP



$(H_n = \sum_{i=1}^{n} \frac{1}{i})$

- $L_k =$ distance to k-th closest vertex
- $L_1 = 0$
- $L_2 = \text{Exp}(n-1)$
- $L_{k+1} = L_k + \text{Exp}\big((n-k)\cdot k\big)$
- $L_k = \sum_{i=1}^{k-1} \text{Exp}\big((n-i)\cdot i\big)$
- $\mathbb{E}(L_k) = \sum_{i=1}^{k-1} \frac{1}{(n-i)\cdot i} = \sum_{i=1}^{k-1} \frac{1}{n}\cdot(\frac{1}{i} + \frac{1}{n-i})$
  $= \frac{1}{n}\cdot\big(H_{k-1} + H_{n-1} - H_{n-k}\big)$

Theorem (Janson 1999)

$$\mathbb{E}\big(d(u,v)\big) = 1\cdot\frac{H_{n-1}}{n-1} \approx 1\cdot\frac{\ln n}{n}$$

$$\mathbb{E}\big(\max_v d(u,v)\big) = \mathbb{E}(L_n) = 2\cdot\frac{H_{n-1}}{n} \approx 2\cdot\frac{\ln n}{n}$$

$$\mathbb{E}\big(\max_{u,v} d(u,v)\big) \approx 3\cdot\frac{\ln n}{n}$$

Expected edge length

Theorem (Janson 1999)

$\mathbb{E}\big(d(u,v)\big) = \frac{H_{n-1}}{n-1}$

Proof.

previous slide: $\mathbb{E}(L_k) = \frac{1}{n} \cdot \big(H_{n-1} + H_{k-1} - H_{n-k}\big)$

$$\mathbb{E}\big(d(u,v)\big) = \frac{1}{n-1} \cdot \sum_{k=2}^{n} \mathbb{E}(L_k)$$

□

Expected edge length

Theorem (Janson 1999)

$\mathbb{E}\big(d(u,v)\big) = \frac{H_{n-1}}{n-1}$

Proof.

previous slide: $\mathbb{E}(L_k) = \frac{1}{n} \cdot \big(H_{n-1} + H_{k-1} - H_{n-k}\big)$

$$\begin{aligned}\mathbb{E}\big(d(u,v)\big) &= \frac{1}{n-1} \cdot \sum_{k=2}^{n} \mathbb{E}(L_k) \\ &= \frac{1}{n} \cdot H_{n-1} + \frac{1}{n \cdot (n-1)} \cdot \sum_{k=2}^{n} H_{k-1} - H_{n-k}\end{aligned}$$

□

Expected edge length

Theorem (Janson 1999)

$\mathbb{E}\big(d(u,v)\big) = \frac{H_{n-1}}{n-1}$

Proof.

previous slide: $\mathbb{E}(L_k) = \frac{1}{n} \cdot \big(H_{n-1} + H_{k-1} - H_{n-k}\big)$

$$\begin{aligned}\mathbb{E}\big(d(u,v)\big) &= \frac{1}{n-1} \cdot \sum_{k=2}^{n} \mathbb{E}(L_k) \\ &= \frac{1}{n} \cdot H_{n-1} + \frac{1}{n \cdot (n-1)} \cdot \sum_{k=2}^{n} H_{k-1} - H_{n-k}\end{aligned}$$

□

Expected edge length

Theorem (Janson 1999)

$\mathbb{E}\big(d(u,v)\big) = \frac{H_{n-1}}{n-1}$

Proof.

previous slide: $\mathbb{E}(L_k) = \frac{1}{n} \cdot \big(H_{n-1} + H_{k-1} - H_{n-k}\big)$

$$\begin{aligned}\mathbb{E}\big(d(u,v)\big) &= \frac{1}{n-1} \cdot \sum_{k=2}^{n} \mathbb{E}(L_k) \\ &= \frac{1}{n} \cdot H_{n-1} + \frac{1}{n \cdot (n-1)} \cdot \sum_{k=2}^{n} H_{k-1} - H_{n-k}\end{aligned}$$

□

Expected edge length

Theorem (Janson 1999)

$\mathbb{E}\big(d(u,v)\big) = \frac{H_{n-1}}{n-1}$

Proof.

previous slide: $\mathbb{E}(L_k) = \frac{1}{n} \cdot \big(H_{n-1} + H_{k-1} - H_{n-k}\big)$

$$\begin{aligned}
\mathbb{E}\big(d(u,v)\big) &= \frac{1}{n-1} \cdot \sum_{k=2}^{n} \mathbb{E}(L_k) \\
&= \frac{1}{n} \cdot H_{n-1} + \frac{1}{n\cdot(n-1)} \cdot \sum_{k=2}^{n} H_{k-1} - H_{n-k} \\
&= \frac{1}{n} \cdot H_{n-1} + \frac{1}{n\cdot(n-1)} \cdot H_{n-1} = \frac{1}{n-1} \cdot H_{n-1}
\end{aligned}$$

□

Expected edge length

Theorem (Janson 1999)

$\mathbb{E}\big(d(u,v)\big) = \frac{H_{n-1}}{n-1} \approx \frac{\ln n}{n}$

Proof.

previous slide: $\mathbb{E}(L_k) = \frac{1}{n} \cdot \big(H_{n-1} + H_{k-1} - H_{n-k}\big)$

$$\begin{aligned}
\mathbb{E}\big(d(u,v)\big) &= \frac{1}{n-1} \cdot \sum_{k=2}^{n} \mathbb{E}(L_k) \\
&= \frac{1}{n} \cdot H_{n-1} + \frac{1}{n \cdot (n-1)} \cdot \sum_{k=2}^{n} H_{k-1} - H_{n-k} \\
&= \frac{1}{n} \cdot H_{n-1} + \frac{1}{n \cdot (n-1)} \cdot H_{n-1} = \frac{1}{n-1} \cdot H_{n-1}
\end{aligned}$$

□

Sums of exponential random variables

Lemma

$X \sim \sum_{i=1}^{m} \mathrm{Exp}(\lambda i)$, then $\mathbb{P}(X \leq t) = (1 - e^{-\lambda t})^m$.

Proof.

- $Y_i \sim \mathrm{Exp}(\lambda)$ independently, order statistics $Y_{(1)} \leq Y_{(2)} \leq \ldots \leq Y_{(m)}$
- $Y_{(i)} - Y_{(i-1)} \sim \mathrm{Exp}(\lambda i)$ (memorylessness)
- X has same distribution as $\max\{Y_1, \ldots, Y_m\} = Y_{(m)}$

□

Sums of exponential random variables

Lemma

$X \sim \sum_{i=1}^{m} \mathrm{Exp}(\lambda i)$, then $\mathbb{P}(X \leq t) = (1 - e^{-\lambda t})^m$.

Proof.

- $Y_i \sim \mathrm{Exp}(\lambda)$ independently, order statistics $Y_{(1)} \leq Y_{(2)} \leq \ldots \leq Y_{(m)}$
- $Y_{(i)} - Y_{(i-1)} \sim \mathrm{Exp}(\lambda i)$ (memorylessness)
- X has same distribution as $\max\{Y_1, \ldots, Y_m\} = Y_{(m)}$

□

Lemma

$$\sum_{i=1}^{k-1} \mathrm{Exp}(ni) \leq \underbrace{\sum_{i=1}^{k-1} \mathrm{Exp}((n-i)i)}_{L_k \sim} \leq \sum_{i=1}^{k-1} \mathrm{Exp}((n-k)i)$$

Concentration of L_k

Lemma

$$\left(1-e^{-(n-k)t}\right)^k \le \mathbb{P}(L_k \le t) \le \left(1-e^{-nt}\right)^k$$

Proof.

- $X \sim \sum_{i=1}^{m} \mathrm{Exp}(\lambda i)$, then $\mathbb{P}(X \le t) = (1-e^{-\lambda t})^m$.
- $\sum_{i=1}^{k-1} \mathrm{Exp}(ni) \le L_k \le \sum_{i=1}^{k-1} \mathrm{Exp}((n-k)i)$

$\square$

Concentration of L_k

Lemma

$\left(1 - e^{-(n-k)t}\right)^k \leq \mathbb{P}(L_k \leq t) \leq \left(1 - e^{-nt}\right)^k$

Proof.

- $X \sim \sum_{i=1}^{m} \text{Exp}(\lambda i)$, then $\mathbb{P}(X \leq t) = (1 - e^{-\lambda t})^m$.
- $\sum_{i=1}^{k-1} \text{Exp}(ni) \leq L_k \leq \sum_{i=1}^{k-1} \text{Exp}((n-k)i)$

□

Corollary

$\mathbb{P}(L_k > t) \leq 1 - \left(1 - e^{-(n-k)t}\right)^k \leq 1 - \left(1 - ke^{-(n-k)t}\right) = ke^{-(n-k)t}$

Proof.

$(1-x)^y \geq 1 - xy$ □

Concentration of L_k

Lemma

$\left(1-e^{-(n-k)t}\right)^k \leq \mathbb{P}(L_k \leq t) \leq \left(1-e^{-nt}\right)^k$

Proof.

- $X \sim \sum_{i=1}^{m} \mathrm{Exp}(\lambda i)$, then $\mathbb{P}(X \leq t) = (1-e^{-\lambda t})^m$.
- $\sum_{i=1}^{k-1} \mathrm{Exp}(ni) \leq L_k \leq \sum_{i=1}^{k-1} \mathrm{Exp}((n-k)i)$

□

Corollary

$\mathbb{P}(L_k > t) \leq 1-\left(1-e^{-(n-k)t}\right)^k \leq 1-\left(1-ke^{-(n-k)t}\right) = ke^{-(n-k)t}$

Proof.

$(1-x)^y \geq 1-xy$ □

Concentration of L_k

Lemma

$\left(1 - e^{-(n-k)t}\right)^k \le \mathbb{P}(L_k \le t) \le \left(1 - e^{-nt}\right)^k$

Proof.

- $X \sim \sum_{i=1}^{m} \mathrm{Exp}(\lambda i)$, then $\mathbb{P}(X \le t) = (1 - e^{-\lambda t})^m$.
- $\sum_{i=1}^{k-1} \mathrm{Exp}(ni) \le L_k \le \sum_{i=1}^{k-1} \mathrm{Exp}((n-k)i)$

□

Corollary

$\mathbb{P}(L_k > t) \le 1 - \left(1 - e^{-(n-k)t}\right)^k \le 1 - \left(1 - ke^{-(n-k)t}\right) = ke^{-(n-k)t}$

Proof.

$(1-x)^y \ge 1 - xy$

□

Balls around nodes

Corollary

$\mathbb{P}(L_k > t) \leq ke^{-(n-k)t}$

ball around v:

$B_t(v) = \{u \in V \mid d(v,u) \leq t\}$

Corollary

$\mathbb{P}\big(|B_t(v)| < k\big) = \mathbb{P}\big(L_k > t\big) \leq ke^{-(n-k)t}$

Global structure

Lemma

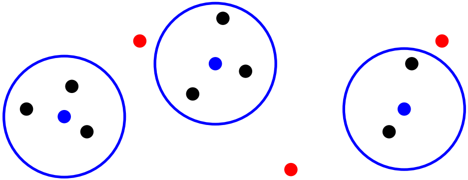
"$\mathbb{P}\big(v \text{ has} < e^{nt} \text{ neighbors within distance } t\big) \leq e^{-nt}$"

example: $$\mathbb{P}\left(v \text{ has} < \log n \text{ neighbors within } \frac{\log\log n}{n}\right) \leq \frac{1}{\log n}$$

Global structure

Lemma

"$\mathbb{P}\big(v$ has $< e^{nt}$ neighbors within distance $t\big) \leq e^{-nt}$"



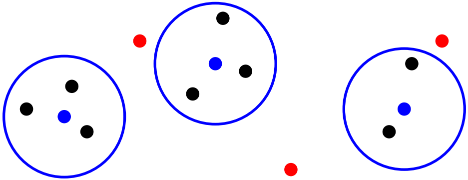
$\frac{n}{e^{nt}}$ components of size e^{nt} and diameter $O(t)$

$\frac{n}{e^{nt}}$ orphans

Global structure

Lemma

"$\mathbb{P}\big(v \text{ has} < e^{nt} \text{ neighbors within distance } t\big) \le e^{-nt}$"



$\frac{n}{e^{nt}}$ components of size e^{nt} and diameter $O(t)$

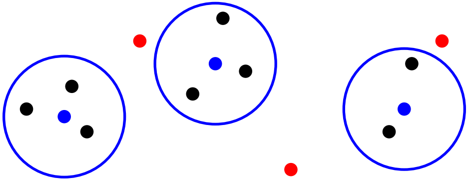
$\frac{n}{e^{nt}}$ orphans

is this really possible without leftover vertices?

Global structure

Lemma

"$\mathbb{P}\big(v \text{ has} < e^{nt} \text{ neighbors within distance } t\big) \leq e^{-nt}$"

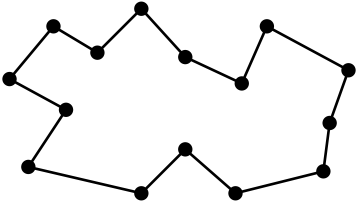


- greedily pick vertex v with $B_t(v) \geq e^{nt}$ without marked neighbor
- mark all vertices in $B_t(v)$ and keep going
- assign remaining vertices to some cluster or keep them as orphans

Outline

1. Properties of RSP
2. Heuristics for TSP
   - Nearest neighbor
   - Insertion heuristics
3. Facility location problem
4. General probability distributions
5. RSP with non-complete graphs
   - Random graphs
   - 2-opt on sparse graphs
6. Conclusions

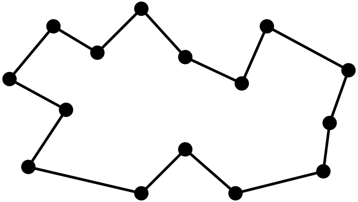
Optimal TSP tour



Theorem

$\mathbb{E}(\text{length of optimal TSP tour}) = \Omega(1)$ (in fact, $\Theta(1)$)

Optimal TSP tour

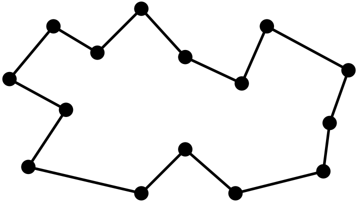


Theorem

$\mathbb{E}$(length of optimal TSP tour) $= \Omega(1)$ (in fact, $\Theta(1)$)

- length of shortest edge at v does not change under shortest path
- length of shortest edge at $v \sim \text{Exp}(n-1)$
- $\mathbb{E}$(length of shortest edge at v) $= \frac{1}{n-1}$
- linearity of expectation

Optimal TSP tour

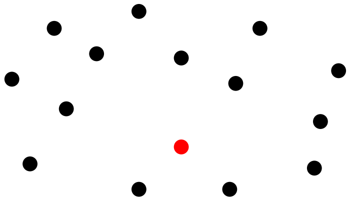


Theorem

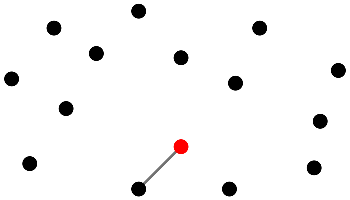
$\mathbb{E}$(length of optimal TSP tour) $= \Omega(1)$ (in fact, $\Theta(1)$)

- length of shortest edge at v does not change under shortest path
- length of shortest edge at $v \sim \mathrm{Exp}(n-1)$
- $\mathbb{E}$(length of shortest edge at v) $= \frac{1}{n-1}$
- linearity of expectation
- holds even without RSP

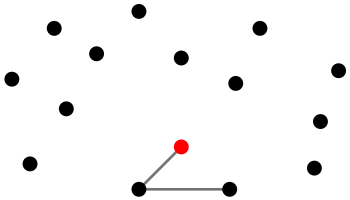
Nearest neighbor for TSP



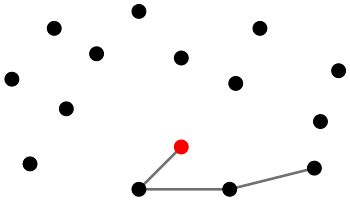
Nearest neighbor for TSP



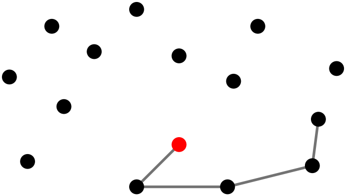
Nearest neighbor for TSP



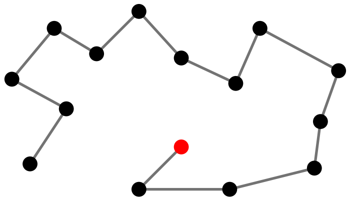
Nearest neighbor for TSP



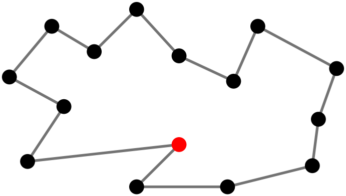
Nearest neighbor for TSP



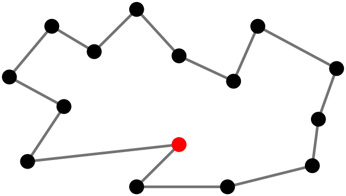
Nearest neighbor for TSP



Nearest neighbor for TSP



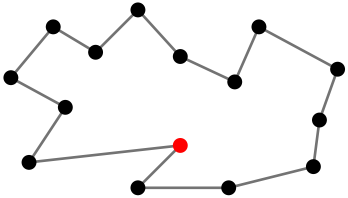
Average-case nearest neighbor without RSP



What is the expected tour length?

distances independent, Exp(1), no RSP, no triangle inequality

Average-case nearest neighbor without RSP

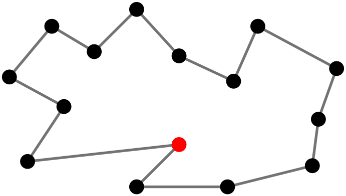


What is the expected tour length?

distances independent, $\text{Exp}(1)$, no RSP, no triangle inequality

1. 1
2. $\log n$
3. $\sqrt{n}$
4. n

Average-case nearest neighbor without RSP

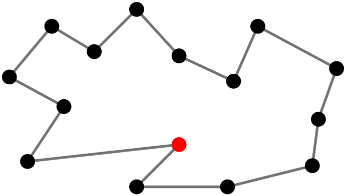


What is the expected tour length?

distances independent, $\mathrm{Exp}(1)$, no RSP, no triangle inequality

- first edge $\sim \mathrm{Exp}(n-1)$

Average-case nearest neighbor without RSP

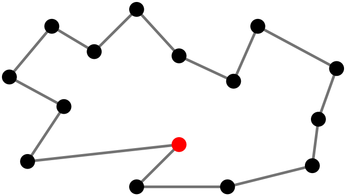


What is the expected tour length?

distances independent, Exp(1), no RSP, no triangle inequality

- first edge $\sim \mathrm{Exp}(n-1)$
- k-th edge $\sim \mathrm{Exp}(n-k)$
- second to last edge $\sim \mathrm{Exp}(1)$

Average-case nearest neighbor without RSP

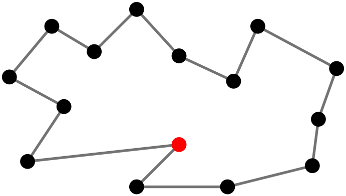


What is the expected tour length?

distances independent, Exp(1), no RSP, no triangle inequality

- first edge $\sim \text{Exp}(n-1)$
- k-th edge $\sim \text{Exp}(n-k)$
- second to last edge $\sim \text{Exp}(1)$
- $\mathbb{E}(\text{last edge}) = \frac{n}{n-1}$

Average-case nearest neighbor without RSP



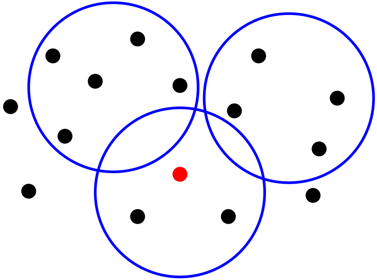
What is the expected tour length?

distances independent, Exp(1), no RSP, no triangle inequality

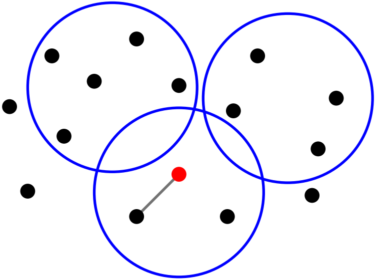
- first edge $\sim \text{Exp}(n-1)$
- k-th edge $\sim \text{Exp}(n-k)$
- second to last edge $\sim \text{Exp}(1)$
- $\mathbb{E}(\text{last edge}) = \frac{n}{n-1}$

► $\mathbb{E}(\text{NN without RSP}) = H_{n-1} + \frac{n}{n-1} = \Theta(\log n)$

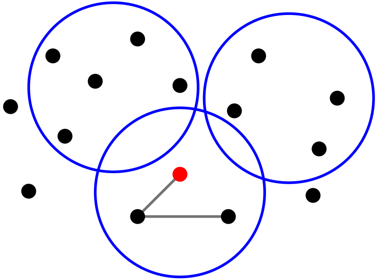
Nearest neighbor for TSP



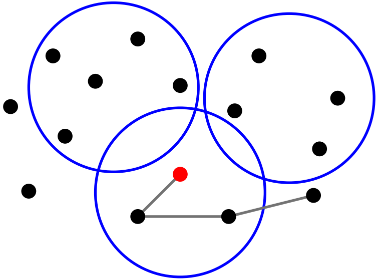
Nearest neighbor for TSP



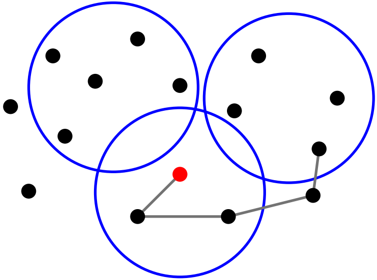
Nearest neighbor for TSP



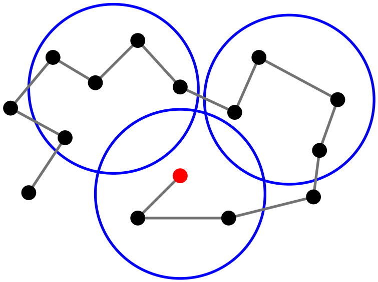
Nearest neighbor for TSP



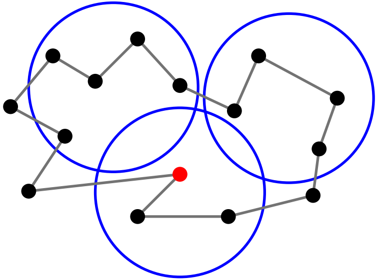
Nearest neighbor for TSP



Nearest neighbor for TSP

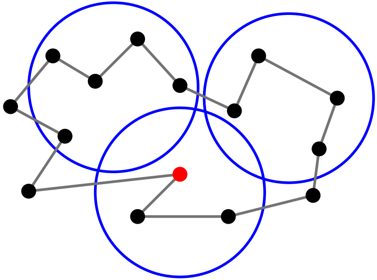


Nearest neighbor for TSP



- $\leq$ 1 long outgoing edge per component/orphan

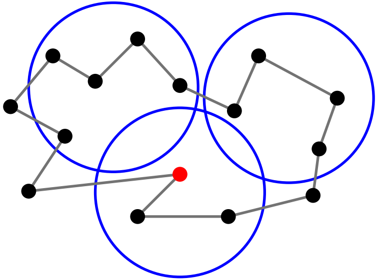
Nearest neighbor for TSP



- ≤ 1 long outgoing edge per component/orphan
- $t_i = \frac{i}{n}$ for $i = 1, 2, 3, \ldots$

  long edges for phase i are estimated as $\leq t_{i+1}$

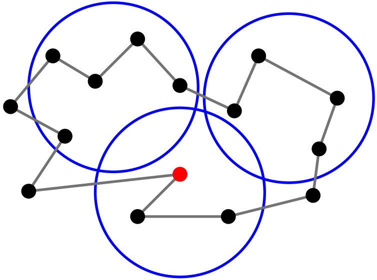
Nearest neighbor for TSP



- ≤ 1 long outgoing edge per component/orphan
- $t_i = \frac{i}{n}$ for $i = 1, 2, 3, \ldots$

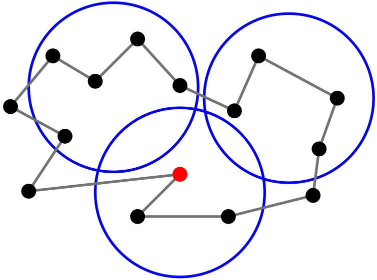
  long edges for phase i are estimated as $\leq t_{i+1}$
- $\leq \frac{n}{e^{nt_i}} = e^{-i}n$ edges of length between t_i and t_{i+1}

Nearest neighbor for TSP



- ≤ 1 long outgoing edge per component/orphan
- $t_i = \frac{i}{n}$ for $i = 1, 2, 3, \dots$
  long edges for phase i are estimated as $\leq t_{i+1}$
- $\leq \frac{n}{e^{nt_i}} = e^{-i}n$ edges of length between t_i and t_{i+1}
- tour length $\leq \sum_{i=0}^{\infty} e^{-i} n t_{i+1} = O(1)$

Nearest neighbor for TSP



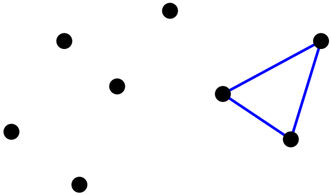
- ≤ 1 long outgoing edge per component/orphan
- $t_i = \frac{i}{n}$ for $i = 1, 2, 3, \ldots$

  long edges for phase i are estimated as $\leq t_{i+1}$
- $\leq \frac{n}{e^{nt_i}} = e^{-i} n$ edges of length between t_i and t_{i+1}
- tour length $\leq \sum_{i=0}^{\infty} e^{-i} n t_{i+1} = O(1)$

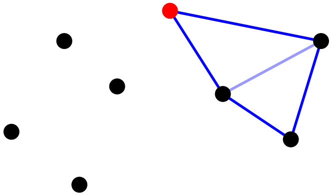
Theorem

nearest neighbor: expected length $O(1)$, expected approximation ratio $O(1)$

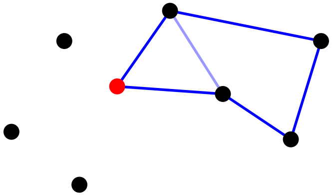
Insertion heuristics



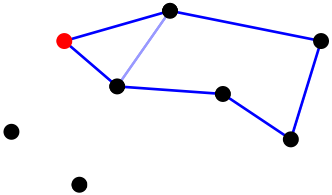
Insertion heuristics



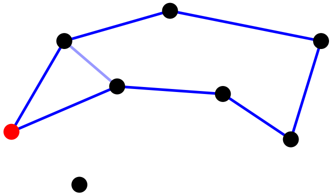
Insertion heuristics



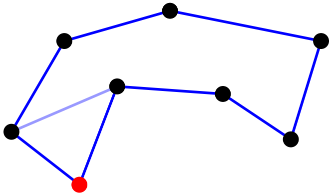
Insertion heuristics



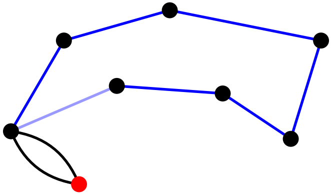
Insertion heuristics



Insertion heuristics



Insertion heuristics



- every component: cheap insertion from second point on
- costs $> t$ only once per component

Theorem

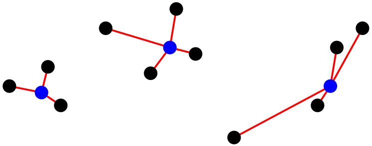
every insertion heuristic achieves expected ratio $O(1)$

Outline

1. Properties of RSP
2. Heuristics for TSP
   - Nearest neighbor
   - Insertion heuristics
3. Facility location problem
4. General probability distributions
5. RSP with non-complete graphs
   - Random graphs
   - 2-opt on sparse graphs
6. Conclusions

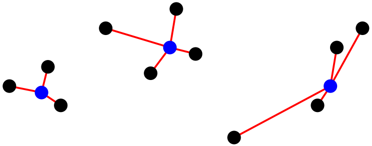
k-center

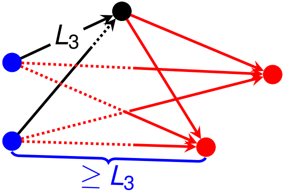
- find $C \subseteq V$ with $|C| = k$
- minimize $\sum_{v \in V} \min_{c \in C} \big(d(v, c) \big)$



k-center

- find $C \subseteq V$ with $|C| = k$
- minimize $\sum_{v \in V} \min_{c \in C} \big(d(v, c)\big)$



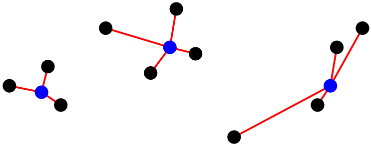


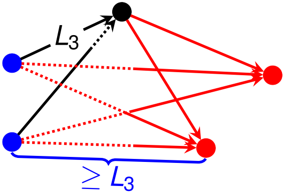
- fixed C
- $L_{k+1}, \ldots, L_n$: distances to C
- $L_1, \ldots, L_k = 0$
- $t_j = L_{j+1} - L_j = \mathrm{Exp}\big(j \cdot (n-j)\big)$

$$\text{costs} = \sum_{i=k+1}^{n} L_i = \sum_{j=k}^{n-1} (n-j) \cdot t_j = \sum_{j=k}^{n-1} \mathrm{Exp}\big(j\big)$$

k-center

- find $C \subseteq V$ with $|C| = k$
- minimize $\sum_{v \in V} \min_{c \in C} \big(d(v, c)\big)$





- fixed C
- $L_{k+1}, \ldots, L_n$: distances to C
- $L_1, \ldots, L_k = 0$
- $t_j = L_{j+1} - L_j = \mathrm{Exp}\big(j \cdot (n-j)\big)$

$$\text{costs} = \sum_{i=k+1}^{n} L_i = \sum_{j=k}^{n-1} (n-j) \cdot t_j = \sum_{j=k}^{n-1} \mathrm{Exp}(j)$$

Theorem

any solution is a $\big(1 + o(1)\big)$-approximation for $k = O(n^{1-\varepsilon})$

Facility location problem

- find $C \subseteq V$
- minimize $\sum_{v \in C} f_v + \sum_{v \in V} \min_{c \in C} \big(d(v, c) \big)$

Facility location problem

- find $C \subseteq V$
- minimize $\sum_{v \in C} f_v + \sum_{v \in V} \min_{c \in C} \big(d(v, c) \big)$
- $f_1 \leq f_2 \leq \ldots \leq f_n$ (adversarial)
- opening k cheapest facilities: $\mathbb{E}(\text{costs}) = \sum_{i=1}^{k} f_i + \sum_{i=k}^{n-1} \frac{1}{i}$

Facility location problem

- find $C \subseteq V$
- minimize $\sum_{v \in C} f_v + \sum_{v \in V} \min_{c \in C} \big(d(v, c)\big)$
- $f_1 \leq f_2 \leq \ldots \leq f_n$ (adversarial)
- opening k cheapest facilities: $\mathbb{E}(\text{costs}) = \sum_{i=1}^{k} f_i + \sum_{i=k}^{n-1} \frac{1}{i}$
- ALG: choose $k = \max\{i \mid f_i < \frac{1}{i-1}\}$ facilities

Facility location problem

- find $C \subseteq V$
- minimize $\sum_{v \in C} f_v + \sum_{v \in V} \min_{c \in C} (d(v, c))$
- $f_1 \leq f_2 \leq \ldots \leq f_n$ (adversarial)
- opening k cheapest facilities: $\mathbb{E}(\text{costs}) = \sum_{i=1}^{k} f_i + \sum_{i=k}^{n-1} \frac{1}{i}$
- ALG: choose $k = \max\{i \mid f_i < \frac{1}{i-1}\}$ facilities
- more challenging: lower bounds for optimal solution
- Cauchy–Schwarz for "decoupling":

$$\mathbb{E}\left(\frac{\text{ALG}}{\text{OPT}}\right) \leq \sqrt{\mathbb{E}(\text{ALG}^2) \cdot \mathbb{E}\left(\frac{1}{\text{OPT}^2}\right)}$$

Facility location problem

- find $C \subseteq V$
- minimize $\sum_{v \in C} f_v + \sum_{v \in V} \min_{c \in C} \big(d(v, c) \big)$
- $f_1 \leq f_2 \leq \ldots \leq f_n$ (adversarial)
- opening k cheapest facilities: $\mathbb{E}(\text{costs}) = \sum_{i=1}^{k} f_i + \sum_{i=k}^{n-1} \frac{1}{i}$
- ALG: choose $k = \max\{i \mid f_i < \frac{1}{i-1}\}$ facilities
- more challenging: lower bounds for optimal solution
- Cauchy–Schwarz for "decoupling":

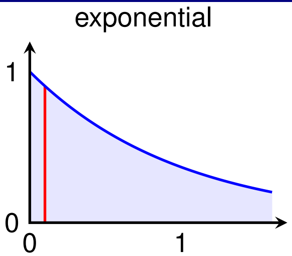
$$\mathbb{E}\left(\frac{\text{ALG}}{\text{OPT}}\right) \leq \sqrt{\mathbb{E}(\text{ALG}^2) \cdot \mathbb{E}\left(\frac{1}{\text{OPT}^2}\right)}$$

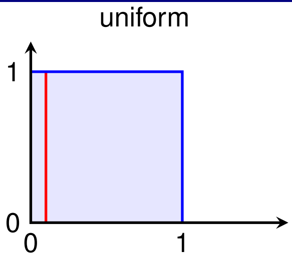
Theorem

ALG is a $\big(1 + o(1)\big)$-approximation for $k = o(n)$

Outline

General probability distributions

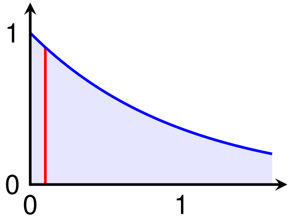




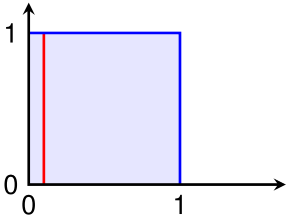
can we transfer the results to uniform/arbitrary distributions?

General probability distributions

exponential



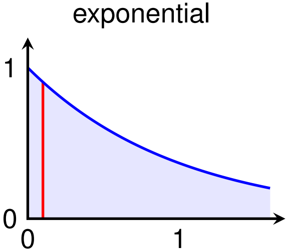
uniform



can we transfer the results to uniform/arbitrary distributions?

- observation: distances decrease with n

General probability distributions



can we transfer the results to uniform/arbitrary distributions?

- observation: distances decrease with n
- density f: differentiable in $(0, \varepsilon)$
- $\mathbb{P}(\text{weight} \leq x) = x + o(x)$
  (every distribution is approximately uniform in $(0, \varepsilon)$)

► results carry over, scale by $1 \pm o(1)$

Outline

RSP generated from non-complete graphs

- balls grow extremely fast: $\left|B_t(v)\right| = \left|\left\{u \in V \mid d(v,u) \leq t\right\}\right| \approx e^{nt}$
- unbounded doubling dimension

RSP generated from non-complete graphs

- balls grow extremely fast: $\left|B_t(v)\right| = \left|\left\{u \in V \mid d(v,u) \leq t\right\}\right| \approx e^{nt}$
- unbounded doubling dimension
- RSP metrics can be generated from any connected graph

  (the metric is always complete!)

► use non-complete/sparse graphs!

RSP generated from non-complete graphs

- balls grow extremely fast: $\left|B_t(v)\right| = \left|\left\{u \in V \mid d(v,u) \leq t\right\}\right| \approx e^{nt}$
- unbounded doubling dimension
- RSP metrics can be generated from any connected graph

  (the metric is always complete!)

► use non-complete/sparse graphs!

- challenges:
  - unknown structure
  - lack of symmetry

RSP with $G_{n,p}$ random graphs

$G_{n,p}$

- n vertices
- $\mathbb{P}\big(\{u, v\} \in E\big) = p$ independently

RSP with $G_{n,p}$ random graphs

$G_{n,p}$

- n vertices
- $\mathbb{P}(\{u, v\} \in E) = p$ independently

RSP on $G_{n,p}$ is a two-stage random process

1. draw random graph – connected w.h.p.
2. draw edge weights for existing edges
3. shortest paths

RSP with $G_{n,p}$ random graphs

$G_{n,p}$

- n vertices
- $\mathbb{P}\big(\{u, v\} \in E\big) = p$ independently

RSP on $G_{n,p}$ is a two-stage random process

1. draw random graph – connected w.h.p.
2. draw edge weights for existing edges
3. shortest paths

can we reuse results for complete graphs?

RSP with $G_{n,p}$ – coupling

$$
\begin{aligned}
& G_{n,p} \text{ with } \mathrm{Exp}(1) \\
\approx\ & G_{n,p} \text{ with } U(0,1) \\
=\ & \text{complete graph with } U(0,1/p)\text{, remove edges of weight} \geq 1 \\
\equiv\ & \text{complete graph with } U(0,1) \text{ (scaling, weight} \geq p \text{ is never used)} \\
\approx\ & \text{complete graph with } \mathrm{Exp}(1)
\end{aligned}
$$

RSP with $G_{n,p}$ – coupling

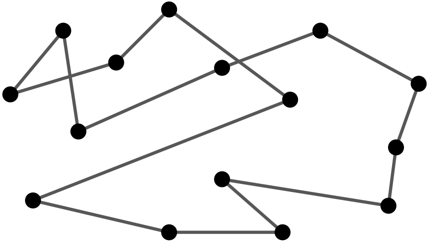
$$\begin{aligned}
& G_{n,p} \text{ with } \mathrm{Exp}(1) \\
\approx\ & G_{n,p} \text{ with } U(0,1) \\
=\ & \text{complete graph with } U(0,1/p)\text{, remove edges of weight } \geq 1 \\
\equiv\ & \text{complete graph with } U(0,1) \text{ (scaling, weight } \geq p \text{ is never used)} \\
\approx\ & \text{complete graph with } \mathrm{Exp}(1)
\end{aligned}$$

all “sum of lengths” results scale by $1/p$

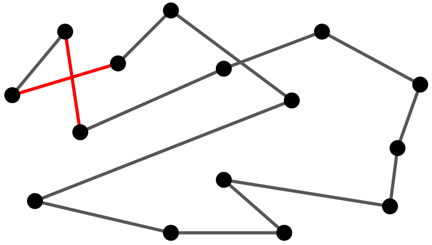
Theorem

“approximation results for complete graphs also hold for $G_{n,p}$”

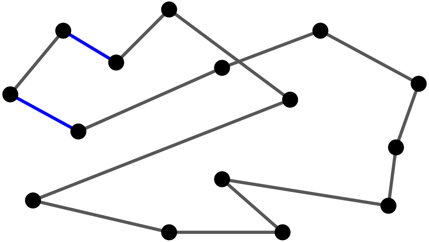
2-opt heuristic for TSP



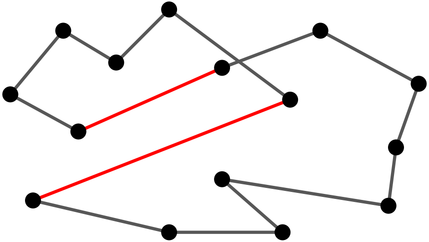
2-opt heuristic for TSP



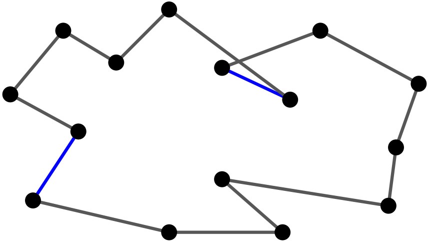
2-opt heuristic for TSP



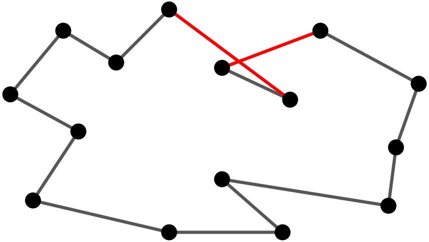
2-opt heuristic for TSP



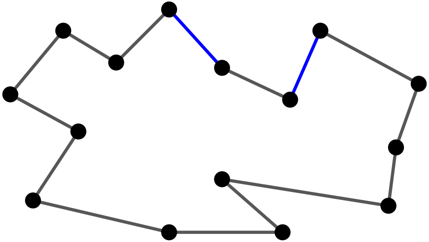
2-opt heuristic for TSP



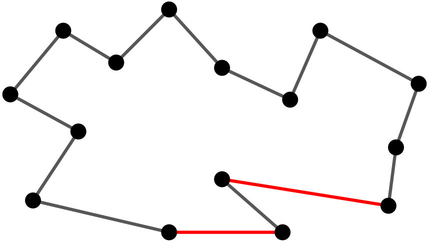
2-opt heuristic for TSP



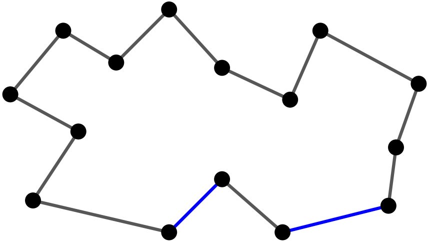
2-opt heuristic for TSP



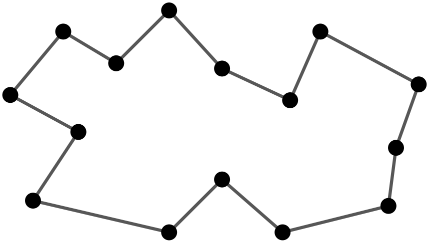
2-opt heuristic for TSP



2-opt heuristic for TSP



2-opt heuristic for TSP – approximation ratio

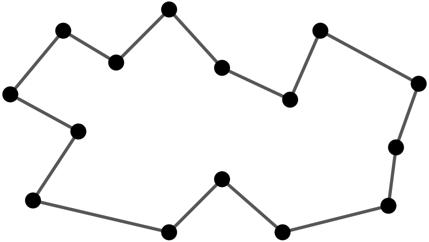


worst-case: $O(\sqrt{n})$

RSP on complete graphs: $O(\log n) \rightsquigarrow$ trivial – why?

RSP on sparse graph: $O(1) \rightsquigarrow$ now

2-opt heuristic for TSP – approximation ratio



worst-case: $O(\sqrt{n})$

RSP on complete graphs: $O(\log n) \rightsquigarrow$ trivial – why?

maximum edge $\approx 3 \cdot \frac{\ln n}{n}$ & optimal tour $= \Omega(1)$

RSP on sparse graph: $O(1) \rightsquigarrow$ now

2-opt – RSP in sparse graphs

- sparse graph: $m = \Theta(n)$ edges
- S_k: sum of k lightest edge weights
- $S_n \leq \text{TSP}$

2-opt – RSP in sparse graphs

- sparse graph: $m = \Theta(n)$ edges
- S_k: sum of k lightest edge weights
- $S_n \leq$ TSP

Lemma

$\mathbb{E}(S_k) = \Omega(k^2/n)$

2-opt – RSP in sparse graphs

- sparse graph: $m = \Theta(n)$ edges
- S_k: sum of k lightest edge weights
- $S_n \leq \text{TSP}$

Lemma

$\mathbb{E}(S_k) = \Omega(k^2/n)$

- ► $\text{TSP} = \Omega(n)$

2-opt – RSP in sparse graphs

Lemma

$\mathbb{E}(S_k) = \Omega(k^2/n)$

Proof.

- $w_1, w_2, \ldots$: edge weigths in increasing order
- $w_1 \sim \mathrm{Exp}(m)$
- $w_{i+1} - w_i \sim \mathrm{Exp}(m-i) \rightsquigarrow w_i \sim \sum_{j=0}^{i-1} \mathrm{Exp}(m-j)$
- $S_k = \sum_{i=1}^{k} w_i \sim \sum_{j=0}^{k-1} (k-j)\, \mathrm{Exp}(m-j) = \sum_{j=0}^{k-1} \mathrm{Exp}\left(\frac{m-j}{k-j}\right)$

  $\geq \sum_{j=0}^{k-1} \mathrm{Exp}\left(\frac{m}{k}\right)$
- $\mathbb{E}(S_k) \geq \frac{k^2}{m}$

□

Approximation ratio of 2-opt in sparse graphs

- $P_{(u,v)} =$ edges of original graph on shortest u–v path
- edges are considered directed for $P_{(u,v)}$

Lemma

$e, f \in T$ with $e \neq f$ and $P_e \cap P_f \neq \emptyset$, then T is not 2-optimal

Approximation ratio of 2-opt in sparse graphs

- $P_{(u,v)} =$ edges of original graph on shortest u–v path
- edges are considered directed for $P_{(u,v)}$

Lemma

$e, f \in T$ with $e \neq f$ and $P_e \cap P_f \neq \emptyset$, then T is not 2-optimal

- upper bound for 2-optimal tour: twice the sum of edge weights
- constant fraction of all edge weights appears in optimal tour

Approximation ratio of 2-opt in sparse graphs

- $P_{(u,v)} =$ edges of original graph on shortest u–v path
- edges are considered directed for $P_{(u,v)}$

Lemma

$e, f \in T$ with $e \neq f$ and $P_e \cap P_f \neq \emptyset$, then T is not 2-optimal

- upper bound for 2-optimal tour: twice the sum of edge weights
- constant fraction of all edge weights appears in optimal tour
- 2-optimal tours have length $O(n)$
  global optimum has length $\Omega(n)$

Theorem

2-opt on sparse graphs achieves approximation ratio $O(1)$

Outline

Summary & open problems

summary

- RSP models random metrics
- TSP: nearest neighbor, insertion, 2-opt
- trivial algorithm for facility location & k-center
  (consider as concentration of measure)
- some first results on incomplete graphs

open problems

- more for sparse graphs
- directed graphs
- other models for random metrics?

Literature

1. Davis, Prieditis: The expected length of a shortest path.
   Information Processing Letters, 1993.
2. Janson: One, two, three times $\log n/n$ for paths in a complete graph...
   Combinatorics, Probability & Computing, 1999.
3. Bringmann, Engels, M., Rao: Random shortest paths: ...
   Algorithmica, 2015.
4. Klootwijk, M.: Probabilistic analysis of facility location on random...
   CiE 2019.
5. Klootwijk, M.: Probabilistic analysis of opt. problems on sparse...
   AofA 2020.
6. Klootwijk, M., Visser: Probabilistic analysis of optimization problems...
   Theoretical Computer Science, 2021.