Range Query on Planar Graphs and Applications on Spatial Sensing with Privacy

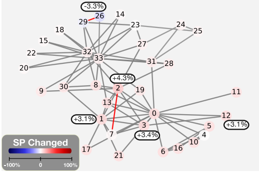
Jie Gao

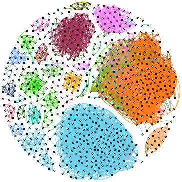
Rutgers University
http://sites.rutgers.edu/jie-gao

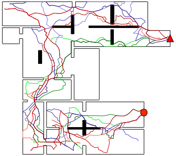
September 16 2020

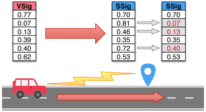
Research Interests: Geometry, Algorithms, Networking

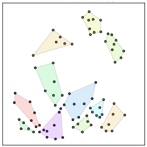
- Trained in algorithm and computational geometry.
- Distributed sensing, spatial data, data privacy.
- Network analysis: social network, contagions, graph neural networks.

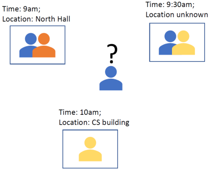






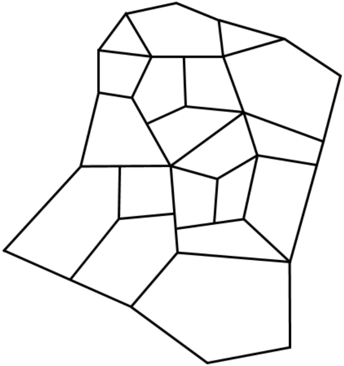






Range Query on a Planar Graph

- A planar graph: each edge e has length $w(e)$ and value $v(e)$.
- Counting range query: report the sum of values along the shortest path between a given pair (u, v).
- Assume that shortesth path is unique.



Naive Solution

- Precompute all $O(n^2)$ query results
  - storage $O(n^2)$
  - query cost $O(1)$ by table lookup.

Naive Solution

- Precompute all $O(n^2)$ query results
  - storage $O(n^2)$
  - query cost $O(1)$ by table lookup.
- No preprocessing:
  - storage $O(n)$,
  - query cost is $O(n \log n)$ – find the shortest path and sum up the values.

Naive Solution

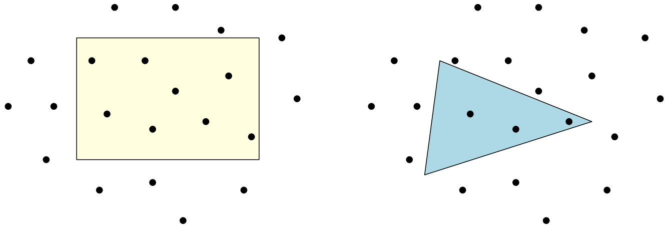
- Precompute all $O(n^2)$ query results
  - storage $O(n^2)$
  - query cost $O(1)$ by table lookup.
- No preprocessing:
  - storage $O(n)$,
  - query cost is $O(n \log n)$ – find the shortest path and sum up the values.

Question: can we design a query data structure such that the storage is $o(n^2)$ and the query cost is $o(n)$?

Classical Range Query Problems

Range query is a well studied problem in computational geometry. Given points in $\mathbb{R}^d$, report the number of points inside

- Orthogonal ranges: rectilinear boxes in $\mathbb{R}^d$.
- Simplex ranges: d-dimensional simplex (e.g., a triangle in 2D).



Storage Query Tradeoff

Orthogonal ranges:

- d-dim range tree: $O(n \log^{d-1} n)$ space, $O(\log^{d-1} n)$ query time.
- kd-tree: $O(n)$ space, $O(n^{1-1/d})$ query time.

Storage Query Tradeoff

Orthogonal ranges:

- d-dim range tree: $O(n \log^{d-1} n)$ space, $O(\log^{d-1} n)$ query time.
- kd-tree: $O(n)$ space, $O(n^{1-1/d})$ query time.

Simplex ranges: with worse bounds

- $O(n^d)$ space, $O(\text{polylog}\, n)$ query time.
- $O(n)$ space, $O(n^{1-1/d})$ query time.

Storage Query Tradeoff

Orthogonal ranges:

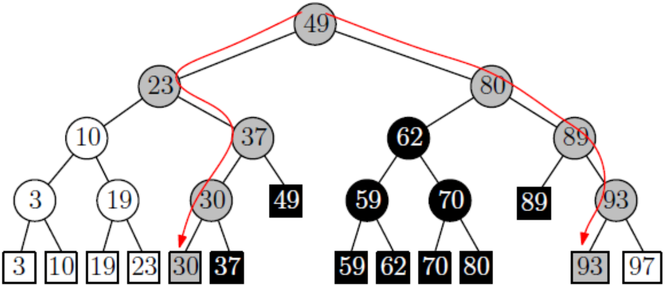
- *d*-**dim range tree**: $O(n \log^{d-1} n)$ space, $O(\log^{d-1} n)$ query time.
- *kd*-tree: $O(n)$ space, $O(n^{1-1/d})$ query time.

Simplex ranges: with worse bounds

- $O(n)$ space, $O(n^{1-1/d})$ query time.
- $O(n^d)$ space, $O(\text{polylog } n)$ query time.

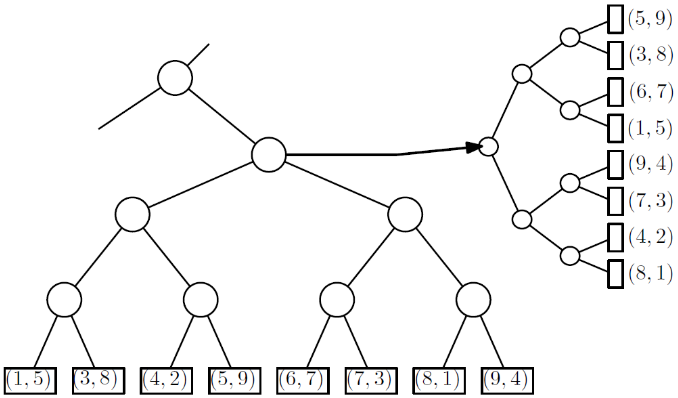
1D Range Tree

Build a binary search tree. Run two queries for the boundary of $[25, 90]$. Take points in between.



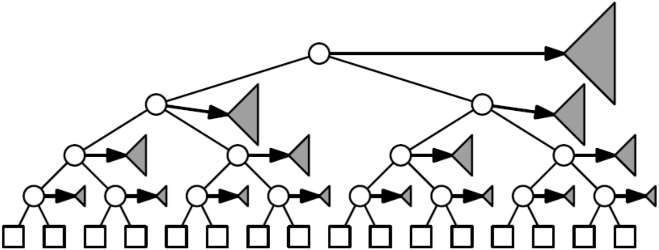
2D Range Tree

Build a 1D range tree on x-coordinates. For each internal node, build a 1D range tree on y-coordinates for all descendants in the main tree.



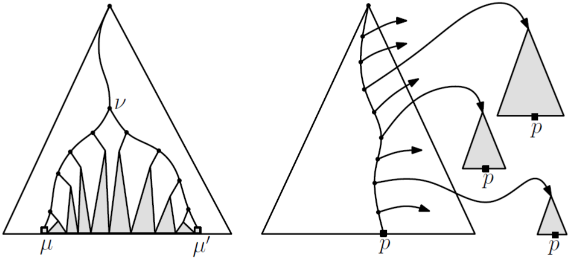
2D Range Tree

Storage?



2D Range Tree

$[\mu, \mu'] \times [\gamma, \gamma']$, query time?

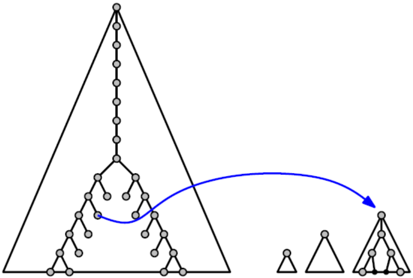


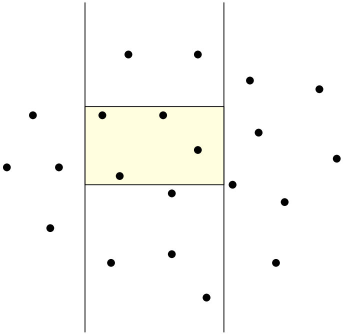
Efficient Range Query

- Precompute answers to k **canonical ranges**, such that each query can be answered by combining the precomputed answers to h of them.

Efficient Range Query

- Precompute answers to k **canonical ranges**, such that each query can be answered by combining the precomputed answers to h of them.
- Ideally storage is $k = \tilde{O}(n)$, and query cost is $h = O(\text{polylog}(n))$.





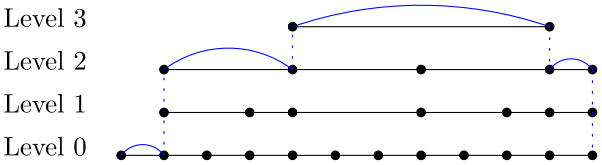
Range Query on Shortest Paths

- How to define 'canonical ranges'?
- How to balance k and h?

Range Query on Shortest Paths: A Simple “1D” Case

Take one shortest path from u to v.

- Build a randomized binary search tree – “skip list”.
- Each node is promoted to the next level with probability $1/2$.
- Canonical range: all edges and shortcuts on all levels. $k = O(n)$
- Query: take the highest possible ‘shortcut’ whenever possible. $h = O(\log n)$.

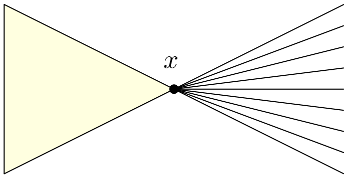


Skip List on a Planar Graph?

What if we run the 'skip list' on a planar graph, for each shortest path P, keep all canonical ranges on P as in the 1D setting?

Skip List on a Planar Graph?

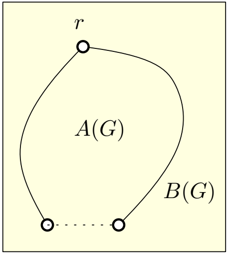
What if we run the 'skip list' on a planar graph, for each shortest path P, keep all canonical ranges on P as in the 1D setting? Unfortunately, this creates too many canonical ranges.

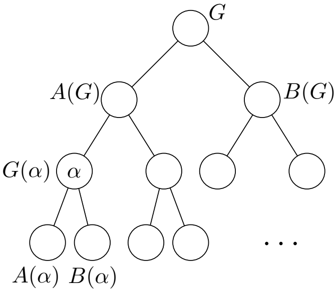


With probability $1 - 1/2^i$, x is not on level i, thus there are ℓ branches and ℓ^2 canonical ranges near x. Take $\ell = n^{1-\varepsilon}$.

Our Solution

Combine a 'space decomposition' data structure with skip list. We use the balanced separator (Lipton & Tarjan) in a planar graph.





Shortest path tree T rooted at r, find a non-tree edge which closes a cycle with edges of T: interior $A(G)$ and exterior $B(G)$.
$|A(G)| \leq 2n/3$; $|B(G)| \leq 2n/3$.

Balanced Separator Hierarchy

Properties of the hierarchical decomposition:

1. Each edge stays on at most one separator in the hierarchical balanced decomposition.
2. Each edge stays in at most $O(\log n)$ subgraphs of the decomposition Φ.

Balanced Separator Hierarchy

Properties of the hierarchical decomposition:

1. Each edge stays on at most one separator in the hierarchical balanced decomposition.
2. Each edge stays in at most $O(\log n)$ subgraphs of the decomposition Φ.

We define two kinds of canonical ranges (p-sums):

- Type-I p-sums on separators

Balanced Separator Hierarchy

Properties of the hierarchical decomposition:

1. Each edge stays on at most one separator in the hierarchical balanced decomposition.
2. Each edge stays in at most $O(\log n)$ subgraphs of the decomposition Φ.

We define two kinds of canonical ranges (p-sums):

- Type-I p-sums on separators $\rightarrow$ '1D'-setting.

Balanced Separator Hierarchy

Properties of the hierarchical decomposition:

1. Each edge stays on at most one separator in the hierarchical balanced decomposition.
2. Each edge stays in at most $O(\log n)$ subgraphs of the decomposition Φ.

We define two kinds of canonical ranges (p-sums):

- Type-I p-sums on separators $\rightarrow$ '1D'-setting.
- Type-II p-sums that stay within a subgraph $G(\alpha)$ for a node α in the balanced separator hierarchy

Balanced Separator Hierarchy

Properties of the hierarchical decomposition:

1. Each edge stays on at most one separator in the hierarchical balanced decomposition.
2. Each edge stays in at most $O(\log n)$ subgraphs of the decomposition Φ.

We define two kinds of canonical ranges (p-sums):

- Type-I p-sums on separators $\rightarrow$ '1D'-setting.
- Type-II p-sums that stay within a subgraph $G(\alpha)$ for a node α in the balanced separator hierarchy $\rightarrow$ only keep the long ones.

Type-II p-sums

Consider a subgraph $G(\alpha)$ with $n(\alpha)$ vertices in the separator hierarchy. A shortest path of two nodes u, v in $G(\alpha)$ is a *canonical path* if the following conditions are met.

- $u, v \in V_i$, with $\log n(\alpha) - q \leq i \leq \log n(\alpha)$, for a value $q = \log \log n$.
- On the shortest path $P(u, v)$, no other vertices are in V_i.

Type-II p-sums

Consider a subgraph $G(\alpha)$ with $n(\alpha)$ vertices in the separator hierarchy. A shortest path of two nodes u, v in $G(\alpha)$ is a *canonical path* if the following conditions are met.

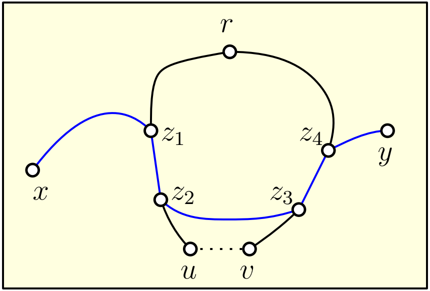
- $u, v \in V_i$, with $\log n(\alpha) - q \leq i \leq \log n(\alpha)$, for a value $q = \log \log n$.
- On the shortest path $P(u, v)$, no other vertices are in V_i.

Lemma: Given a subgraph $G(\alpha)$ for a node α in the separator hierarchy Φ, the expected number of canonical paths defined in $G(\alpha)$ is upper bounded by $O(\log^2 n)$.

Query

The shortest path $P(x, y)$ visits the separator $Z(G)$ at most twice $\rightarrow$ at most 5 segments.

- Two segments handled by Type-I p-sums.
- Three segments handled by Type-II p-sums.

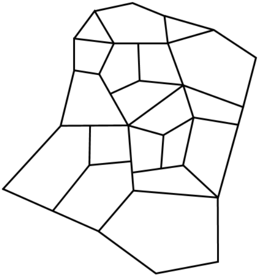


Space and Query Tradeoff

On a planar graph G of n vertices in which each edge carries a sensing reading. We can preprocess the graph G with total storage $O(n \log^2 n)$ such that we can answer range query on any shortest path by querying $O(\log n)$ p-sums.

Applications: Data Privacy in Spatial Sensing

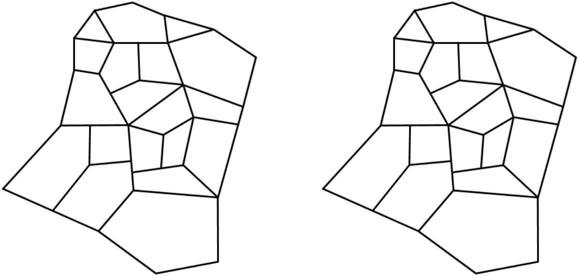
Consider the value $v(e)$ as sensor data on a street segment (e.g., target detection). Answer range queries – how many targets were detected along a shortest path, while protecting the privacy of individual sensor readings.



Differential Privacy

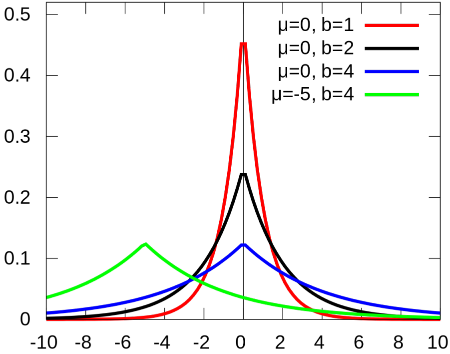
A randomized range query response mechanism M is ε-differentially private if for any two adjacent datasets D and D' (i.e., differ by one sensor reading), for any range $R \in \mathcal{R}$ and any measurable subset $H \in \mathrm{Range}(M)$,

$$\Pr[M_D(R) \in H] \leq \exp(\varepsilon) \cdot \Pr[M_{D'}(R) \in H].$$



Laplace Mechanism

Laplace mechanism: add noise with distribution Lap(b), and its probability density is given as: $P(x|b) = \frac{1}{2b}\exp(-\frac{|x|}{b})$.



Laplace Mechanism

The level of noise is usually determined in terms of *sensitivity*.

Laplace Mechanism

The level of noise is usually determined in terms of *sensitivity*.

The sensitivity of a function f, written as Δf, is the largest possible difference in the output of f between any pair of adjacent databases:

$$\max_{(D,D')} |f(D) - f(D')|.$$

Laplace Mechanism

The level of noise is usually determined in terms of *sensitivity*.

The sensitivity of a function f, written as Δf, is the largest possible difference in the output of f between any pair of adjacent databases:

$$\max_{(D,D')} |f(D) - f(D')|.$$

To achieve ε-differential privacy, a noise of $\mathrm{Lap}(\Delta f / \varepsilon)$ suffices.

Laplace Mechanism

The level of noise is usually determined in terms of *sensitivity*.

The sensitivity of a function f, written as Δf, is the largest possible difference in the output of f between any pair of adjacent databases:

$$\max_{(D,D')} |f(D) - f(D')|.$$

To achieve ε-differential privacy, a noise of $\mathrm{Lap}(\Delta f/\varepsilon)$ suffices.

Add an independent noise of $\mathrm{Lap}(1/\varepsilon)$ on each edge.

Sum of Independent Laplace Distributions

Given γ_i as independent random variables following Laplace distribution $\text{Lap}(b_i)$. Suppose $Y = \sum_i \gamma_i$. Then, $|Y|$ is at most $O(\sqrt{\sum_i b_i^2} \log(1/\delta))$ with probability at least $1 - \delta$, $0 < \delta < 1$.

Sum of Independent Laplace Distributions

Given γ_i as independent random variables following Laplace distribution $\mathrm{Lap}(b_i)$. Suppose $Y = \sum_i \gamma_i$. Then, $|Y|$ is at most $O(\sqrt{\sum_i b_i^2} \log(1/\delta))$ with probability at least $1-\delta$, $0 < \delta < 1$.

If a query range consists of m edges, where each edge adds an independent noise from $\mathrm{Lap}(1/\varepsilon)$, then the total noise of the output is only bounded by $O(\frac{1}{\varepsilon}\sqrt{m} \log \frac{1}{\delta})$ with probability $1-\delta$.

Idea: Adding Noises to p-sums

- x : The number of p-sum values to which each item contributes;
- y : The number of p-sum values needed to answer a range query.

Each p-sum needs a noise of distribution $\mathrm{Lap}(x/\varepsilon)$ added to its pre-computed value. This ensures that ε-differential privacy with $O(\frac{x\sqrt{y}}{\varepsilon})$ error is achieved with high probability.

Idea: Adding Noises to p-sums

- x : The number of p-sum values to which each item contributes;
- y : The number of p-sum values needed to answer a range query.

Each p-sum needs a noise of distribution $\mathrm{Lap}(x/\varepsilon)$ added to its pre-computed value. This ensures that ε-differential privacy with $O(\frac{x\sqrt{y}}{\varepsilon})$ error is achieved with high probability.

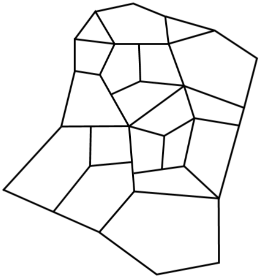
Lemma Each edge (u, v) stays on at most $O(\log^3 n \log\log n)$ p-sums on average.

Differential Privacy for Spatial Sensing

The range query along any shortest path can be answered with error $O(\frac{1}{\varepsilon}\log^{3.5} n \log\log n \cdot \log\frac{1}{\delta})$ with probability at least $1-\delta$, with ε-differential privacy on any event.

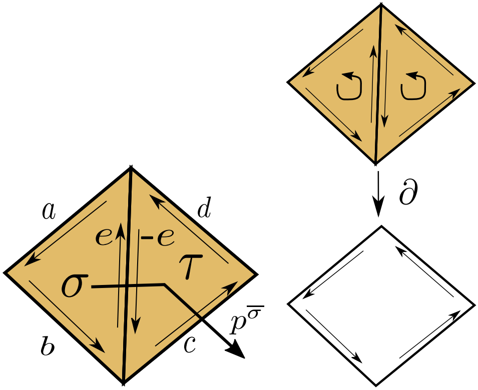
Extension: Range Query on Faces/Regions

The events occur at the faces of the administrative, or city block-wise planar subdivisions of the domain, G and the query ranges of interest are the sets of faces $U \subseteq F$.



Discrete Differential Forms

Translate values at faces to values at edges.
For a 2D range R, apply range query on boundary of R.



Acknowledgement

- Differentially Private Range Counting in Planar Graphs for Spatial Sensing, Abhirup Ghosh, Jiaxin Ding, Rik Sarkar, Jie Gao, Proceedings of the 39th Annual IEEE International Conference on Computer Communications (INFOCOM'20), July 6-9, 2020.