### CPSC 768: Scalable and Private Graph Algorithms

**Lecture 25: Learning-Augmented Algorithms** 

### Quanquan C. Liu quanquan.liu@yale.edu

#### Announcements

- Final project report and presentation: April 24<sup>th</sup> (last day of class)
  - Final project presentation is a 30 min presentation
- Last day of Open Problem Sessions: April 26<sup>th</sup> (last week of classes)
  - Will be turned into a reading group/continue with OPS, stay tuned!

- Classical Algorithms (intro to algorithms courses)
  - Worst-case guarantees
  - Limited adaptivity to input (special classes, closely related inputs)



- Classical Algorithms (intro to algorithms courses)
  - Worst-case guarantees
  - Limited adaptivity to input (special classes, closely related inputs)
- Machine Learning Based Approaches
  - Stronger performance due to adaptivity (learning) inputs
  - No worst-case guarantees





- Classical Algorithms (intro to algorithms courses)
  - Worst-case guarantees
  - Limited adaptivity to input (special classes, closely related inputs)
- Machine Learning Based Approaches
  - Stronger performance due to adaptivity (learning) inputs
  - No worst-case guarantees
- Learning-Augmented Algorithms
  - Adaptive
  - Often has worst-case guarantees









Use machine learning advice to inform algorithmic procedure

- Use machine learning advice to inform algorithmic procedure
  - What part of the input can we reasonably learn?

- Use machine learning advice to inform algorithmic procedure
  - What part of the input can we reasonably learn?
  - How do we use the learned advice to obtain better theoretically proven bounds?

- Use machine learning advice to inform algorithmic procedure
  - What part of the input can we reasonably learn?
  - How do we use the learned advice to obtain better theoretically proven bounds?
  - Example:

- Use machine learning advice to inform algorithmic procedure
  - What part of the input can we reasonably learn?
  - How do we use the learned advice to obtain better theoretically proven bounds?
  - Example:
    - Frequency estimation in a stream

- Use machine learning advice to inform algorithmic procedure
  - What part of the input can we reasonably learn?
  - How do we use the learned advice to obtain better theoretically proven bounds?
  - Example:
    - Frequency estimation in a stream—learn infrequent elements

- Use machine learning advice to inform algorithmic procedure
  - What part of the input can we reasonably learn?
  - How do we use the learned advice to obtain better theoretically proven bounds?
  - Example:
    - Frequency estimation in a stream
      - Learn infrequent elements

#### Consistency

 If the predictions are high quality, then algorithms performs much better than worst-case algorithm

#### Consistency

 If the predictions are high quality, then algorithm performs much better than worst-case algorithm

#### Competitiveness

 If the predictions are low quality, then algorithm does not perform any worse than worst-case algorithm

#### Consistency

 If the predictions are high quality, then algorithm performs much better than worst-case algorithm

#### Competitiveness

 If the predictions are low quality, then algorithm does not perform any worse than worst-case algorithm

#### Robustness

 Performance of algorithm degrades gracefully as function of prediction error

Often cannot satisfy all three!

- Consistency
  - If the predictions are high quality, then algorithm performs much better than worst-case algorithm
- Competitiveness
  - If the predictions are low quality, then algorithm does not perform any worse than worst-case algorithm

#### Robustness

 Performance of algorithm degrades gracefully as function of prediction error

• Suppose we have a **sorted** list of integers

1	3	6	7	8	9	11	15	16	18	19	25	36	78	98	99

- Suppose we have a **sorted** list of integers
- Problem: where is 15?

1	3	6	7	8	9	11	15	16	18	19	25	36	78	98	99

- Suppose we have a **sorted** list of integers
- Problem: where is 15?
- Prediction: predicted index of 15

1	3	6	7	8	9	11	15	16	18	19	25	36	78	98	99

- Suppose we have a **sorted** list of integers
- Problem: where is 15?
- Prediction: predicted index of 15

1	3	6	7	8	9	11	15	16	18	19	25	36	78	98	99

- Suppose we have a **sorted** list of integers
- Problem: where is 15?
- Prediction: predicted index of 15; how can we use it?

1	3	6	7	8	9	11	15	16	18	19	25	36	78	98	99

- Suppose we have a **sorted** list of integers
- Problem: where is 15?
- Prediction: predicted index of 15; doubling search from predicted index!

1	3	6	7	8	9	11	15	16	18	19	25	36	78	98	99

- Suppose we have a **sorted** list of integers
- Problem: where is 15?
- Prediction: predicted index of 15; doubling search from predicted index!

First decide go left or right

- Suppose we have a **sorted** list of integers
- Problem: where is 15?
- Prediction: predicted index of 15; doubling search from predicted index!

- Suppose we have a **sorted** list of integers
- Problem: where is 15?
- Prediction: predicted index of 15; doubling search from predicted index!

- Suppose we have a **sorted** list of integers
- Problem: where is 15?
- Prediction: predicted index of 15; doubling search from predicted index!

- Suppose we have a **sorted** list of integers
- Problem: where is 15?
- Prediction: predicted index of 15; doubling search from predicted index!

- Suppose we have a **sorted** list of integers
- Problem: where is 15?
- Prediction: predicted index of 15; doubling search from predicted index!

- Suppose we have a **sorted** list of integers
- Problem: where is 15?
- Prediction: predicted index of 15; doubling search from predicted index!

- Suppose we have a sorted list of integers
- Problem: where is 15?
- Prediction: predicted index of 15; doubling search from predicted index!

Start doubling search

What is the runtime?

Why is the prediction reasonable?

- Suppose we have a sorted list of integers
- Problem: where is 15?
- Prediction: predicted index of 15; doubling search from predicted index!

Start doubling search

What is the runtime?  $O(\log(L_1 - error))$ 

#### Learned Binary Why is the prediction reasonable?

- Suppose we have a sor
- Many real-world instances give you a • **Problem:** where is 15?
- prediction—e.g. library books Prediction: predicted i... мыным predicted index!

Start doubling search

What is the runtime?  $O(\log(L_1 - error))$ 

### Learned Binary Satisfies all three desired properties!

- Suppose we have a sorted list of integers
- Problem: where is 15?
- Prediction: predicted index of 15; doubling search from predicted index!

Start doubling search

What is the runtime?  $O(\log(L_1 - error))$ 

#### Learned Bloom Filter

- A bloom filter is a data structure that provides approximate membership queries
  - Given: a set of elements  $S = \{x_1, \dots, x_n\}$

#### Learned Bloom Filter

- A bloom filter is a data structure that provides approximate membership queries
  - Given: a set of elements  $S = \{x_1, ..., x_n\}$
  - Goal: given *x*, determine if *x* is in *S*

#### Learned Bloom Filter

- A bloom filter is a data structure that provides approximate membership queries
  - Given: a set of elements  $S = \{x_1, \dots, x_n\}$
  - Goal: given *x*, determine if *x* is in *S*
- Bloom filter gives low space, randomized solution
- A bloom filter is a data structure that provides approximate membership queries
  - Given: a set of elements  $S = \{x_1, \dots, x_n\}$
  - Goal: given *x*, determine if *x* is in *S*
- Bloom filter gives low space, randomized solution
- Many applications!

Bloom filter contains:

- Bloom filter contains:
  - Array of *m* bits, which represents *S*

- Bloom filter contains:
  - Array of m bits, which represents S
  - k independent hash functions  $h_1, ..., h_k$ , each hashing elements in S to  $\{1, ..., m\}$

- Bloom filter contains:
  - Array of m bits, which represents S
  - k independent hash functions  $h_1, ..., h_k$ , each hashing elements in S to  $\{1, ..., m\}$
- Initialization: for each  $x \in S$  and  $i \in \{1, ..., k\}$ , set bit  $h_i(x)$  in array to 1

- Bloom filter contains:
  - Array of m bits, which represents S
  - k independent hash functions  $h_1, ..., h_k$ , each hashing elements in S to  $\{1, ..., m\}$
- Initialization: for each  $x \in S$  and  $i \in \{1, ..., k\}$ , set bit  $h_i(x)$  in array to 1
- On query of x':

• Report in *S* if and only if  $h_i(x') = 1$  for all  $i \in \{1, ..., k\}$ 

- Initialization: for each  $x \in S$  and  $i \in \{1, ..., k\}$ , set bit  $h_i(x)$  in array to 1
- On query of x':
  - Report in *S* if and only if  $h_i(x') = 1$  for all  $i \in \{1, ..., k\}$

1	1	0	0	0	1	0	0	0	1	0	1	0	0	1	0

- Initialization: for each  $x \in S$  and  $i \in \{1, ..., k\}$ , set bit  $h_i(x)$  in array to 1
- On query of x':
  - Report in *S* if and only if  $h_i(x') = 1$  for all  $i \in \{1, ..., k\}$



- Guarantees:
  - No false negatives: if return NO, not in S
  - False positives: returns YES but not in S



- Guarantees:
  - No false negatives: if return NO, not in S
  - False positives: returns YES but not in S
    - **Probability:** For query x and for each  $i \in [k]$ ,  $P[h_i(x) = 1] \le \frac{kn}{m}$

- Guarantees:
  - No false negatives: if return NO, not in S
  - False positives: returns YES but not in S
    - **Probability:** For query *x* and for each  $i \in [k]$ ,  $P[h_i(x) = 1] \le \frac{kn}{m}$
    - $P[all \ 1s] \le \left(\frac{kn}{m}\right)^k$

- Guarantees:
  - No false negatives: if return NO, not in S
  - False positives: returns YES but not in S
    - **Probability:** For query x and for each  $i \in [k]$ ,  $P[h_i(x) = 1] \le \frac{kn}{m}$
    - $P[all \ 1s] \le \left(\frac{kn}{m}\right)^k$
    - Suppose pick  $k = \log_2 n$  and  $m = 3n \log_2 n$ , then probability of failure is  $\left(\frac{n \log_2 n}{3n \log_2 n}\right)^{\log_2 n} = \left(\frac{1}{3}\right)^{\log_2 n} < \frac{1}{n}$

• ML Oracle:

- ML Oracle:
  - Some trained function f(x) provides useful information

- ML Oracle:
  - Some trained function f(x) provides useful information
- What function would be useful for us?

- ML Oracle:
  - Some trained function f(x) provides useful information
- What function would be useful for us?
  - Function that tells us whether  $x \in S!$

- ML Oracle:
  - Some trained function f(x) provides useful information
- What function would be useful for us?
  - Function that tells us whether  $x \in S!$
- Caveat:
  - We can't learn this exactly!

- ML Oracle:
  - Some trained function f(x) provides useful information
- What function would be useful for us?
  - Function that tells us whether  $x \in S!$
- Caveat:
  - We can't learn this exactly!
  - Will be non-zero false positive and negative rate

• [Kraska, Beutal, Chi, Dean, Polyzotis, SIGMOD '18]

• [Kraska, Beutal, Chi, Dean, Polyzotis, SIGMOD '18]



• [Kraska, Beutal, Chi, Dean, Polyzotis, SIGMOD '18]



• [Kraska, Beutal, Chi, Dean, Polyzotis, SIGMOD '18]



#### No false negatives!

• [Kraska, Beutal, Chi, Dean, Polyzotis, SIGMOD '18]



#### No false negatives!



#### No false negatives!

### Sandwiched Learned Bloom Filter

• [Mitzenmacher, NeurIPS '18]



#### No false negatives!

### Sandwiched Learned Bloom Filter

• [Mitzenmacher, NeurIPS '18]



### False positive rate same or less than Bloom filter!

### Sandwiched Learned Bloom Filter

• [Mitzenmacher, NeurIPS '18]

Better analysis and performance in practice!



False positive rate same or less than Bloom filter!

• Why do we need them?

# **Types of Dynamic Algorithms**

- Incremental/Decremental vs. Fully Dynamic
  - Incremental/decremental algorithms:
    - Only edge insertions/deletions, respectively
- Sometimes large gap in runtimes
  - Polynomial or exponential gaps in runtimes

## **Types of Dynamic Algorithms**

**Best Fully Dynamic** 

**Best Partially Dynamic** 

Planar Digraph APSP	$\widetilde{O}\left(n^{2/3} ight)$	[FR06, Kle05]	$\widetilde{O}(\sqrt{n})$	[DGWN22]
Triconnectivity	$O(n^{2/3})$	[GIS99]	$\widetilde{O}\left(1 ight)$	[HR20, PSS17]
k-Edge Connectivity	$n^{o(1)}$	[JS22]	$\widetilde{O}(1)$	$[CDK^+21]$
Dynamic DFS Tree	$\widetilde{O}\left(\sqrt{mn} ight)$	[BCCK19]	$\widetilde{O}\left(n ight)$	$[\mathrm{BCCK19},\mathrm{CDW^{+}18}]$
Minimum Spanning Forest	$\widetilde{O}(1)$	[HDLT01]	$\widetilde{O}(1)$	[Epp94]
APSP	$egin{array}{l} \left(rac{256}{k^2} ight)^{4/k} ext{-}  ext{Approx} \ \widetilde{O}\left(n^k ight)  ext{ update} \ \widetilde{O}(n^{k/8})  ext{ query} \end{array}$	[FGNS23]	$(2r-1)^k$ -Approx $\widetilde{O}\left(m^{1/(k+1)}n^{k/r} ight)$	$[CGH^+20]$
AP Maxflow/Mincut	$O(\log(n)\log\log n) ext{-}\operatorname{Approx} \ \widetilde{O}\left(n^{2/3+o(1)} ight)$	$[CGH^+20]$	$O\left(\log^{8k}(n) ight)$ -Approx. $\widetilde{O}\left(n^{2/(k+1)} ight)$	[Gor19, GHS19]
MCF	$egin{array}{lll} (1+arepsilon) ext{-Approx} \ \widetilde{O}(1)  ext{ update} \ \widetilde{O}(n)  ext{ query} \end{array}$	[CGH+20]	$egin{aligned} O(\log^{8k}(n)) ext{-}\operatorname{Approx.} \ \widetilde{O}\left(n^{2/(k+1)} ight)  ext{ update} \ \widetilde{O}(P^2)  ext{ query} \end{aligned}$	[Gor19, GHS19]
Strongly Connected Components	$\Omega(m^{1-\varepsilon})$ query or update	[AW14]	$\widetilde{O}(m)$	[Rod13]
Uniform Sparsest Cut	$2^{O(\log^{5/6}(n))}$ -Approx $2^{O(\log^{5/6}(n))}$ update $O(\log^{1/6}(n))$ query	[GRST21]	$egin{aligned} &O\left(\log^{8k}(n) ight) ext{-}\mathrm{Approx}\ &\widetilde{O}\left(n^{2/(k+1)} ight)\ &O(1)  ext{ query} \end{aligned}$	[Gor19, GHS19]
Submodular Max	$1/4 ext{-Approx} \ \widetilde{O}(k^2)$	[DFL <sup>+</sup> 23]	$0.3178 ext{-}\operatorname{Approx} \ \widetilde{O}\left(\operatorname{poly}(k) ight)$	$[FLN^+22]$

[**L**S '23]

- Why do we need them?
- Can we use predictions to bridge the gap?

- Why do we need them?
- Can we use predictions to bridge the gap?
  - Yes!

- Why do we need them?
- Can we use predictions to bridge the gap?
  - Yes!
- Predictions on the edge updates [L-Srinivas '23]

- Why do we need them?
- Can we use predictions to bridge the gap?
  - Yes!
- Predictions on the edge updates [L-Srinivas '23]
  - For each edge update, give prediction on when update occurs

- Why do we need them?
- Can we use predictions to bridge the gap?
  - Yes!
- Predictions on the edge updates [L-Srinivas '23]
  - For each edge update, give prediction on when update occurs
  - Assume one edge insertion/deletion occurs on a day, give prediction on the day of the edge insertion/deletion

## **Offline-Dynamic Connectivity**

- Geometric representation of the problem
- Divide-and-conquer: process each subproblem


Pick a uniformly at random divider for subproblems



**CPSC 768** 



Pick a uniformly at random divider for subproblems



**CPSC 768** 

#### Run offline divide-and-conquer algorithm on subproblems



**CPSC 768** 

Run offline divide-and-conquer algorithm on subproblems



Predicted

 Time
 Image: Image:

Run offline divide-and-conquer algorithm on subproblems



Predicted

 Time
 Image: I(e\_1) I(e\_2) I(e\_3) Q(u,w) I(e\_4) I(e\_5) D(e\_3) D(e\_3) D(e\_4) Q(u,w)

 Actual
  $I(e_1) I(e_2) I(e_3) Q(u,w) I(e_5) I(e_4) D(e_3) D(e_3) D(e_4) Q(u,w)$ 

# Recompute computation of largest subtree containing errors and children





#### Predicted

 Time
 Image: Image:

### Random Partition Tree Data Structure Purpose of the random partition tree: in expectation size of subproblem (largest subtree) equal to error





#### Predicted

 Time
 Image: Image:

### Random Partition Tree Data Structure Purpose of the random partition tree: in expectation size of subproblem (largest subtree) equal to L<sub>1</sub>error

**Bad event: very large subtree for small error** 



#### Predicted

 Time
 Image: I(e\_1)
 I(e\_2)
 I(e\_3)
 Q(u,w)
 I(e\_4)
 I(e\_5)
 D(e\_3)
 D(e\_4)
 Q(u,w)

 Actual
  $I(e_1)$   $I(e_2)$   $I(e_3)$  Q(u,w)  $I(e_5)$   $I(e_3)$   $D(e_3)$   $D(e_4)$  Q(u,w) 





• Streaming, distributed, parallel and dynamic techniques can be used in a variety of problems across models

- Streaming, distributed, parallel and dynamic techniques can be used in a variety of problems across models
- Modern algorithms must take into account stronger adversaries
  - Privacy-violating
  - Decentralized adversaries

- Streaming, distributed, parallel and dynamic techniques can be used in a variety of problems across models
- Modern algorithms must take into account stronger adversaries
  - Privacy-violating
  - Decentralized adversaries
- New tools to help us make algorithms practical!

- Streaming, distributed, parallel and dynamic techniques can be used in a variety of problems across models
- Modern algorithms must take into account stronger adversaries
  - Privacy-violating
  - Decentralized adversaries
- New tools to help us make algorithms practical!
  - Models that model modern architecture
  - ML-based methods

- Streaming, distributed, parallel and dynamic techniques can be used in a variety of problems across models
- Modern algorithms must take into account stronger adversaries
  - Privacy-violating
  - Decentralized adversaries
- New tools to help us make algorithms practical!
  - Models that model modern architecture
  - ML-based methods
- Graphs at the forefront of these developments!

Exciting new fields with lots of potential for development and broad impact

- Streaming, distributed, parallel and dynamic techniques can be used in a variety of problems across models
- Modern algorithms must take into account stronger adversaries
  - Privacy-violating
  - Decentralized adversaries
- New tools to help us make algorithms practical!
  - Models that model modern architecture
  - ML-based methods
- Graphs at the forefront of these developments!