

COMP 251 Study guide

Francis Piché

April 26, 2018

Contents

1 Disclaimer	9
2 Preliminaries	9
I Recursive Algorithms	9
3 Divide + Conquer Algorithms	9
3.1 MergeSort	10
3.2 Binary Search	10
3.3 Run Time of Divide + Conquer in General	11
3.4 Aside on Recurrences: Domain Transformation	11
4 Master Theorem	12
4.1 Tree Method to Prove Master Theorem	15
5 Multiplication	15
5.1 Grade School Multiplication	15
5.2 Russian Peasant Multiplication	15
5.3 Divide + Conquer Multiplication	16
5.4 Fast Fourier Transforms	17
5.5 Multiplying Matrices	17
5.6 Fast Exponentiation	18
6 The Median Problem	18
6.1 The Selection Problem	18
6.2 Median of Medians	19
7 Finding the Closest Pair of Points in the Plane	20
7.1 Exhaustive Search	20
7.2 2-D case	20
7.3 Widening the Bottleneck	21
7.4 The Finished Algorithm	22
7.5 The Runtime (Enhanced)	22
II Graph Algorithms	22
8 Theorems About Undirected Graphs	22
8.1 Handshaking Lemma	22
8.2 Leaf Existence	23
8.3 Number of edges in a Tree	23
8.4 Halls Theorem	24

9 Breadth First Search	24
9.1 Generic Search Algorithm	24
9.1.1 Revised Generic Search Algorithm	24
9.1.2 The Running Time	25
9.1.3 Validity	25
9.2 Search Trees	25
9.3 Choices of Bags	26
9.4 BFS Trees	26
9.4.1 Structure	26
9.4.2 BFS on Bipartite Graphs	26
10 Depth First Search	27
10.1 DFS Trees	27
10.2 Recursive DFS	27
10.3 Ancestral Edges	28
10.4 Previsit and Postvisit	28
10.5 Directed Graph BFS Tree Structure	29
10.6 Example: Directed Acyclic Graphs	29
10.7 Example: Topological Ordering	30
III Greedy Algorithms	30
11 Scheduling	30
11.1 Task Scheduling	30
11.1.1 Running Time	31
11.2 Class Scheduling	31
11.2.1 First Start	32
11.2.2 Shortest-Duration	32
11.2.3 Minimum Conflict	32
11.2.4 Last Start	32
12 The Shortest Path Problem	33
12.1 Dijkstra's Shortest Path Algorithm	34
12.1.1 Special Case, All Arcs Have Distance 1	34
12.1.2 Shortest Path Graph	34
12.1.3 Shortest Path Tree	34
12.1.4 The Running Time	36
13 Huffman Codes	36
13.1 Data Encoding	36
13.1.1 Morse Code	36
13.2 Prefix Codes	37
13.3 Binary Tree Representations	37
13.3.1 Letter to Leaf Assignment	38

13.3.2 Tree Shape	38
13.4 The Key Formula	38
13.5 The Algorithm	39
13.5.1 Proof Of Correctness	39
13.5.2 Running Time	39
14 Minimum Spanning Tree Problem	39
14.1 Kruskal's Algorithm	39
14.2 Prim's Algorithm	40
14.3 Boruvka's Algorithm	40
14.4 Running Times	40
14.4.1 Kruskal's Running Time	40
14.4.2 Prim's Running Time	40
14.4.3 Boruvka's Running Time	40
14.5 Proof That They All Work	41
14.6 The Cycle Property	42
14.7 The Reverse Delete Algorithm	42
14.7.1 Runtime of Reverse Delete	42
14.7.2 Proof of Reverse Delete	42
15 The Clustering Problem	43
15.1 Maximum Spacing Clustering	43
15.2 Reverse-Delete Clustering Algorithm	43
15.2.1 Proof Of Reverse-Delete Clustering	43
16 The Set Cover Problem	44
16.1 The Greedy Set Cover Algorithm	45
16.2 Approximation Algorithms	45
16.3 Proof that Greedy Set Cover Almost Works	46
16.4 Running Time	47
16.5 The Hitting Set Problem	47
17 Matroids	47
17.1 The Hereditary Property	48
17.2 The Greediest Algorithm	48
17.2.1 The Running Time	48
17.2.2 Does It Work?	48
17.3 The Augmentation Property	49
17.4 What is a Matroid?	49
17.4.1 Examples of Matroids	49
17.5 Characterization of Matroids	50
IV Dynamic Programming	52

18 Fibonacci Numbers	53
18.1 Closed Form	53
18.2 Recursive Tree of Fibonacci Formula	53
18.2.1 Why is it so slow?	54
19 Dynamic Strategies	55
19.1 Top-Down (Memoization)	55
19.2 Bottom-Up	55
19.3 Top-down vs Bottom-Up	55
19.4 Difference Between Divide + Conquer and DP	55
19.5 Formula for Running Time of Dynamic Program	55
20 Interval Scheduling Revisited	55
20.1 Weighted Interval Selection	56
20.2 Dynamic Approach	56
20.2.1 Why This Ordering?	56
20.3 Weighted Interval Selection Algorithm	57
20.3.1 The Running Time	58
20.4 Key Points	58
21 Classifying Dynamic Problems	58
21.1 Structure of Sub-problems	58
21.1.1 One-Sided Interval	58
21.1.2 Box-Structure	58
21.1.3 Two-Sided Interval	59
21.1.4 Tree Structure	59
22 Knapsack Problem	59
22.1 Recursive Formula	59
22.2 Dynamic Program	59
22.3 Running Time	59
23 The Humpty-Dumpty Problem	60
23.1 Dynamic Approach	60
23.2 Running Time	60
24 Shortest Paths Revisited	60
24.1 Negative Cycles	61
24.2 The Dynamic Program	61
24.3 The Runtime	62
24.4 Finding negative cycles	62
25 Finishing Dynamic Programming	63
25.1 The Independent Set Problem	63
25.2 Independent Set On a Tree	63
25.3 The Running Time	64

26 Shortest Paths Problem Yet Again	64
26.1 Floyd-Warshall	64
26.2 Program	65
26.3 Running Time	65
V Network Flows	65
27 Maximum Flows in a Network	65
27.1 The Value of a Flow	65
27.2 Relation to Paths	66
27.3 Finding Maximum Flows	67
27.4 Ford-Fulkerson Algorithm	68
28 The Maxflow Mincut Theorem	68
28.1 When and How can the Algorithm Use an Arc?	68
28.2 The Residual Graph	68
28.2.1 Bottleneck Capacity	68
28.3 Ford-Fulkerson Revisited	69
28.4 s-t Cuts	69
28.4.1 The Cut Lemma	69
28.5 The Capacity of a Cut	70
28.6 Maxflow-Mincut Theorem	71
28.7 Running Time	71
29 The Vertex Matching and Cover Problem	72
29.1 Auxiliary Network	72
29.2 Running Time of Max Matching	72
29.3 Minimum Cut	73
29.3.1 Structure of Minimum Cut	73
29.4 Vertex Cover Problem	73
29.5 Supply/Demand	73
29.6 Multiple Sources and Sinks	74
29.7 Lower Bounds	74
30 Applications of Max-Flow	74
30.1 Flight Scheduling	74
30.2 Open Pit Mining	75
30.3 Foreground- Background Segmentation	76
31 Flow Decomposition and Fast Flow Algorithms	77
31.1 The Flow Decomposition Theorem	78
31.2 Finding the correct paths	79
31.3 Strongly Polynomial	81

VI Data Structures	81
32 Heaps	81
32.1 Insertion and UpHeap	81
32.2 Extract Min and DownHeap	82
32.3 Building a Heap	82
32.3.1 Building A Heap is Linear time	82
33 Applications of Heaps	83
33.1 HeapSort	83
33.2 Shortest Path Problem	83
33.3 Minimum Spanning Tree Problem	83
33.4 Data Compression Problem	84
34 Dictionaries and Hash Tables	84
34.1 Dictionaries	84
34.2 Hash Table	84
34.2.1 Hash Function	84
34.2.2 Universal Hash Functions	85
34.2.3 Random Hash Functions	85
34.2.4 Deterministic Hash Functions	85
34.2.5 Proof That It Works	86
34.3 String Searching	86
34.3.1 Searching for Multiple Strings	87
35 Binary Search Trees	87
35.1 BST Property	88
35.2 Operations and Observations	88
35.2.1 Search(r, k)	88
35.2.2 Insert(x)	88
35.2.3 Building a BST	88
35.2.4 Find-Min and Find-Max	88
35.2.5 Sorting Using a BST	89
35.2.6 Successor and Predecessor	89
35.2.7 Delete	90
35.3 Balanced BST's	90
36 Data Structures for Disjoint Sets	91
36.1 Set-Representatives + Storage	91
36.2 Operations	91
36.2.1 Make-Set(v)	91
36.2.2 Find-Set(v)	91
36.2.3 Merge(u, v)	92
36.3 Runtime	92
36.4 Union By Rank	92

36.5 Kruskals Revisited	92
36.6 Upper-Bound on Tree Depths	92
36.7 New Running Time for Kruskals	93
36.8 Path Compression	93
37 Data Structures for Range Queries	93
37.1 Range Min/Max Problem	93
37.1.1 Pre-Computation	94
37.1.2 Overlapping Ranges	94
37.2 Sparse Table	95
37.3 Range Sum Query Problem	96
37.4 Segment Trees	96
38 Learning	97
38.1 Multi-Armed Bandits	97
38.2 Expert Advisors	98
38.2.1 The Learning Model	98

1 Disclaimer

The material of this document was transcribed from Prof. Adrian Vetta's lecture recordings for COMP251 in the winter of 2018. Extra notes, clarifications or interpretations added by me may not be correct. All images are taken directly from Prof. Adrian Vetta's lecture slides. I claim no ownership over these images or the content taken from the slides.

2 Preliminaries

In this course an algorithm is considered **good** if it:

- Works
- Runs in polynomial time. Meaning it runs, in $O(n^k)$ time. Where n is (always) the size of the problem. (Number of elements in a list to be sorted etc.)
- Scales multiplicatively with computational power. (If your computer is twice as fast, the problem is solved at least twice as fast)

A bad algorithm is one that:

- Doesn't always work
- Runs in exponential time or greater. Meaning: $O(k^n)$ time.
- Does not scale well with computational power. (Your computer is twice as fast, but barely any performance boost).

Part I

Recursive Algorithms

I won't be going into detail on the specifics of things like how recursion works, MergeSort, BinarySearch, solving recurrences, Big O , etc. as it's considered prerequisite material. If you need some review, my COMP250 study guide is still publicly available.

3 Divide + Conquer Algorithms

Examples:

- MergeSort
- BinarySearch

3.1 MergeSort

The MergeSort algorithm involves splitting a list of n elements in half, sorting each half recursively, and merging the sorted lists back into one. It takes time $T(\frac{n}{2})$ to sort the list of half size, and time $O(n)$ to merge the list back together. So the recurrence relation for MergeSort is given by:

$$T(n) = 2T\left(\frac{n}{2}\right) + cn$$

where c is some constant.

Theorem 1. MergeSort runs in time $O(n \log(n))$.

Proof. Add **dummy numbers** (extra "padding" to the list), until n is a power of two. $n = 2^k$. We can do this because $O()$ gives an **upper bound**, and adding numbers will make our solution take longer than the real one. Doing this will make solving the recurrence easier.

Unwinding the formula:

$$\begin{aligned} T(n) &= 2(2(T\left(\frac{n}{4}\right) + c\frac{n}{2}) + cn \\ &= 2^2(T\left(\frac{n}{4}\right) + 2cn) \\ &= 2^3(T\left(\frac{n}{8}\right) + 3cn) \\ &= 2^4(T\left(\frac{n}{16}\right) + 4cn) \end{aligned}$$

Notice we have a pattern emerging.

$$= 2^k(T(1)) + kcn$$

Recall $2^k = n$, so $k = \log_2(n)$ and $T(1) = 1$ so:

$$= n + n \log_2(n)$$

Which is $O(n \log n)$. □

3.2 Binary Search

Binary search involves splitting your sorted list into two, and searching that half. So our recurrence is given by:

$$T(n) = T\left(\frac{n}{2}\right) + c$$

where c represents the constant work (comparisons, setting new bounds etc.)

Theorem 2. Binary Search is $O(\log_2(n))$.

Proof. Again we add dummy numbers so that n is a power of two. $n = 2^k$

We begin with our recurrence:

$$\begin{aligned} T(n) &= T\left(\frac{n}{2}\right) + c \\ &= T\left(\frac{n}{4}\right) + c + c \\ &= T\left(\frac{n}{8}\right) + c + c + c \\ &= T\left(\frac{n}{2^k}\right) + kc \\ &= T(1) + \log_2(n) \end{aligned}$$

since $k = \log_2(n)$ which is $O(\log_2(n))$. □

3.3 Run Time of Divide + Conquer in General

Divide and Conquer is a technique of solving problems that involves taking one large problem of size n , and breaking it down into a smaller problems of size $\frac{n}{b}$, and solving those problems recursively. They are then combined to produce a solution in time poly-time: $O(n^d)$.

So the run-time of a divide and conquer algorithm is:

$$T(n) = aT\left(\frac{n}{b}\right) + O(n^d)$$

In the case of MergeSort, $a = 2$, $b = 2$, $d = 1$.

In the case of BinarySearch, $a = 1$, $b = 2$, $d = 0$.

3.4 Aside on Recurrences: Domain Transformation

Note that the recurrence for MergeSort is really:

$$T'(n) \leq T'(\lfloor n/2 \rfloor) + T'(\lceil n/2 \rceil) + cn$$

Which we simplified by adding dummy entries. However, we can also say this:

$$T'(n) \leq 2T'\left(\frac{n}{2} + 1\right) + cn$$

But the $+1$ doesn't fit with our previous method.

We'll use **domain transformation** to solve this, starting with:

$$\begin{aligned} T(n) &= T'(n+2) \\ &\leq T'\left(\frac{n+2}{2} + 1\right) + c(n+2) \end{aligned}$$

plugging in our expression from above

$$\leq T'\left(\frac{n+2}{2} + 1\right) + c'(n)$$

absorbing the $+2$ into c .

$$= T'\left(\frac{n}{2} + 2\right) + c'(n)$$

simplifying the fraction.

$$= T\left(\frac{n}{2}\right) + c'n$$

from our domain transformation at the beginning. Solving this the usual way, we get:

$$T(n) = O(n \log(n))$$

But again from our domain transformation:

$$T(n) = T'(n+2)$$

, so

$$T'(n) = T(n-2) = O(n \log(n))$$

So we've shown that $T'(n)$ has the same upper bound as $T(n)$.

4 Master Theorem

Theorem 3. If $T(n) = aT(n/b) + O(n^d)$ for constants $a > 0$, $b > 1$, $d \geq 0$, then:

$$\begin{cases} O(n^d) & \text{if } a < b^d \\ O(n^d \log(n)) & \text{if } a = b^d \\ O(n^{\log_b(a)}) & \text{if } a > b^d \end{cases}$$

These cases are just a few that occur often in practice when dealing with divide + conquer algorithms.

Proof. First we'll need two things. One is the geometric series, and the other is a law of logarithms. Professor Vetta proved them in class, and honestly I doubt you'd be asked to prove them on an exam, but it's good proof practice to go through them so I'll do it here.

$$\sum_{k=0}^l x^k = \frac{1 - x^{l+1}}{1 - x}$$

Proof:

Starting with:

$$(1 - x) \sum_{k=0}^l x^k$$

We can expand it out:

$$= \sum_{k=0}^l x^k - \sum_{k=0}^l x^{k+1}$$

Simplifying the sigma notation:

$$= \sum_{k=0}^l x^k - \sum_{k=1}^{l+1} x^k$$

All terms will cancel except:

$$= x^0 - x^{l+1} = 1 - x^{l+1}$$

Divide through by $1 - x$

$$= \frac{1 - x^{l+1}}{1 - x}$$

Our second fact to derive is this law of logs:

$$x^{\log_b(y)} = y^{\log_b(x)}$$

Using the power rule of logarithms:

$$\log_b(x)\log_b(y) = \log_b(y^{\log_b(x)})$$

similarly,

$$\log_b(x)\log_b(y) = \log_b(x^{\log_b(y)})$$

so,

$$\log_b(x^{\log_b(y)}) = \log_b(y^{\log_b(x)})$$

Now we're ready for the proof.

Assume n is a power of b , and split up the problem into all it's chunks.

$$T(n) = n^d + a\left(\frac{n}{b}\right)^d + a^2\left(\frac{n}{b^2}\right)^d + \dots + a^l\left(\frac{n}{b^l}\right)^d$$

(this is just if you'd "unwound" the whole recursion down to it's simplest form like we did in the MergeSort/Binary Search proofs.)

Each term is the amount of work it will take at each level of the recursion.

Notice you can factor out:

$$\begin{aligned} &= n^d \left(1 + a\left(\frac{1}{b}\right)^d + a^2\left(\frac{1}{b^2}\right)^d + \dots + a^l\left(\frac{1}{b^l}\right)^d\right) \\ &= n^d \left(1 + \left(\frac{a}{b}\right)^d + \left(\frac{a}{b^d}\right)^2 + \dots + \left(\frac{a}{b^d}\right)^l\right) \end{aligned}$$

That looks like a geometric series! So let's look at the cases:

Case 1: $a < b^d$

Applying the geometric series formula:

$$\begin{aligned} &= n^d \sum_{k=0}^l \left(\frac{a}{b^d}\right)^k \\ &= n^d \frac{1 - \left(\frac{a}{b^d}\right)^{l+1}}{1 - \frac{a}{b^d}} \end{aligned}$$

we can remove the $\frac{a}{b^d}^{l+1}$ term with this inequality (since the term doesn't depend on n):

$$\leq n^d \frac{1}{1 - \frac{a}{b^d}}$$

which is $O(n^d)$.

Case 2: $a = b^d$

Since $\frac{a}{b^d} = 1$:

$$= n^d(1 + 1 + 1 + \dots + 1)$$

There are $l + 1$ terms, but we said n was a power of b , ($n = b^l$) so, $l = \log_b(n)$, thus:

$$= n^d(\log_b(n) + 1)$$

which is $O(n^d \log_b(n))$

Case 3: $a > b^d$

Again from geometric series, and multiplying through by -1:

$$n^d \frac{\left(\frac{a}{b^d}\right)^{l+1} - 1}{\frac{a}{b^d} - 1}$$

Again this inequality holds:

$$\leq n^d \frac{\left(\frac{a}{b^d}\right)^{l+1}}{\frac{a}{b^d} - 1}$$

Which is $O(n^d \left(\frac{a}{b^d}\right)^l)$ which we can simplify:

$$\left(\frac{n}{b^l}\right)^d a^l$$

but $n = b^l$, so:

$$\begin{aligned} &= (1)a^l \\ &= a^{\log_b(n)} \end{aligned}$$

now by our second fact:

$$= n^{\log_b(a)}$$

which is $O(n^{\log_b(a)})$

□

It's **much** more important to understand the proof than it is to memorize the theorem.

4.1 Tree Method to Prove Master Theorem

A more intuitive way to think of the proof is with a *Recursion Tree*.

The root node of the tree has label n , and each node has a children (except the leaves). a is called the *branching factor*. Each child is labelled $\frac{n}{b^d}$ where d is the depth. The labels represent the size of the sub problems.

The number of nodes at each level is a^d .

Case 1 is when the root level "dominates" all other levels, so the running time is just $O(f(n))$ where $f(n)$ is the amount of work at the root level.

Case 2 is when all levels are roughly the same weight. So the total running time is just $O(f(n)l)$ where l is the number of levels.

Case 3 is when the leaves dominate, so the running time is $O(a^l)$ since the leaves each take time $O(1)$, and there are a^l of them.

5 Multiplication

5.1 Grade School Multiplication

This takes n^2 multiplications when you multiply two n-digit numbers. so the runtime is $\Omega(n^2)$

5.2 Russian Peasant Multiplication

Super weird looking algorithm but it works!

```
Mult(x,y){
    if x = 1 then output y
    if x is odd then output y + Mult(floor(x/2),2y)
    if x is even then output Mult(x/2, 2y)
}
```

This actually comes from if you take the binary representation of x : say $x = 46_{10}$ then $x = 101110_2$. The bits that are 1's will have the y added step, and the zero bits will just have the doubling step. Weird right?

Notice that this means the number of steps is just the number of bits in x . The number of digits in the result will be at most $2n$, so if we need to then add these, we add at most n numbers of $2n$ digits so takes time $O(n^2)$

5.3 Divide + Conquer Multiplication

Notice that a number x can be written as:

$$x = x_n x_{n-1} \dots x_{\frac{n}{2}+1} x_{\frac{n}{2}} \dots x_2 x_1$$

where the x_i are the digits.

Then we have:

$$x = 10^{\frac{n}{2}} x_L + x_R$$

where n is the number of digits, x_L is the first $\frac{n}{2}$ digits, and x_R is the last $\frac{n}{2}$

So by expanding:

$$xy = (10^n x_L y_R + 10^{\frac{n}{2}}(x_L y_R + x_R y_L) + x_R y_R$$

Notice that this now involves four products of $\frac{n}{2}$ digit numbers. So the recursion is:

$$T(n) = 4T\left(\frac{n}{2}\right) + O(n)$$

We have $a = 4, b = 2, d = 1$, which is case 3 of the master theorem.

Which means the running time is:

$$O(n^{\log_2(4)})$$

which simplifies to:

$$O(n^2)$$

Thanks to Gauss, we can actually use this fact:

$$x_L y_R + x_R y_L = x_R y_R + x_L y_L - (x_R - x_L)(y_R - y_L)$$

which is actually only 3 unique products. (adding is cheap)

So our new running time is:

$$T(n) = 3T\left(\frac{n}{2}\right) + O(n)$$

which is case 3 of the master theorem, so

$$\begin{aligned} & O(n^{\log_2(3)}) \\ &= O(n^{1.59}) \end{aligned}$$

5.4 Fast Fourier Transforms

These are $O(n \log(n))$ for multiplying n-bit numbers. They'll be studied more in-depth at the end of the course (time-permitting).

5.5 Multiplying Matrices

There are n multiplications to calculate each entry of the result matrix, and there are n^2 entries, so $O(n^3)$

Using divide + conquer, divide into 4 sub-matrices:

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1n} \\ x_{21} & x_{22} & x_{23} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ x_{d1} & x_{d2} & x_{d3} & \dots & x_{dn} \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

So if we let:

$$x = \begin{bmatrix} A & B \\ C & D \end{bmatrix} y = \begin{bmatrix} E & F \\ G & H \end{bmatrix}$$

then:

$$XY = \begin{bmatrix} AE + BG & AF + BH \\ CE + DG & CF + DH \end{bmatrix}$$

So multiplying involves eight products with $\frac{n}{2} \times \frac{n}{2}$ and the recurrence is:

$$T(n) = 8T\left(\frac{n}{2}\right) + O(n^2)$$

which is Case 3 of the master theorem, so runtime is $O(n^{\log_2 8})$ which is $O(n^3)$, no improvement.

There actually is a trick to do better.

Claim:

$$XY = \begin{bmatrix} AE + BG & AF + BH \\ CE + DG & CF + DH \end{bmatrix}$$

is the same as:

$$\begin{bmatrix} S_1 + S_2 - S_4 + S_6 & S_4 + S_5 \\ S_6 + S_7 & S_2 - S_3 + S_5 - S_7 \end{bmatrix}$$

where:

$$S_1 = (B - D)(G + H)$$

$$S_2 = (A + D)(E + H)$$

$$S_3 = (A - C)(E + F)$$

$$S_4 = (A + B)H$$

$$S_5 = A(F - H)$$

$$S_6 = D(G - E)$$

$$S_7 = (C + D)E$$

which is only 7 products! (The additions are negligible)

So we have:

$$T(n) = 7T\left(\frac{n}{2}\right) + O(n^2)$$

Which is Case 3 of the master theorem, so $O(n^{\log_2(7)})$ which is $O(n^{2.81})$

5.6 Fast Exponentiation

Method of taking exponents in a fast way, since doing:

$$x * x * x * x \dots * x$$

is super slow.

```
FastExt(x,n){
    if n=1 output x
    else
        if n is even output FastExp(x, floor(n/2))^2
        if n is odd output FastExp(x, floor(n/2))^2*x
}
```

So our recurrence looks like:

$$T(n) = T\left(\text{floor}\left(\frac{n}{2}\right)\right) + O(1)$$

(since we're halving the problem, and doing some constant work at each step)

This is Case 2 of the Master Theorem, so the runtime is $O(\log_2 n)$

6 The Median Problem

6.1 The Selection Problem

Want to find the k th smallest number in a set S .

Select(S, k)

If $|S| = 1$ then output x_1 .

Else:

Set S_L = all numbers less than x_1

Set $S_R = \text{all numbers greater than } x_1$

If $|S_L| = k - 1$ then output x_1 (since if you have $k-1$ things smaller than x_1 , that can only mean x_1 is the k th smallest element)

If $|S_L| > k - 1$ then output $\text{Select}(S_L, k)$ (since that means the k th smallest element must be within that set)

If $|S_L| < k - 1$ then output $\text{Select}(S_R, k-1-|S_L|)(-1)$ since you know its not x_1 , and $-|S_L|$ since you know its not in any of those, so you want the $k-1-|S_L|$ -th element of S_R .)

The runtime of this algorithm is almost entirely dependent on the choices of pivots, since if you get a "bad" pivot every time, then you would recurse on a set of size $n-1$.

$$T(n) = (n-1) + T(n-1)$$

$$= \frac{1}{2}(n(n+1))$$

which is $O(n^2)$.

We could instead choose our pivot randomly.

The pivot would separate the list into sizes from $\frac{n}{4}$ to $\frac{3n}{4}$ with probability $\frac{1}{2}$, and so the pivot would be good half the time. So the expected running time is:

$$\begin{aligned} T(n) &\leq \frac{1}{2}T\left(\frac{3n}{4}\right) + \frac{1}{2}T(n) + O(n) \\ \frac{1}{2}T(n) &\leq \frac{1}{2}T\left(\frac{3n}{4}\right) + O(n) \\ T(n) &\leq T\left(\frac{3n}{4}\right) + O(n) \end{aligned}$$

which satisfies Case 1 of the master theorem which is $O(n)$.

But what if we want to be **certain** that the worst case will never happen?

6.2 Median of Medians

Divide the set S into groups of size 5. Sort each group and find the median of each group. If you were to find the median of these medians, there would always be less than $\frac{7}{10}n$ elements in your two groups, which is pretty good. The reason this comes up is:

There's $\frac{n}{5}$ groups overall. Imagine the everything was sorted. Each group of 5 is sorted, and the groups are sorted by their medians. So there's $\leq \frac{n}{5} * \frac{1}{2} = \frac{n}{10}$ groups to the left of the

median of medians. There's 3 elements above than the median in its own group, so there's $\leq \frac{3n}{10}$ elements smaller (to the top left) than the median, which means there's $\leq \frac{7n}{10}$ elements larger (to the bottom right) than the median.

So the max size of the sets is $\frac{7n}{10}$

Finding the median of the medians is done recursively, by partitioning into 5 groups, and putting a recursive call on finding the pivot.

So the recursive formula is:

$$T(n) \leq T\left(\frac{7n}{10}\right) + T\left(\frac{n}{5}\right) + O(n)$$

Notice the Master Theorem doesn't apply here, instead we need to use the recursion tree method.

First our problem of size n is broken into two problems, one of size $\frac{7n}{10}$ and the other size $\frac{2n}{10}$. Continuing down recursively, we actually get one side of the tree ending before the other. Namely, the $\frac{7n}{10}$ side will reach the leaves later than the $\frac{3n}{10}$ side.

However, up until the point that this end is reached, we're doing $(\frac{9}{10})^l n$ work at each level. Beyond this, the work needed at each level only decreases, so it's $\leq (\frac{9}{10})^l n$. These terms are geometrically decreasing, so the first term dominates, and we get $O(n)$.

7 Finding the Closest Pair of Points in the Plane

How fast can we solve this?

7.1 Exhaustive Search

Calculate the distance between every pair of points, choose the shortest pairwise distance. $O(n^2)$. Is there a faster algorithm?

In one-Dimension, notice that the closest pair of points needs to be next to each other on the line. So we only need to find how far each **pair** is. ($n - 1$ distances to calculate).

7.2 2-D case

Simply taking the closest in their x-coordinate (or y coordinate) doesn't work since they could be close in x but very far in y.

A divide + conquer approach is to separate the points into two groups of size $\frac{n}{2}$, so we

want our dividing line to pass through the median x-coordinate.

We can now recursively search for the closest pairs in each group.

But what if the closest pair is **between** the two groups?

So we have to check to see if there's a better solution with an endpoint in each group. How can we do this efficiently? (This is the bottleneck step).

7.3 Widening the Bottleneck

Notice that by solving the subproblems recursively we can find the smallest distance between two points in both the left and right subproblems call this δ . So we know that if a better solution exists, it will be within δ from the dividing line.

This seems much better! But what if all the points are within δ of the dividing line? Well then this doesn't help much.

There's actually a trick we can do.

We can break up the area into squares of size $\frac{\delta}{2}$, and no two points will lie in the same square. This is because if two points are in the same square, then there are on the same side of the dividing line. These points are within $\delta\sqrt{2}/2$ (by construction of the boxes) from each other, but this is $< \delta$, so this contradicts the minimality of δ .

We can now use this fact to derive another fact:

Suppose there's a point on either side of the dividing line with distance less than δ . We can prove that there will be at most 10 points between them in the y-ordering. (Within the area filled with boxes).

Proof

Since the squares are of size $\frac{\delta}{2}$, then the two points are either on the same row, or one is within two rows above the other. (or else it would be further than δ) Now, since there can only be one point per-box, there's at most 10 points between them. (count the boxes for yourself!)

Now recall the 1-D case, we can now just look at every pairwise distance on a group where the points are at most 11 apart (rather than the ones that are next to each other as before). So you need to find the distance between a given point, and the next 11 distances.

So at most $11n$ distances to calculate.

7.4 The Finished Algorithm

- Find the point with the median x-coordinate
- Partition using this point
- Recursively find the closest pair of points in each half
- Find the closest pair within the small range given by δ , by checking the nearest 11 points (in the y-ordering) for each point.
- Among the three pairs found, (left, right, crossing) output the closest pair.

7.5 The Runtime (Enhanced)

Two subproblems of size $\frac{n}{2}$, and the work at each level is: finding the median $O(n)$, partitioning $O(n)$, making the smaller group (within δ of dividing line) $O(n)$, applying the 1-D algorithm $O(n)$. So our recurrence looks like:

$$T(n) = 2T\left(\frac{n}{2}\right) + O(n)$$

which is case 2 of the master theorem, so $n \log(n)$.

Part II

Graph Algorithms

For a review of basic graph terminology, see my COMP250 study guide. And I'm going to assume you took MATH240 and know the basics of graph theory from there. Lecture 7 had a review of these basics, but I won't include them here. I will only go over the theorems that were proved.

8 Theorems About Undirected Graphs

8.1 Handshaking Lemma

Theorem 4. In an undirected graph, there are an even number of vertices with odd degree.

Proof. We start with:

$$2 | E | = \sum_{v \in V} \deg(v)$$

Since each edge is double counted when summing the degrees. (Each edge (u, v) contributes 1 degree to u and 1 degree to v)

This is the same as:

$$= \sum_{v \in Odd} \deg(v) + \sum_{v \in Even} \deg(v)$$

(the sum of the vertices of even degree plus the odd degree ones)

Rearranging we get:

$$\sum_{v \in Odd} = 2 |E| - \sum_{v \in Even} \deg(v)$$

Now we know at the sum of the even degrees is even, and we know that $2x$ is even for any x . So the right hand side is always even. Therefore the left hand side must be even. But for the sum of odd numbers to be even, there must be an even number of odd terms.

□

8.2 Leaf Existence

Theorem 5. Lemma: A tree T with $n \geq 2$ vertices has at least one leaf.

Proof. A tree is connected, which means there's no vertices with degree 0. A leaf is a vertex with degree 1, so to get a contradiction, assume every vertex has degree ≥ 2 .

Take the longest path $P = v_1, v_2, v_3 \dots v_{l-1}, v_l$

But every vertex has degree greater than 2, so v_l has a neighbour $x \neq v_{l-1}$ so v_l forms an edge with something in the path which would create a cycle and thus be a contradiction (since all trees have no cycles). If the neighbour was not on the path, then P was not the longest path. □

8.3 Number of edges in a Tree

Theorem 6. A tree with n vertices has $n - 1$ edges

Proof. By induction:

Base Case:

A tree on one vertex has zero edges

linebreak Induction Step:

Assume that any tree on $n - 1$ vertices has $n - 2$ edges. Take a tree with $n \geq 2$ vertices. By the previous lemma, there exists a leaf vertex v . Let $T' = T - v$. Then T' is a tree on $n - 1$ vertices, which has $n - 2$ edges by the induction hypothesis. Adding back v , we get that T is a tree on $n - 1$ vertices with $n - 2$ vertices. □

8.4 Halls Theorem

Theorem 7. A bipartite graph, with $|X| = |Y|$ contains a perfect matching $\Leftrightarrow \forall S \subseteq X, |\Gamma(S)| \geq |S|$

Proof. (\Rightarrow)

If there is a set $S \subseteq X$ with $|\Gamma(S)| < |S|$, then the graph cannot have a perfect matching, since there would not be enough things in the neighbourhood for the things in S to match to.

(\Leftarrow)

Take a maximum cardinality matching M in the graph. If M is perfect we're done, if not, then there exists a vertex x_0 who is not matched in X . If Halls condition holds, then x_0 has a neighbour y_0 . Suppose y_0 is matched to x_1 . Again if Halls condition holds, then x_0, x_1 have another neighbour say y_1 .

We now repeat this process until eventually it terminates (it will since we have a finite number of vertices). It will terminate when we reach x_k who is unmatched.

We now create an m-alternating path from y_k to x_0 . This path is m-augmenting, so we augment, and receive a larger matching. Contradiction, M was not maximal.

□

9 Breadth First Search

9.1 Generic Search Algorithm

```
Put root into bag
  while bag not-empty
    remove v from the bag
      if v is unmarked
        mark v
        for each arc(v,w)
          put w into the bag
```

A vertex is discovered when it is marked. Notice that there can be multiple copies of a vertex in the bag, and that this actually won't affect our performance.

9.1.1 Revised Generic Search Algorithm

Instead of adding vertices, we'll add arcs.

```
Put (*,r) into a bag
  while bag not-empty
    remove (u,v) from the bag
```

```

if v is unmarked
  mark v
  set p(v) to u //keep track of "predecessor" of v
  for each arc(v,w)
    put (v,w) into the bag

```

Keeping track of the predecessor will be useful later.

9.1.2 The Running Time

We look at each arc out of v only once, when v is first marked. The arc is then added to the bag once, and removed once. So, we get a runtime proportional to the number of arcs.

$$\Rightarrow O(m)$$

9.1.3 Validity

Theorem 8. Let G be a connected, undirected graph. Then the search algorithm finds every vertex in G .

Proof. We need to show that every vertex v is marked by the algorithm. We will use induction on the length of the smallest path from the vertex to the root.

Base Case:

$k = 0$ then v is the root, and only the root exists, so trivially true.

Induction Step:

Assume true for a path of length $k - 1$ from the root. Now assume there is a path P with k edges from v to r . So let

$$P = \{v = v_k, v_{k-1}, \dots, v_1, v_0 = r\}$$

Then there is a path:

$$Q = \{u = v_{k-1}, \dots, v_1, v_0 = r\}$$

So by the induction hypothesis, u is marked. Then after we mark u , all edges incident to u would have been added, so (u, v) would have been added. And so later, (u, v) would be removed and v would be marked. \square

We can prove that for directed graphs, every vertex that has a directed path from r is marked in the same way.

9.2 Search Trees

Theorem 9. The predecessor edges made by the search algorithm on a connected, undirected graph G is a tree rooted at r .

Proof. By induction on the number of marked vertices, k .

Base Case: $k = 1$

Induction Step:

Assume true for the first $k - 1$ vertices. Let v be the k th vertex to be marked. Assume v was marked when we removed the edge (u, v) . This means that u is the predecessor of v . But (u, v) was added to the bag when we marked u , so u must be in the set of the first $k - 1$ vertices to be marked. Thus, by the induction hypothesis, when we add the edge $(p(v), v) = (u, v)$, we are adding a leaf, so the new graph formed is still a tree. \square

9.3 Choices of Bags

We can use a Queue to get BFS, if we use a Stack we get DFS, if we use a Priority Queue, we get minimum spanning tree.

9.4 BFS Trees

The edges are added to the queue in order of their distance from r . The vertices are marked in order of their distance from r .

Theorem 10. For any vertex v , the path from v to r given by the search tree T of predecessor edges is a shortest path.

Proof. Left as exercise. \square

9.4.1 Structure

The structure of these trees can be broken down into "layers", where each layer is the set of vertices at a given distance from the root.

Any vertex $v \in S_l$ is at distance l from r in T , and the same is true in the whole graph G . This implies that for every edge in the graph that is not in the tree (u, v) , u and v are either in the same layer or in adjacent layers. If this was not the case, say u was in S_3 and v was in S_6 , then we could get from the root to v in less than 6 steps.

9.4.2 BFS on Bipartite Graphs

Theorem 11. A graph G is bipartite \Leftrightarrow it contains no odd length cycles.

Proof. \Rightarrow

Assume G contains an odd length cycle C .

$$C = \{v_0, v_1, \dots, v_{2k}\}$$

Without loss of generality we can assume $v_0 \in Y$, therefore $v_1 \in X$, and so on. We eventually get down to $v_{2k} \in X$ but it is a cycle so v_{2k} forms an edge with v_0 , and we said $v_0 \in Y$, which means $v_{2k} \in Y$, it can't be both in X and in Y , contradiction.

\Leftarrow

Assume G has no odd length cycles. Choose a root vertex r , and run BFS. Let X be the set of all odd layers of the BFS tree. Let Y be the set of all even layers of the BFS tree. Since every edge in the graph goes to either adjacent layers or the same layer, we know that if we have no edges in the same layer, then we'll have that every edge goes from X to Y .

Assume there's a non-tree edge (u, v) with u and v in the same layer. Let z be the closest common ancestor of u and v in the search tree. Let P be the path from u to z in the tree, and let Q be the path from v to z in the tree. The length of P is the same as Q since u and v are in the same layer. But then the cycle

$$C = P \cup Q \cup (u, v)$$

has an odd number of edges. So (u, v) cannot exist. \square

10 Depth First Search

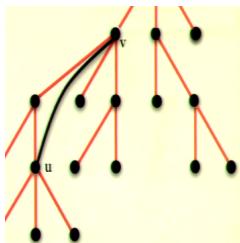
We use the generic search algorithm using a stack.

10.1 DFS Trees

The DFS tree is much different than the BFS tree. DFS partitions the edges of an undirected graph into two types:

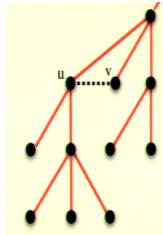
Tree Edges: Predecessor edges in the DFS tree at T

Back Edges: Edges where one endpoint is the ancestor of the other endpoint in T .



Here (u, v) is a back edge.

We cannot have Cross Edges: Where neither endpoint is an ancestor of the other.



10.2 Recursive DFS

We can also do DFS recursively:

```
RecursiveDFS(r)
    mark r
    for each edge (r, v)
```

```

if v is unmarked
  set p(v) = r
  RecursiveDFS(v)

```

10.3 Ancestral Edges

Theorem 12. Let T be a DFS tree in an undirected graph G . Then for every edge (u, v) either u is an ancestor of v in T or v is an ancestor of u .

Proof. Wlog assume u is marked before v . At the time u is marked, the algorithm will recurse on each arc incident to u .

Case 1: v is unmarked when the $\text{RecursiveDFS}(u)$ examines (u, v) .

Then the parent of v is then u , and so (u, v) is an ancestral tree edge.

Case 2: v is already marked. But v was marked after u , so it was marked during $\text{RecursiveDFS}(u)$. So we have a series of vertices

$$\{u = w_0, w_1 \dots w_{l-1}, w_l = v\}$$

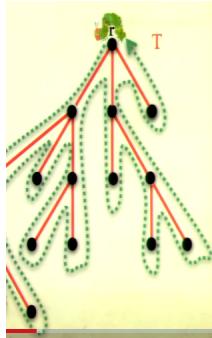
where $p(w_k) = w_{k-1}$ (the parent of each vertex is the previous vertex). This means that u is an ancestor of v , so (u, v) is a back edge.

□

Corollary Every non-tree edge is a back edge.

10.4 Previsit and Postvisit

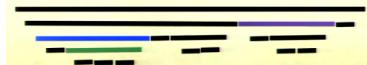
The way DFS explores the vertices of a graph is given by this picture:



We can add a "clock" that will keep track of the order in which the vertices were visited.

$\text{Pre}(v)$ is the time at which we arrive at a subtree rooted at v . $\text{Post}(v)$ is the time at which we leave a subtree rooted at v .

So we can represent each vertex by an interval of time. If we take the interval for every vertex, we get what's called a **Laminar Family**. Meaning, every interval is either completely disjoint, or completely overlapping.



If we draw an edge between each interval and the smallest interval that contains it, we actually build up the DFS tree again!

10.5 Directed Graph BFS Tree Structure

Now we can have four types of edges:

Tree arcs: Same as before

Forward Arcs: Arcs (u, v) where u is an ancestor of v

Backward Arcs: Arcs (u, v) where v is an ancestor of u

Cross Arcs: Non-Ancestral arcs (u, v) where u is marked after v .

Note that for cross arcs, the other way around is not possible because after visiting v , we must visit all descendants of it, before moving back up the tree and going down the other branch containing u .

We still have intervals. In a tree arc (u, v) , the interval of v is contained in the interval of u . In a forward arc, the same is true. In a backward arc however, the interval of u is contained in the interval of v . In cross arcs, the intervals of u and v are disjoint.

So we have this list of properties:

For tree arcs:

$$\text{post}(v) < \text{post}(u)$$

For forward arcs:

$$\text{post}(v) < \text{post}(u)$$

For backward arcs:

$$\text{post}(u) < \text{post}(v)$$

For cross arcs:

$$\text{post}(u) < \text{post}(v)$$

So the only different one is for backward arcs.

10.6 Example: Directed Acyclic Graphs

How can we determine if a graph is acyclic?

Theorem 13. A directed graph G is acyclic \Leftrightarrow DFS produces no backward arcs.

Proof. \Rightarrow

Suppose DFS gives a backward arc (u, v) . By definition, then u is a descendant of v in the DFS tree T . Then there exists a path:

$$P = \{v = v_0, v_1, \dots, v_k = u\}$$

which means $P \cup (u, v)$ is a directed cycle in G .

\Leftarrow

Assume DFS gives no backward arcs. Suppose there's a directed cycle:

$$C = \{v_0, v_1, \dots, v_k, v_0\}$$

Since there's no backward arcs we have that:

$$\text{post}(v_0) > \text{post}(v_1) > \dots > \text{post}(v_k) > \text{post}(v_0)$$

But $\text{post}(v_0)$ can't be greater than itself.

□

Corollary There is a linear time algorithm to test whether or not a graph is acyclic. Just run DFS and check if any arc is a backward arc.

10.7 Example: Topological Ordering

A topological ordering is when the vertices of a graph can be horizontally ordered such that every arc is from right to left.

Theorem 14. A directed graph G has a topological ordering \Leftrightarrow DFS produces no backward arcs.

Proof. \Rightarrow If DFS produces a backward arc then G contains a cycle C . Let the cycle:

$$C = \{v_0, v_1, \dots, v_k, v_0\}$$

where, wlog, v_0 is the leftmost vertex of the cycle in the order. But then v_0, v_1 goes from left to right, which is not allowed.

\Leftarrow

Assume DFS gives no backward arcs. Then for every arc (u, v) we have:

$$\text{post}(u) > \text{post}(v)$$

so simply order the vertices by their post numbers.

□

Part III

Greedy Algorithms

11 Scheduling

11.1 Task Scheduling

A firm can process 1 task a time. The job of customer i takes t_i time. We want to minimize the sum of the waiting times. Any job cannot be started until the previous one is finished.

First, we sort the jobs by length, shortest to longest. Simply schedule them in that order.

We must now prove this works.

Theorem 15. The greedy algorithm outputs an optimal schedule.

Proof. We will use an exchange argument.

Let the greedy algorithm schedule in the order: $\{1, 2, \dots, n\}$

Assume there's a better schedule S . Then there must be a pair of jobs i and j such that:

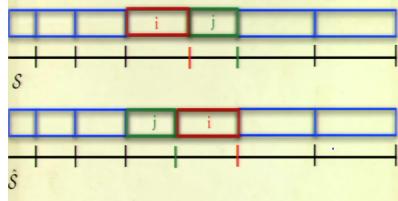
Job i is scheduled immediately before job j by schedule S

Job i is longer than job j

If we don't have this property, then it's sorted. The waiting time of job i is currently when job i finishes, and the waiting time of job j is when job j finishes.

Swap jobs i and j . Everything else stays the same. Specifically, the waiting time of every unchanged job stays the same.

Now the new waiting time of job j is better than both of the old ones, and the waiting time of job i is the same as the waiting time of the old job j .



So this configuration is better, and that contradicts the assumption that S was an optimal schedule. \square

11.1.1 Running Time

All we did was sort, so it's $O(n \log(n))$

11.2 Class Scheduling

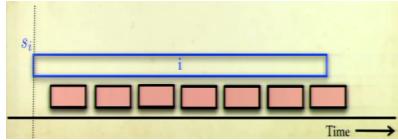
There is one classroom. There's a set $I = \{1, 2, \dots, n\}$ of classes that want to use the room. Class i has a start time s_i and a finish time f_i . The goal is to book as many classes as possible.

This is also known as the interval selection problem.

11.2.1 First Start

Select the class that starts earliest, iterate on the remaining classes that do not conflict with this one.

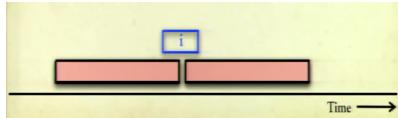
This doesn't work because the first class might also be the longest.



11.2.2 Shortest-Duration

Select the shortest class first, then iterate.

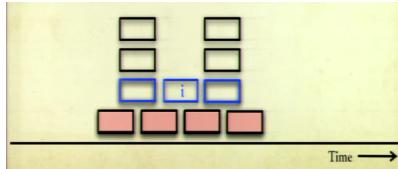
Doesn't work because might just be in an unlucky position.



11.2.3 Minimum Conflict

Select the class that conflicts with the fewest number of classes first. Then iterate.

Doesn't work because we might get a configuration like this:



Here the optimal solution has 4 classes, but we chose the configuration that has only 3, we chose i since i conflicts with only 2, whereas the red ones all conflict with 3 or 4. Next time we iterate, we just have two stacks of 3, so we can choose any from the left or right, and we're stuck with having at most 3 classes in our solution.

11.2.4 Last Start

Select the class that starts last, and iterate on the classes that do not conflict with this selection.

This one works!

It is symmetric to selecting the class that finishes first, then iterating on the classes that don't conflict with this selection.

Here's the pseudocode:

```
FirstFinish(I)
Let class1 be the class with the earliest finish time
Let X be the set of classes that clash with class1
output {1} U FirstFinish(I\X)
```

Lemma There is some optimal solution that selects Class 1.

Proof

Recall the classes are indexed such that $f_1 \leq f_2 \leq \dots \leq f_n$.

Take an optimal schedule S and assume Class 1 is not in it. Let i be the lowest index class in S .

We claim we can replace i with 1. We know that f_i is before s_j for any $j \in S$. But f_1 is before s_j , so we know Class 1 doesn't conflict with any class in $S - \{j\}$.

So this ordering is at least as good as before, and contains class 1.

Theorem 16. The first-finish algorithm outputs an optimal schedule.

Proof. By induction on the cardinality of the optimal solution $|opt(I)|$

Base Case:

Let the solution have size 1, then this is trivially the optimal solution.

Induction Step:

Assume true for size k . Let the optimal solution have size $k + 1$. First finish outputs $\{1\} \cup FirstFinish(I - X)$. But by the lemma, 1 is part of some optimal solution, S^* .

This means $S^* - \{1\}$ is an optimal solution with size k . So by the induction hypothesis, we have an optimal solution. \square

There are at most n iterations of the algorithm, and it takes n time to find the class that finishes earliest in each iteration. So $O(n^2)$, but you could get it down to $O(n \log n)$ if you're careful about implementation.

12 The Shortest Path Problem

If every arc a has a length associated to it (a weight), l_a , then the length of a path P is:

$$l(P) = \sum_{a \in P} l_a$$

How do we then find the shortest path from s to every other vertex in the graph.

It turns out, we can find all the shortest paths in one go!

12.1 Dijkstra's Shortest Path Algorithm

We'll need to keep track of a few things. We need to keep track of the "tentative" distance from the root to our current vertex. We also need to keep track of which nodes we're done with.

- 1.) Initially, assign the first vertex's tentative distance to 0, set everything else to ∞ .
- 2.) For each unmarked arc coming out of the current vertex, calculate the distance from the current vertex to the next, and add it to the distance from the root to the current. (To get the total). If this value is less than the current tentative distance of that vertex, replace it.
- 3.) Once you've calculated all neighbors of the current vertex, mark it as visited.
- 4.) Set the current vertex to be the one with the lowest tentative distance.

12.1.1 Special Case, All Arcs Have Distance 1

In this case, we actually get exactly Breadth First Search! Try it out ;).

12.1.2 Shortest Path Graph

Let S^k be the set of vertices in S at the end of the k th iteration, where S is the set of vertices we're "done with".

Let T^k be the set of arcs in T at the end of the k th iteration, where T is the set of arcs we fix to be in our final result.

Notice that all arcs in T^k are between vertices in S^k because when we add a vertex to S , we add the arc between it and its predecessor (another vertex in S), to T . This means that $G^k = (S^k, T^k)$ is a directed graph.

Finally, G^n is the final output of the algorithm.

12.1.3 Shortest Path Tree

Theorem 17. The Graph G^k is a directed tree rooted as s .

Proof. Base Case: $k=1$

S^1 only contains s , and T^1 is empty. One vertex is a trivial tree.

Induction Step:

Assume true for G^{k-1} . Let v_k be the vertex added to S at the k th iteration. So $S^k = S^{k-1} \cup \{v_k\}$. This means that $T^k = T^{k-1} \cup (\text{pred}(v_k), v_k)$. So v_k has in-degree 1, and out-degree 0, which means v_k is a leaf. So G^k is still a directed tree rooted at s . \square

Theorem 18. G^k gives the true shortest path distances from s to every vertex in S^k .

Proof. Base Case: $k = 1$

Trivially true. The label, $d(s)$ is 0, which is the shortest path distance, $d^*(s)$

Induction Step:

Assume true for G^{k-1} . That is, $d^{k-1}(v) = d^*(v) \forall v \in S^{k-1}$

Let v_k be the vertex added to S in the k th iteration. Take the shortest path P from s to v_k that uses as many arcs in common to G^k as possible. Now assume for a contradiction that $d^k(v_k) < d^*(v_k)$. (basically this path P follows the tree, then eventually jumps out to get to v_k and we're assuming this is faster than just following the tree).

Let x be the last vertex of G^{k-1} in P . Let $y \notin S^k$ be the vertex after x in P . If this y doesn't exist, (there's no vertex after x) then we're done, since that means $P \subseteq G^k$. Assume it does exist.

Since y is on the shortest path from s to v_k and the arc-lengths are non-negative, each sub-path is also a shortest path, then the path from s to y is shorter than the path from s to v_k .

Since we're assuming P is the optimal path:

$$d^*(y) \leq d^*(v_k) < d^k(v_k)$$

But since $x \in S^{k-1}$ we have:

$$d^k(y) \leq d^{k-1}(x) + l(x, y)$$

(basically meaning that the distance from s to y is at most the distance from x to y .) And by our induction hypothesis:

$$= d^*(x) + l(x, y)$$

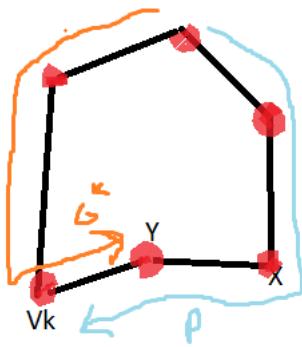
and by our assumption:

$$= d^*(y)$$

So we've now proved:

$$d^k(y) \leq d^*(y) < d^k(v_k)$$

A picture to see what's going on:



This is a contradiction because we've shown that the path down the left to y is shorter than the path to v_k , but we assumed the optimal path was from the right, and that meant that the distance to v_k had to be larger than the distance to y . \square

12.1.4 The Running Time

There are n iterations, there are at most n distance updates at each iteration. So at most $O(n^2)$ but again we can improve it to $O(m \log n)$ using a heap.

13 Huffman Codes

13.1 Data Encoding

Suppose we want to encode the alphabet in binary. How many bits do we need to encode every letter?

Five bits since $2^5 \geq 26$

How do we measure the quality of an encoding? A natural measure would be the length of the encoding. But what if some letters are used very often? We would want these to have a smaller size.

Let f_i be the frequency at which a letter i appears in the alphabet. Then:

$$\text{cost} = \sum_{i \in A} l_i f_i$$

13.1.1 Morse Code

Morse code follows this idea. It uses less bits for the frequently used letters, and less bits for the less common ones. But there's problem with it. It cannot be binary because it's ambiguous whether 101 means 101 or 1, 0, 1. So Morse code is actually ternary. It uses

pauses to signify the end of a letter.

How can we get around this?

13.2 Prefix Codes

A coding system is prefix-free if no codeword is a prefix of another codeword. Morse is not prefix free, since in 1101, 1 means t, 11 means m , 110 means g and 1101 means q. So t is a prefix of m is a prefix of g is a prefix of q.

13.3 Binary Tree Representations

We can use a binary tree T to represent a prefix-free binary code. Each left edge has label 0 and each right edge as label 1.

The leaf vertices are the letters of the alphabet. The codeword for a letter are the labels on the path from root to leaf.

Theorem 19. A binary coding system is prefix-free \Leftrightarrow it has a binary tree representation.

Proof. (\Leftarrow)

In a binary tree representation the letters are at the leaves. This means that the path P_x from the root to leaf x and the path P_y from the root to a leaf y must diverge at some point.

So the codeword for x cannot be a prefix of the codeword for y .

(\Rightarrow)

Given a binary coding system, we can define a binary tree recursively. A letter whose code word started with a 0 is placed in the left subtree. Otherwise it is placed in the right subtree. Then just recurse on the next letter. \square

Observe that the cost of the tree is:

$$\text{cost}(T) = \sum_{i \in A} f_i d_i(T)$$

where d_i is the depth of the node i .

Proof

We have the definition of cost:

$$\text{cost}(T) = \sum_{i \in A} f_i l_i(T)$$

The length of the word is just the sum of the edges in the word

$$= \sum_{i \in A} \sum_{e \in P_i} 1$$

Which is exactly the same as the depth in the tree.

$$= \sum_{i \in A} f_i d_i(T)$$

13.3.1 Letter to Leaf Assignment

How should we assign letters to leaves? The least frequent letters should be at the deepest leaf. So, we can just sort all the frequencies, and start adding each least frequent letter to the deepest leaf.

13.3.2 Tree Shape

But what should the shape of the tree be?

Let $n_e = \sum_{i \in A: e \in P_i} f_i$ be the number of letters (weighted by frequency) whose root-to-leaf paths use edge e in T . (How many letters use this edge)

$$\text{cost}(T) = \sum_{e \in T} n_e$$

Proof

We start with our first observation:

$$\text{cost}(T) = \sum_{i \in A} f_i d_i(T)$$

The depth is just the sum of the edges in the path from root to the vertex.

$$= \sum_{i \in A} f_i \sum_{e: e \in P_i} 1$$

Changing the order of summation:

$$= \sum_{e \in T} \sum_{i \in A: e \in P_i} f_i$$

Which is exactly our definition.

$$= \sum_{e \in T} n_e$$

13.4 The Key Formula

The key to designing a good coding system is the following formula:

Theorem 20. Let \hat{T} be the tree formed from T by removing a pair of sibling-leaves a and b and labelling it's parent by z where $f_z = f_a + f_b$ then:

$$\text{cost}(T) = \text{cost}(\hat{T}) + f_a + f_b$$

Observation 1 is telling us that the least frequent letters should be siblings, and observation 2 tells us how to find the optimal shape of the tree.

13.5 The Algorithm

```
Huffman(A,f)
  if A has two letters then
    encode one letter with 0 and the other with 1
  else
    let a and b be the most infrequent letters
    merge a and b into a new node z with frequency z = a + b
    recurse on the new set
    create the tree by adding a and b as children of z in the completed tree
```

13.5.1 Proof Of Correctness

Theorem 21. The Huffman Coding Algorithm gives the minimum cost encoding.

Proof. By Induction on the size of A .

Base Case: $|A| = 2$

Each letter has codeword length 1.

Induction Step:

Assume works for $|A| = k$. Take $|A| = k + 1$.

First, the algorithm merges the two smallest leaves into one vertex, so there are now $k - 1$ elements. We know that an optimal solution for \hat{A} where $|\hat{A}| = k - 1$ exists, by our induction hypothesis. Then, the algorithm can extend this to a solution for A by attaching leaves labeled a and b as children of the vertex in \hat{A} , z , that we merged earlier. \square

13.5.2 Running Time

There are $n-2$ iterations, each iteration takes $O(n)$ to find the two least frequent letters and update the alphabet. So, $O(n^2)$. Again, with heaps we can get it to $O(n \log(n))$.

14 Minimum Spanning Tree Problem

Given a graph, each edge e has a cost c_e , where all edge costs are distinct.

So the cost of a tree T is:

$$c(T) = \sum_{e \in T} c_e$$

14.1 Kruskal's Algorithm

Sort the edges $\{e_1, e_2, \dots, e_m\}$ by cost, least to greatest.

Set $T = \emptyset$

For each $i = \{1, 2, \dots, m\}$

Let $e_i = (u, v)$

if u and v are in different components of the tree, then add this edge to T .

14.2 Prim's Algorithm

Set $T = \{a\}$

If $V(T) \neq V(G)$ then

Let e be the minimum cost edge in $\delta(T)$ (edges leaving a vertex in T)

At this edge to T . (and it's vertices)

14.3 Boruvka's Algorithm

Set $T = \emptyset$

If T has more than one component $\{S_1, S_2, \dots, S_l\}$ then

For $i = \{1, 2, \dots, l\}$ let e_i be the minimum cost edge in $\delta(S_i)$

Add all of these edges to the tree.

14.4 Running Times

14.4.1 Kruskal's Running Time

It takes $O(m \log m)$ to sort the edges, and there's m iterations of the loop. Within the loop we have to search the tree to see if u and v are in different components. This takes time $O(n)$.

So we have:

$$O(m \log m + mn) = O(mn)$$

14.4.2 Prim's Running Time

We have n iterations of the loop. Within the loop, we have to exhaustively search for the minimum edge in time $O(m)$.

So:

$$O(mn)$$

14.4.3 Boruvka's Running Time

We have at most n components, finding the minimum edge takes $O(m)$, and there are $\leq \log n$ iterations.

So:

$$O(mn \log n)$$

14.5 Proof That They All Work

First, notice that for a chicken to cross a road and get to a chicken coop, it must cross the road an odd number of times.

We'll also need this fact:

Theorem 22. The Cut Property of a minimum spanning tree is this: Assume the edge costs are distinct. If e is the cheapest edge in some cut $\delta(S)$ then e is in the minimum spanning tree.

Proof. Let $e = (u, v)$ be the cheapest edge in a cut $\delta(S)$ recall that $\delta(S)$ is the edges leaving a subset of vertices S .

Let T^* be a minimum spanning tree, and to get a contradiction, assume $e \notin T^*$.

Since T^* is a spanning tree there is a unique path P in T^* from u to v .

Observation. If $\hat{e} \in P$ then $(T^* \setminus \hat{e} \cup e)$ is a spanning tree. (Basically we can replace an edge from the path joining u, v with u, v itself).

If a chicken is walking along P must cross the cut $\delta(S)$.

Observation. There is at least one edge $\hat{e} \in P \cap \delta(S)$.

We know from the beginning that e is the lowest cost edge, so

$$\begin{aligned} c_{\hat{e}} &> c_e \\ \Rightarrow (T^* \setminus \hat{e}) \cup e \end{aligned}$$

is a cheaper spanning tree than T^* . This contradicts the assumption that T^* was the minimum cost tree.

Essentially what we did was, replace \hat{e} , by e since we know that doing so gives us a cheaper spanning tree.

□

This proves all our algorithms simultaneously.

In the case of Prim's algorithm, we literally added edges based on if they were the minimum edge in the cut $\delta(T)$, so it directly uses this theorem.

In Baruvka's algorithm, we add edges from the cut $\delta(S)$ for each component, so again using the theorem.

In Kruskal's algorithm, we add edges if u, v are in different components. Let S be one of the two. Then e_i is the cheapest edge in $\delta(S)$ since we look at edges by order of cost. So this one also works.

14.6 The Cycle Property

Theorem 23. Assume distinct edge costs. If e is the most expensive edge in some cycle C , the e is not in the MST.

Proof. Let $e = (u, v)$ be the most expensive edge in the cycle C .

So $P = C - e$ is a path from u to v .

Assume for a contradiction that e is in the MST T^* .

Let $(S, V - S)$ be the cut introduced by $T^* - e$.

Observation 1: If $\hat{e} \in \delta(S)$ then $(T^* - e) \cup \hat{e}$ is a spanning tree.(basically we can replace e with \hat{e} and get a spanning tree, since \hat{e} also joins the two sets.)

Observation 2: There is at least one edge $\hat{e} \in P \cap \delta(S)$ (there's an edge that crosses the cut that is also part of the path.)

But $c_{\hat{e}} < c_e$ so we can replace e by \hat{e} which contradicts the fact that T^* was the MST. \square

14.7 The Reverse Delete Algorithm

Sort the edges by cost, most expensive to cheapest.

For each edge:

If $G \setminus \{e_i\}$ is connected then set $G = G \setminus \{e_i\}$

So basically take the most expensive edge, and if the graph is still connected without it, then throw it away.

14.7.1 Runtime of Reverse Delete

There are m iterations, and at each one need to check that graph is still connected (using BFS or DFS) in $O(m)$. So the running time is $O(m^2)$

14.7.2 Proof of Reverse Delete

First notice that $G \setminus \{e_i\}$ being connected means there was a cycle including e_i , and since we've sorted in reverse order, e_i is the most expensive edge, so by the cycle property, this algorithm works.

15 The Clustering Problem

Given a collection of objects, O we want to partition the objects into a set of clusters $\{S_1, \dots, S_k\}$. A "good" clustering has similar objects in the same clusters.

We represent the problem by a weighted graph G .

There is a vertex for each object O , and an edge between each pair of objects.

The weight $d_{ij} \geq 0$ of an edge represents the disimilarity of object i and object j .

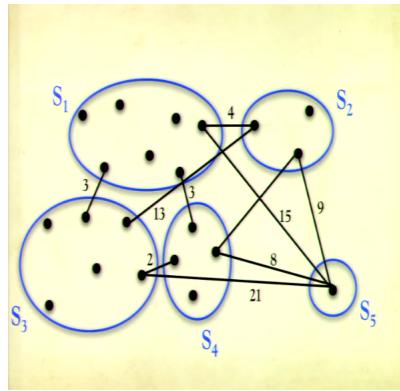
The quality of a clustering has no optimal definition. (Depends on application)

15.1 Maximum Spacing Clustering

Maximize the distances between the clusters. In other words, partition the vertices into k clusters so that the minimum distance between two vertices in different clusters is maximized.

Given a clustering $\{S_1, S_2, \dots, S_k\}$ we define the distance between two clusters as:

$$d(S_l, S_m) = \min_{i \in S_l, j \in S_m} d_{ij}$$



So here we just want to maximize the minimum black line (here it's 2) so the quality is (2).

15.2 Reverse-Delete Clustering Algorithm

Sort the edges by cost, highest to lowest.

For each edge:

If $G \setminus \{e_i\}$ has k components or less, then set $G = G \setminus \{e_i\}$

Notice, this is exactly the MST problem, where in MST, $k = 1$

15.2.1 Proof Of Reverse-Delete Clustering

First, observe this:

Theorem 24. A connected graph contains a spanning tree as a subgraph

Proof. Simply grow a BFS tree from any root vertex. \square

Next, observe this fact:

Theorem 25. We can remove an edge, and the number of components increases by at most 1.

Proof. Originally, u, v are in the same component S_1 . S_1 contains a spanning tree T .

Case 1: e is not in T . Then S_1 remains a component after deletion of e

Case 2: e is in T for every spanning tree in S_1 . Then S_1 is broken into two components on the deletion of e . \square

Now our algorithm:

Proof. Let e_l be the edge whose deletion causes the number of components to increase from $k - 1$ to k .

This means that the algorithm deleted all the edges up to e_l .

When we delete e_l we have the clustering $S = \{S_1, \dots, S_k\}$

But this means that only the edges up to e_l can cross between the clusters. Since we organized these to be largest to smallest, that means e_l is the shortest edge between two clusters. So the quality is determined by e_l .

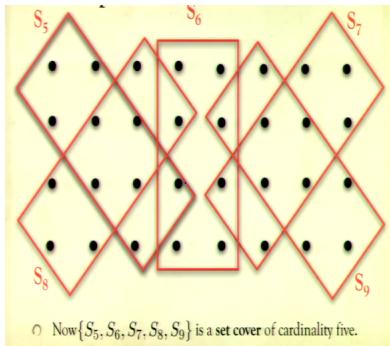
We now need to show that this is the optimal solution.

Any other clustering $S^* = \{S_1^*, \dots, S_k^*\}$ with k components must separate the endpoints of at least one edge with an endpoint in the edges up to e_l from below. (edges smaller than e_l).

But then we'll have separated two clusters by an edge shorter than e_l which is worse than S . \square

16 The Set Cover Problem

Given a collection of n items, I , and another collection of sets $S = \{S_1, \dots, S_m\}$ we want to find the smallest collection of sets in S that contain all of the elements of I .



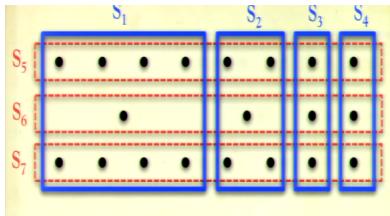
16.1 The Greedy Set Cover Algorithm

Repeat until $I = \phi$

Let $\hat{S} = \text{argmax } |S \cap I|$ (pick the set that covers the most items in I)

Set $S = S - \{\hat{S}\}$ and $I = I - \hat{S}$

This doesn't work!



But we'll see that it's pretty close.

16.2 Approximation Algorithms

An algorithm A is an α -approximation algorithm if for any instance I :

It runs in time $\text{poly}(|I|)$

It always outputs a feasible solution S .

It always guarantees: $\text{cost}(S) \leq \alpha * \text{OPT}$ where OPT is the optimal solution, and α is the desired approximation. (here we're looking at a minimization problem)

The greedy set cover algorithm is an approximation algorithm. So we want to find α for it, so we know how good it actually is.

Observation

If the optimal set cover has cardinality k then for any $X \subseteq I$, then there is some set S that covers at least $\frac{1}{k}|X|$ items of X .

Proof Let the optimal solution be $\{S_1^*, \dots, S_k^*\}$

Let the sets $\{S_1^*, \dots, S_k^*\}$ cover every item in I .

So they cover every item of any subset X of I . So since X is covered by k sets, there must be some set that covers greater than one k th fraction of X

16.3 Proof that Greedy Set Cover Almost Works

Theorem 26. If the optimal set cover has cardinality k then the greedy algorithm finds a solution of cardinality at most $k\ln(n)$.

Proof. wlog let the greedy algorithm output $\{S_1, \dots, S_T\}$ and let the optimal solution be $\{S_1^*, \dots, S_k^*\}$.

We want to show that $T \leq k\ln(n)$

Let I_t be the uncovered items in the start of step t For example, I_1 is just I .

Since $I_t \subseteq I$, by the observation before, there's a set that covers at least $\frac{1}{k}|I|$ items in I_t

This means in step t it must pick a set that covers at least $\frac{1}{k}|I|$ items in I_t .

$$\Rightarrow |I_{t+1}| \leq |I_t| - \frac{1}{k}|I_t|$$

$$= \left(1 - \frac{1}{k}\right)|I_t|$$

.

$$\leq \left(1 - \frac{1}{k}\right)\left(1 - \frac{1}{k}\right)\dots\left(1 - \frac{1}{k}\right)|I_t|$$

We iterate t times, so

$$= \left(1 - \frac{1}{k}\right)^t |I_1|$$

And since I_1 is just n :

$$= \left(1 - \frac{1}{k}\right)^t n$$

Key fact: $1 - x < e^{-x} \forall x \neq 0$

So if we let $\frac{1}{k} = x$, then we have:

$$|I_{t+1}| < (e^{-\frac{1}{k}})^t n$$

$$= e^{-\frac{t}{k}} n$$

Now setting $t = k \ln(n)$, then:

$$\begin{aligned} |I_{t+1}| &< (e^{-\frac{k \ln(n)}{k}})^t n \\ &= e^{-\ln(n)} n \\ &= 1 \\ \Rightarrow |I_t| &= |I_{k \ln(n)+1}| < 1 \end{aligned}$$

which means that it's empty, which means we're done at step t . So there was $t = k \ln(n)$ steps, we finished, meaning there's at most $k \ln(n)$ sets that the algorithm picked. \square

So this is a $\log(n) - \text{approximation}$ algorithm for this problem. This is actually a bad approximation, but it's the best we've come up with unless we solve P=NP.

16.4 Running Time

There are n iterations, and there's at most n distance updates at each iteration, so $O(n^2)$

16.5 The Hitting Set Problem

Given a collection S of m elements, and a collection I of $\{I_1, \dots, I_n\}$ sets that are subsets of S .

We want to select as few elements as possible such that there is at least one element selected in every set.

In other words, we want to find the smallest set X of elements such that every set I is "hit".

It was left as an exercise to show that this is exactly the same as the set cover problem.

17 Matriods

Given a set E of elements where each element has a weight $w_e \geq 0$. There is a collection \mathcal{F} of feasible subsets of E . Feasible meaning a valid solution to the problem.

Each set $F \in \mathcal{F}$ is a valid solution with weight:

$$w(F) = \sum_{e \in F} w_e$$

So the weight of a solution is the sum of the element weights.

The problem is then to find a feasible set in \mathcal{F} with the maximum weight.

17.1 The Hereditary Property

\mathcal{F} satisfies the hereditary property if:

$$F \in \mathcal{F} \Rightarrow \hat{F} \in \mathcal{F}, \forall \hat{F} \subseteq F$$

Basically subsets of a feasible solution are also a feasible solution.

This arises in the interval selection problem. E is the set of intervals, $F \in \mathcal{F}$ if F is a disjoint collection of intervals.

The Maximum Weight Spanning tree problem also has this property. Where E is the set of edges, and $F \in \mathcal{F}$ if F is a forest. (If you solve each component of the forest, you can put them together into the larger solution). (Subsets of forests are also forests)

17.2 The Greediest Algorithm

A generic algorithm for a hereditary set system:

Sort the elements by weight $\{w_1 \geq w_2 \geq \dots \geq w_m\}$.

Set $T = \emptyset$

For $i = \{1, 2, \dots, m\}$

If $T \cup e_i \in \mathcal{F}$ then set $T \leftarrow T \cup e_i$.

Basically, if making my solution bigger by adding the current element is still in the valid solution, then do so.

17.2.1 The Running Time

Sorting time is $O(m \log m)$, and there are m iterations. Each test for feasibility takes time T . So $O(m \log m + mT)$. So as long as test for feasibility is fast, then our algorithm is fast.

17.2.2 Does It Work?

This is essentially Kruskal's algorithm, which works in that scenario.

This does **NOT** work for the interval selection problem.

When does it work then?

17.3 The Augmentation Property

\mathcal{F} satisfies the augmentation property if:

$$F, \hat{F} \in \mathcal{F} \text{ and } |F| > |\hat{F}|$$

$$\Rightarrow \exists e \in \hat{F} \text{ such that } \hat{F} \cup e \in \mathcal{F}$$

Basically, if there are two feasible sets where one is strictly larger than the other, there is an element in the bigger set that could be added to the smaller set, and still overall be a feasible solution.

17.4 What is a Matroid?

A matroid is a non-empty set system $M = (E, \mathcal{F})$ that satisfies both the Hereditary and Augmentation Properties.

Notice that that hereditarity and non-emptyness $\Rightarrow \emptyset \in \mathcal{F}$

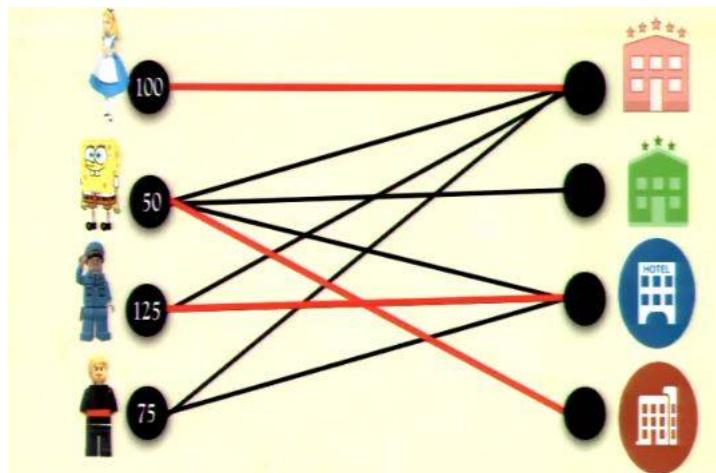
17.4.1 Examples of Matroids

E is the set of edges in a graph. $F \in \mathcal{F}$ if F is a forest. Basically, subsets of forests are still forests, and you can augment forests by adding an edge in a bigger forest.

E is a finite set of vectors in a vector space. $F \in \mathcal{F}$ if F is a collection of linearly independent vectors.

E is the set of left vertices in a bipartite graph. $F \in \mathcal{F}$ if there is a matching in the graph that matches each vertex in F to a distinct vertex on the right. (Halls theorem!)

The online auction problem:



Here you want to maximize profits.

The Job Scheduling Problem with Deadlines:

One job can be processed per day, each job has a deadline, and a late cost. Minimize the losses and complete the most jobs before the deadlines.

As an exercise, prove that these are Matroids.

17.5 Characterization of Matroids

The greediest algorithm works when matroids are the structure we're dealing with!

Theorem 27. A hereditary, non-empty set system M is a matroid \Leftrightarrow the greediest algorithm outputs the optimal solution in M for any set of weights w .

Proof. (\Rightarrow)

First, the greediest algorithm works on matroids:

Let the algorithm output $\{e_1, \dots, e_l\}$ where:

$$w(e_1) \leq w(e_2) \leq \dots \leq w(e_l)$$

Let the optimal solution be $\{e_1^*, \dots, e_l^*\}$ where:

$$w(e_1^*) \leq w(e_2^*) \leq \dots \leq w(e_k^*)$$

Notice that since we have a matroid, the augmentation property holds. So we have $l \geq k$, otherwise, the algorithm would have selected another element. (The greediest algorithm needs to select at least as many elements as the true solution).

Now we want to show that $w(e_i) \geq w(e_i^*)$ (the greedy solution is at least as good as the optimal one).

Suppose not. Let j be the smallest index with $w(e_j) < w(e_j^*)$. Now consider:

$$\hat{F} = \{e_1, \dots, e_{j-1}\} \text{ and } F = \{e_1^*, \dots, e_j^*\}$$

So by the augmentation property, $\exists e_i^* \in F$ such that $\hat{F} \cup e_i \in \mathcal{F}$

Or in English, \hat{F} is a subset of a feasible solution, so it's a feasible solution itself, so we can add e_i from the optimal solution and still be feasible.

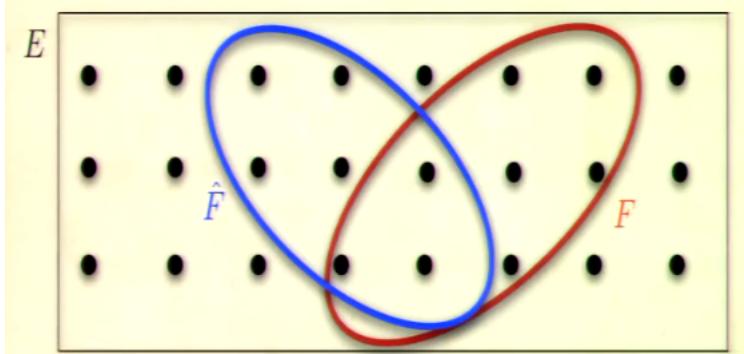
But since we ordered everything, $w(e_i^*) \geq w(e_j^*) > w(e_j)$. This is a contradiction because then the greediest algorithm should have chosen e_i^* instead of e_j , since choosing e_j produces a lesser result.

(\Leftarrow) Take a hereditary set system M that is not a matroid. Then:

$$\exists F, \hat{F} \in \mathcal{F} \text{ with } |F| > |\hat{F}| \text{ but } \nexists e \in F - \hat{F} \text{ s.t. } \hat{F} \cup e \in \mathcal{F}$$

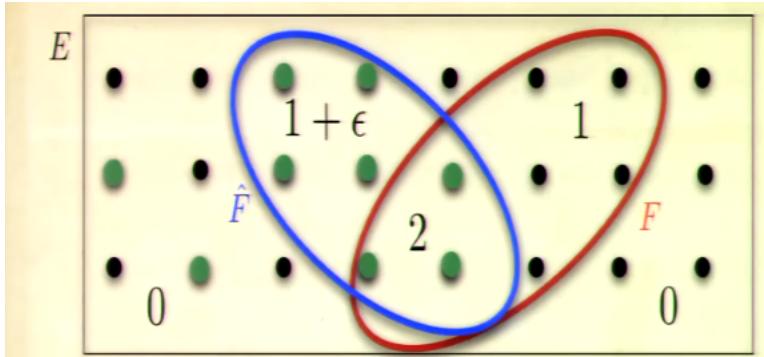
Or in English, you have two feasible solutions with one larger than the other, but you cannot augment the smaller one with an element of the larger and still have a feasible solution.

Here's a counter example:



- Now here is a collection of weights that cause the greedy algorithm to fail: $w(e) = \begin{cases} 2 & \text{if } e \in F \cap \hat{F} \\ 1+\epsilon & \text{if } e \in \hat{F} \setminus F \\ 1 & \text{if } e \in F \setminus \hat{F} \\ 0 & \text{if } e \notin F \cup \hat{F} \end{cases}$

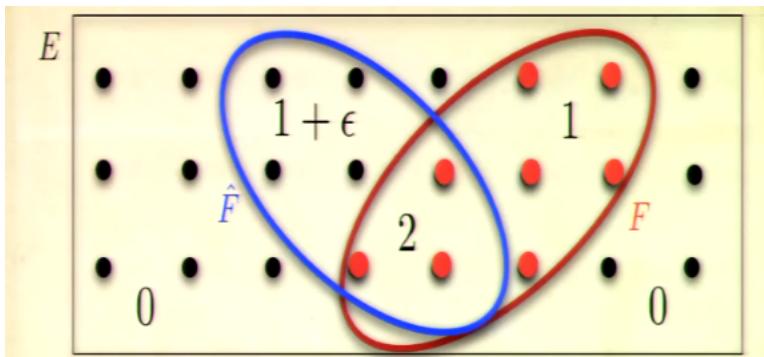
The greedy algorithm produces this solution:



- As the greediest algorithm runs:
 - It first selects all the elements in $F \cap \hat{F}$
 - It next selects all the elements in $\hat{F} \setminus F$
 - Finally it (possibly) selects some elements in $E \setminus (F \cup \hat{F})$
- So the algorithm outputs a solution with weight:

$$2 \cdot |F \cap \hat{F}| + (1 + \epsilon) \cdot |\hat{F} \setminus F| + 0$$

But this would have been better:



- So the algorithm outputs a solution with weight:

$$2 \cdot |F \cap \hat{F}| + (1 + \epsilon) \cdot |\hat{F} \setminus F| + 0 = 2 \cdot |F \cap \hat{F}| + (1 + \epsilon) \cdot |\hat{F} \setminus F| < 2 \cdot |F \cap \hat{F}| + 1 \cdot |F \setminus \hat{F}|$$

□

Part IV

Dynamic Programming

18 Fibonacci Numbers

Base Cases:

$$F(0) = 1, F(1) = 1$$

Recurrence:

$$F(n) = F(n - 1) + F(n - 2) \forall n \geq 2$$

Models the growth of asexual populations.

18.1 Closed Form

The above recurrence can be solved and this form is obtained:

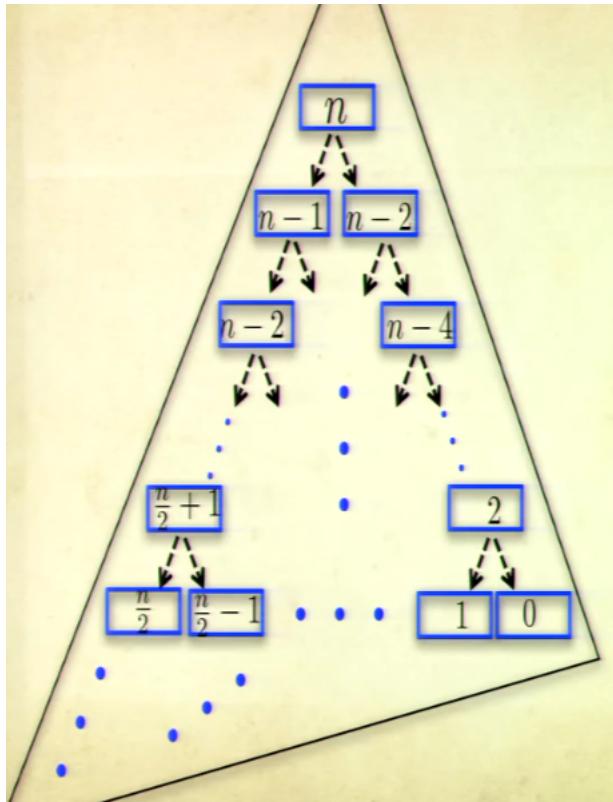
$$F(n) = \frac{1}{\sqrt{5}} \left(\left(\frac{1 + \sqrt{5}}{2} \right)^n - \left(\frac{1 - \sqrt{5}}{2} \right)^n \right)$$

Which turns out to be $\Theta(1.618^n)$.

But how long would it take to solve the recurrence by unwinding the recursive formula? (Not by using the techniques in MATH240).

18.2 Recursive Tree of Fibonacci Formula

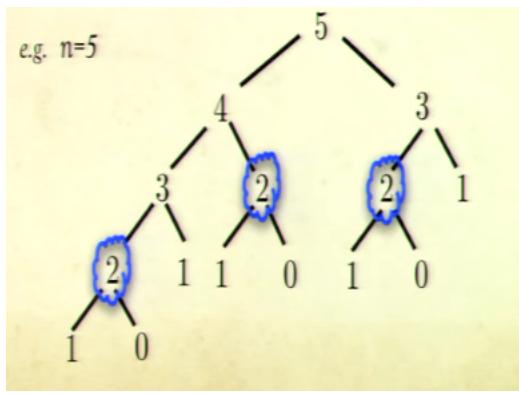
Well, at each level we're doing a constant amount of work. So we'll just add a $+1$ every time. We split into two sub-problems, so the work increases as 2^l at each level. But how many levels?



Well, we can see that the right side of the tree is decreasing by 2 each time, so there are $\frac{n}{2}$ levels on that side. So the running time is at least $\Omega(2^{\frac{n}{2}}) = \Omega(\sqrt{2}^n)$. That's so slow!

18.2.1 Why is it so slow?

The problem is that the recursive formula solves the same subproblem many times.



Here it solves for $n = 2, 3$ times. The solution is storage.

19 Dynamic Strategies

19.1 Top-Down (Memoization)

Our Fibonacci example:

If $n < 1$ then return 1.

If $F(n)$ is undefined

then set $F(n) \leftarrow F(n - 1) + F(n - 2)$

So really you only solve problems once.

There are $n - 1$ additions, which may be n digits long, so $O(n^2)$ (adding n-digit numbers is $O(n)$).

19.2 Bottom-Up

Just start at the base cases and accumulate up to n . Essentially with iteration.

19.3 Top-down vs Bottom-Up

Sometimes you don't need to solve all of the subproblems, so top-down solves only the ones that absolutely need to be solved. So is faster in those cases.

19.4 Difference Between Divide + Conquer and DP

Much like divide and conquer, we're dividing up our problems into sub-problems, solving them and putting them together into a solution. However in DP, these sub-problems may overlap, until the Divide and Conquer problems.

19.5 Formula for Running Time of Dynamic Program

Let k be the number of sub-problems in the recursive formula, and let l be the number of sub-problems overall, then the running time is:

$$O(kl)$$

20 Interval Scheduling Revisited

When we used greedy algorithms, we said it would be useful to have a more general version.

20.1 Weighted Interval Selection

There is a set $I = \{1, 2, \dots, n\}$ of intervals. Each interval has a start time s_i , finish time f_i , and now a value v_i . The goal is to maximize the value of our set S , where S is the set of disjoint intervals with the maximum total value $\sum_{i \in S} v_i$.

Recall the first fit algorithm worked for non-weighted algorithms. Does it still work for the weighted interval selection? NO!

20.2 Dynamic Approach

Recall that for a dynamic program we want the solution \mathcal{P} to be computable using a set of sub-problems, each with a natural ordering from smallest to largest. We also need there to be a polynomial number of sub-problems.

So how do we define the sub-problems?

In this situation, we have n intervals, $\{1, 2, \dots, n\}$. We'll simply define a sub-problem $\{1, 2, \dots, l\}$, and find the maximum value for this set of intervals.

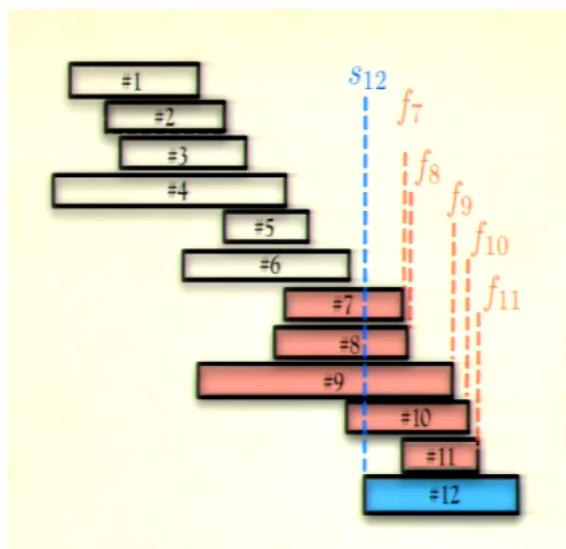
But we can change the labels $\{1, 2, \dots, n\}$ to whatever we want, so what's the best labeling of the intervals?

Again, the best way to do it is by finish times.

20.2.1 Why This Ordering?

First, notice that any interval finishing after the start interval n , clashes with it.

Next, notice that intervals that clash with n are consecutive.



Proof. Let interval l be the lowest index interval finishing after n starts. Then, by our first observation, l clashes with n . But for any interval $i > l$, we have that $f_i \geq f_l$, since we are ordered by finish times. This means that $f_i > s_n$, which means that i also clashes with n . So the intervals $\{l, l+1, \dots, n-1\}$ all clash with n . Conversely, no interval i with index $i < l$ since, $s_i < f_i \leq s_n < t_n$. Which concludes our proof.

□

Let $h(n)$ be the highest index interval that is disjoint from n . If we select interval n , then $\{h(n)+1, h(n)+2, \dots, n-1\}$ are conflicting.

So if we select interval n , then we can only pick the next to be from 1 to $h(n)$. (This is our subproblem). What if we don't select interval n ? Then we try to find an optimal solution from groups 1 to $n-1$.

Note that this same property holds for any $h(j) < j$

So the recursive formula for our algorithm is:

$$\text{optimal}(j) = \max\{\text{optimal}(j-1), v_j + \text{optimal}(h(j))\} \quad \forall j \geq 1$$

Basically, take either the optimal solution for not picking j , or take j , (and thus its value, v_j) and find the optimal solution for the things up to $h(j)$, and see which one is better.

However, as we saw with Fibonacci numbers, solving this with the formula above will be too slow. So we'll need a dynamic program.

20.3 Weighted Interval Selection Algorithm

Sort the intervals by finish times.

Set $\text{opt}(0) = 0$

For $j = \{1, 2, \dots, n\}$

Let $h(j) < j$ be the highest index interval that is disjoint from j .

Set $\text{opt}(j) = \max\{\text{opt}(j-1), v_j + \text{opt}(h(j))\}$ If $\text{opt}(j) + \text{opt}(j-1)$ then $\text{parent}(j) = j-1$, else $\text{parent}(j) = h(j)$

This requires storing a pointer to the parent of each interval, so that we can know which intervals to select in the end by following the pointers back.

20.3.1 The Running Time

Recall, the running time of a DP in general is the number of sub-problems multiplied by the time required to use the recursive formula at each step.

We have n sub-problems, and to solve a sub-problem we examine the solutions to 2 smaller subproblems (constant time), so this is $O(n)$

20.4 Key Points

The four steps we did in this problem, can be used in almost any dynamic program. First, understand the underlying structure of the program. Second, set up a recursive formula. Third, solve the recursion by a bottom-up dynamic program. Last, to recover the solution, follow the pointers backwards.

Note that most of the work is in step 1.

21 Classifying Dynamic Problems

With the 4 steps in mind, we need to define the sub-problems clearly. We need them to be computable using these sub-problem, we want an ordering of the sub-problems, a polynomial number of sub-problems, and a recurrence giving the solution in terms of the subproblems.

21.1 Structure of Sub-problems

There are 4 main classes of dynamic sub-problem structures. 1.) One-sided interval 2.) Box Structure 3.) Two-sided interval 4.) Tree Structure.

21.1.1 One-Sided Interval

The input is $\{x_1, x_2, \dots, x_n\}$ and the sub-problems are of the form $[j] = \{x_1, \dots, x_j\}$ Basically the subproblems are just pieces of the input list. For example, the weighted interval selection problem.

21.1.2 Box-Structure

The input is $\{x_1, x_2, \dots, x_n\}, \{y_1, y_2, \dots, y_n\}$. And the sub-problems are of the form: $[i, j] = \{x_1, x_2, \dots, x_i, y_1, y_2, \dots, y_j\}$. Essentially the one-sided interval, but on two dimensions. Example: Knapsack problem, humpty dumpty.

21.1.3 Two-Sided Interval

The input is $\{x_1, x_2, \dots, x_n\}$ and the subproblems look like: $\{x_i, x_{i+1}, \dots, x_j\}$ basically an interval that doesn't start at 1. Example: RNA secondary structure problem.

21.1.4 Tree Structure

The input is a rooted tree with a vertex set $\{x_1, x_2, \dots, x_n\}$ and the sub-problems are sub-trees rooted at x_j .

22 Knapsack Problem

Say you have a bag with capacity W , there are n items to put in it. Item i has value v_i and weight w_i . We want to get the maximum total value into the knapsack.

$$\max \sum_{i=1}^n v_i x_i$$

such that we're less than W , and x_i is 0 or 1.

The dimensions of the box-structure are the items, and the capacity. What happens if we select item n ? We lose capacity, so $W < -W - w_n$, and the set of items loses n . If we don't pick n , we cast it aside, so the capacity is unchanged, and the set of items loses n .

22.1 Recursive Formula

$$opt(j, w) = \max\{opt(j - 1, w), v_j + opt(j - 1, w - w_j)\}$$

22.2 Dynamic Program

Set $opt(0, w) = 0 \forall w \leq W$

Set $opt(j, w) = 0 \forall j, w$

For $j = \{1, 2, \dots, n\}$

For $w = \{1, 2, \dots, W\}$ Set $opt(j, w) = \max\{opt(j - 1, w), v_j + opt(j - 1, w - w_j)\}$ Output $opt(n, W)$

We also need to add pointers if we want to output the optimal solution.

22.3 Running Time

There are nW sub-problems, and we need to examine the solutions to two sub-problems for each one, so $O(nW)$. This is a *pseudo-polynomial*, algorithm, meaning its polynomial in the number of weights. If W is small, it's all good.

23 The Humpty-Dumpty Problem

There is a skyscraper with T storeys. You have k eggs. What's the maximum story from which an egg does not break? Solve with minimum number of egg-drop tests.

You could try from floor 0 and work your way up, but that might take T tests.

You could try binary search, but then $k \leq \log(T)$

23.1 Dynamic Approach

The two dimensions are: the number of eggs remaining, the number of storeys remaining. Let $d(T, k)$ be the number of drops needed to guarantee we can solve the problem.

If the egg breaks, then we only have $k - 1$ eggs left, and $t - 1$ storeys left. So $d(T, v) \leq 1 + d(t - 1, k - 1)$

Set $d(\tau, 1) = \tau$

Set $d(0, k) = 0$

For $\tau = \{1, 2, \dots\}$

For $k = \{1, 2, \dots, T\}$

Set $d(\tau, k) = 1 + \min\{\max\{d(t - 1, k - 1), d(T - t, k)\}\}$

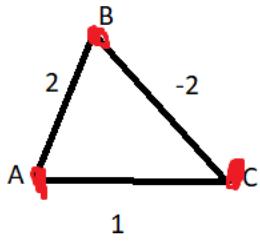
23.2 Running Time

There are Tk sub-problems, and to solve a sub-problem we need to examine the solutions to $2T$ smaller problems, so $O(T^2k)$.

24 Shortest Paths Revisited

In Dijkstras algorithm, we assumed that the length of all arcs were positive. But in the dynamic version of the algorithm, we'll be able to deal with negative arc lengths.

Negative lengths are a problem because Dijkstras fails. As per this example:



Dijkstras says the shortest path from A to C is to go directly. But we can see that it's shorter to go through B.

24.1 Negative Cycles

Negative cycles are a problem because then there is no shortest path. You could just keep going around the cycle and making your path "shorter" (negative weight) to infinity.

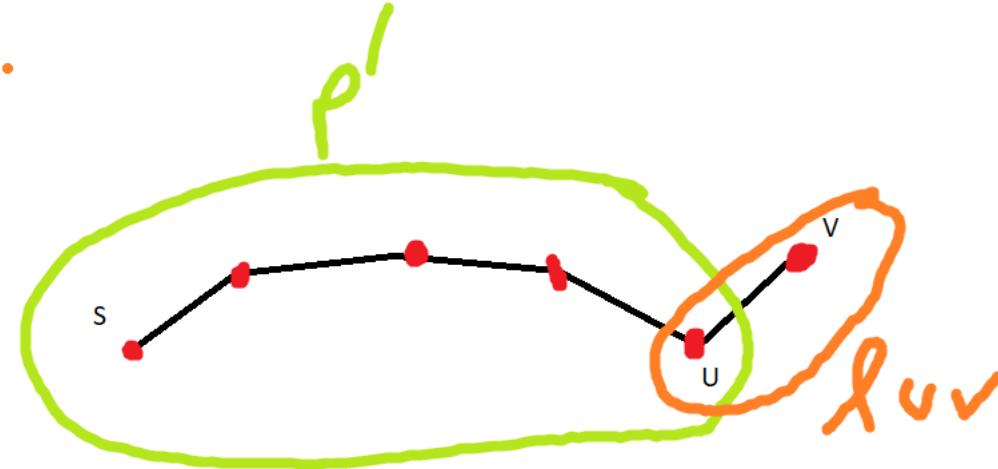
So rather than a shortest walk, we'll be looking for shortest paths. That is, no repeated edges allowed.

24.2 The Dynamic Program

We have two dimensions. Vertices: $\{1, 2, \dots, n\}$ and the number of arcs in a path: $\{0, 1, \dots, n-1\}$. We want $d(v, k)$, the shortest path from s to v using at most k arcs. If we plug $k = n - 1$, then we get our solution.

Notice that a path using $< k$ arcs has this property: $d(v, k) = d(v, k - 1)$. If it uses exactly k arcs, then the best path has k arcs.

Also notice this:



The length of the path P from s to v is the same as the length of P' , with l_{uv} . That is,

$$l(P) = l(P') + l_{uv}$$

Our recurrence is then:

$$d(v, k) = \min[d(v, k - 1), \min_{u \in \Gamma^-}(d(u, k - 1)) + l_{uv}]$$

where Γ^- is the set of in-neighbors of u .

This recurrence basically says our two observations.

24.3 The Runtime

We have n choices for the set of vertices, n choices for k so there are n^2 subproblems. On each one, we need to lookup n things in our table (at most n in-neighbors) so the runtime is $O(n^3)$ However in practice it's more like $O(mn)$, since they graphs aren't always so dense.

24.4 Finding negative cycles

We make two claims. The first is that $d(v, n) = d(v, n-1)$ if there are no negative cycles. The second is that if $d(v, n) = d(v, n-1)$, then there are no negative cycles. The proofs weren't covered in class, so I'll consider them non-essential but are on the slides on the Facebook page.

We can use this claim to have a way to find negative cycles. If we run the Bellman Ford algorithm (that's what this algorithm is called) for an extra iteration at the end and the shortest path becomes magically shorter, then there is a negative cycle. That is, if $d(v, n) < d(v, n-1)$

25 Finishing Dynamic Programming

In the tree structure, the subproblems are sub-trees.

The input is a rooted tree with vertex set $\{x_1, x_2, \dots, x_n\}$ and the subproblem is a sub-tree rooted at x_j .

We can order the subproblem based on $i < j \Leftrightarrow x_i$ is a descendant of x_j .

25.1 The Independent Set Problem

An independent set is a collection of pairwise non-adjacent vertices in a graph. The problem is to find an independent set of maximum cardinality in a graph.

We can answer this recursively.

What if we select the vertex? Then the algorithm cannot choose any neighbor of v . The vertices remaining are $V - (v \cup \Gamma(v))$

If not chosen, the vertices remaining are $V - \{v\}$

$$opt(G) = \max\{opt(G - v), 1 + opt(G - (v \cup \Gamma(v)))\}$$

Base cases: $opt(\emptyset) = 0$

The issue here is that the natural ordering is unclear, and we don't have a polynomial number of subproblems, since there is a sub-problem for each possible subset of V . So exponential number of subsets to look at.

It's quite possible there's no efficient algorithm for the independent set problem.

25.2 Independent Set On a Tree

What if we can solve the problem on the tree?

We now have that there is a natural ordering, and a polynomial number of subproblems.

If the algorithm doesn't select r , where r is the root, then it can choose from $V - \{r\}$. If it does, then we must select any vertices that are not children of r . Again, we're left with

disjoint subtrees.

There are only n subtrees, since there are n possible choices of root.

$$opt(T_v) = \max\left\{\sum_{u \in C} opt(T_u), 1 + \sum_{w \in C^2(v)} w \cdot opt(T_w)\right\}$$

Where C^2 is the children of the children of v .

Base case: $opt(\phi) = 0$.

Note, this also works if each vertex has a weight or value, by replacing 1 with the appropriate weight.

25.3 The Running Time

There are n sub-problems, to examine each sub-problem we examine at most n smaller problems, so $O(n^2)$.

However, each vertex has only one parent and one grandparent, so it will be used as a "lookup" only twice, so we actually have $O(n)$.

26 Shortest Paths Problem Yet Again

This will give the shortest paths for EVERY pair of vertices. We could just run the Bellman-Ford algorithm n times, but this gives runtime of $O(n^2m)$, and we can probably do better.

26.1 Floyd-Warshall

If we order the vertices in an arbitrary order $\{1, 2, \dots, n\}$ we create a subproblem from each triplet of vertices $\{u, v, k\}$. Here, $opt(u, v, k)$ is the shortest path from u to v using at most the vertices from $\{1, 2, \dots, k\}$

If $k \notin \mathcal{P}$ then clearly $opt(u, v, k) = opt(u, v, k - 1)$.

If $k \in \mathcal{P}$ then $opt(u, v, k) = opt(u, k, k - 1) + opt(k, v, k - 1)$. To see this, notice we can go from u to k , then from k to v and it's the same as just going from u to v . Combining these two shortest paths gives the overall shortest path.

This gives a 3D-box structure. The subproblems are ordered by their k values.

$$opt(u, v, k) = \min\{opt(u, v, k - 1), opt(u, k, k - 1) + opt(k, v, k - 1)\}$$

Base cases: $opt(u, v, 0)$ is 0 if $u = v$, l_{uv} if are connected, and ∞ if there isn't an arc.

26.2 Program

Set all base cases:

For $k = \{1, 2, \dots, n\}$ For $u = \{1, 2, \dots, n\}$ For $v = \{1, 2, \dots, n\}$

Exercise: How would you add pointers to keep track of the shortest paths themselves?

26.3 Running Time

There are n^3 subproblems, and we look at 3 smaller problems so $O(n^3)$.

Part V

Network Flows

27 Maximum Flows in a Network

In the shortest path problem, we want to find a minimum length path from a source s to a sink t .

In the maximum flow problem we have a good produced at a source s , and we want to send as much of the good as possible to a sink vertex t . Each arc $a = (i, j)$ has a capacity $u_a = u_{ij}$. The capacity is the maximum amount of good that we can send over a given period.

A network flow f satisfies two properties:

For a flow f_a , $0 \leq f_a \leq u_a$

The flow into a vertex v that is not the source or the sink, equals the flow out of that vertex. $\sum_{a \in \delta^-(v)} f_a = \sum_{a \in \delta^+(v)} f_a$

27.1 The Value of a Flow

The value of a flow f is the quantity of flow that reaches a sink t .

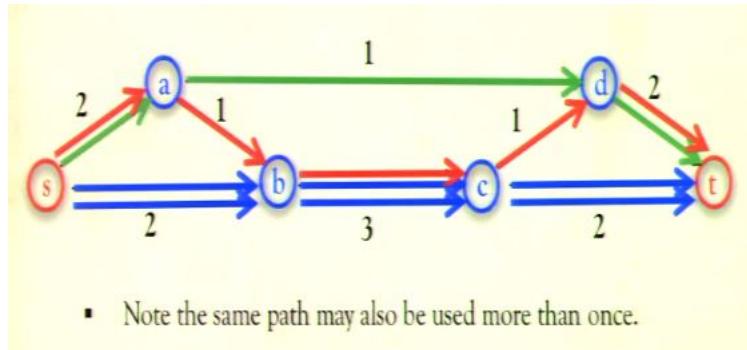
$$|f| = \sum_{a \in \delta^-(t)} f_a = \sum_{a \in \delta^+(s)} f_a$$

Assume no arc into s and no arcs out of t .

The problem is then to find the maximum flow.

27.2 Relation to Paths

If we think of a path, it's a flow since the flow in is the same as the flow out, and similarly unions of flows satisfy the same property.

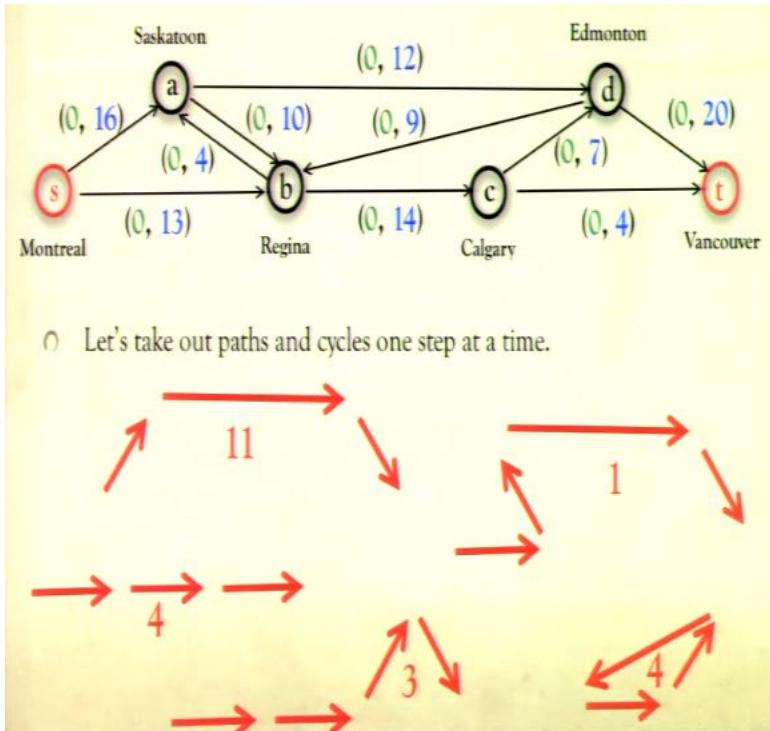


If this union of paths satisfies the capacity constraint on each arc, then it's an $s - t$ flow.

Cycles also satisfy flow conservation, and unions of cycles also. So if the union of cycles also satisfy capacity constraint, then they're an $s-t$ flow. The issue is that the flow may never reach t . (Trapped in cycle)

The union of a set of paths and a set of cycles satisfies flow conservation, so if it also satisfies capacity, then they form an $s - t$ flow. The converse is also true, every $s - t$ flow is made up of $s - t$ paths and cycles.

We can decompose the flow into paths and cycles like so:



Let's take out paths and cycles one step at a time.

Now if you add up all the paths and cycles you get the original flow.

Decompositions are not unique!

Also note that cycles can be removed without affecting the value of the flow.

27.3 Finding Maximum Flows

So we can search for the maximum flows by searching for directed paths.

We can use backward arcs as negative values, to reduce the amount of flow used on that arc in the forward direction. This lets us correct mistakes in our decomposition.

We need to be careful to maintain positive flow values, and maintain the arc doesn't exceed capacity.

So we can increase the flow on a on augmenting path (containing forward and backward arcs) P by:

$$b(P, f) = \min[min_{i,j} forward u_{ij} - f_{ij}, min_{i,j} backward f_{ij}]$$

This is called the bottle neck capacity.

27.4 Ford-Fulkerson Algorithm

Set $f = 0$

Repeat

Find an augmenting path P wrt f

Augment the flow on the path by $b(P,f)$

We'll see later that this works, and how to find these augmenting paths.

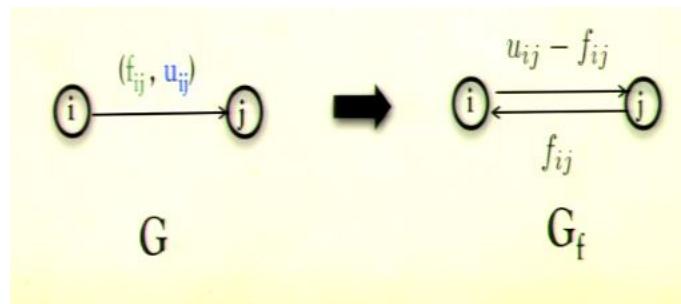
28 The Maxflow Mincut Theorem

28.1 When and How can the Algorithm Use an Arc?

An arc $a = (i, j)$ can be used forwards if there is capacity to spare in the arc. $f_{ij} < u_{ij}$. We can use it backwards if it has already been used forwards, ie: $f_{ij} > 0$.

28.2 The Residual Graph

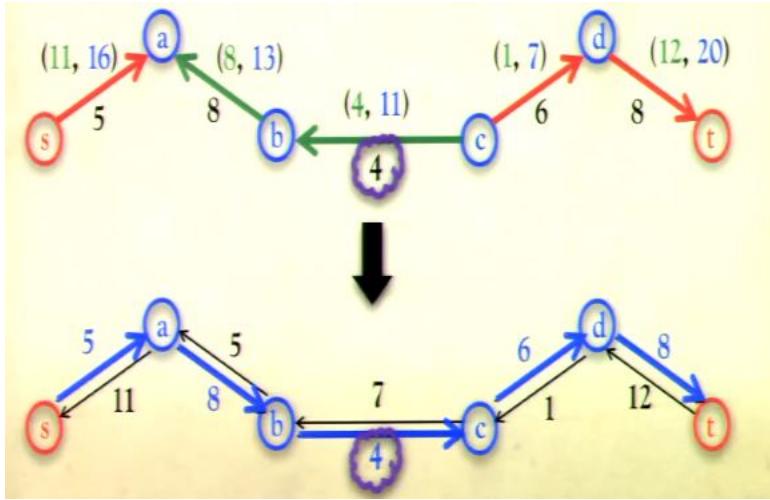
Given $G = (V, A)$ and a flow f , we define the residual graph G_f . It contains an arc (i, j) of capacity $u_{ij} - f_{ij}$ and one of f_{ji} , for each arc of G .



Essentially, we can think of each arc of G actually being two arcs.

28.2.1 Bottleneck Capacity

We can now just look for the lowest value arc in the directed path from s to t .



28.3 Ford-Fulkerson Revisited

We can now:

Set $f=0$

Repeat

Find a directed s - t path P in the residual graph G_f

Augment the flow on the path P by its bottleneck capacity $b(P,f)$

When you get to a point where there's no more paths from s to t in the residual graph.
Is this maximal though?

28.4 s - t Cuts

We say that $(\mathcal{S}, V - \mathcal{S})$ is an s - t cut if $s \in \mathcal{S}$ and $t \notin \mathcal{S}$

Let f^* be the flow output when the algorithm terminates. Let S^* be the set of vertices reachable from s in the residual graph G_f .

First, notice that $(\mathcal{S}^*, V - \mathcal{S}^*)$ is an s - t cut. Since we know the algorithm terminates when there is not path from s to t , so at the final stage of the algorithm, t is not in \mathcal{S}^* . Also note that there are no arcs leaving \mathcal{S}^* in G_f . If something would leave it, then it would be able to go add an extra vertex to S^* .

28.4.1 The Cut Lemma

Given a flow f and an s - t cut, then the value of the flow is;

$$|f| = \sum_{a \in \delta^+(S)} f_a - \sum_{a \in \delta^-(S)} f_a$$

Proof. The value of the flow is the amount of flow leaving the source.

$$\begin{aligned} |f| &= \sum_{a \in \delta^+(s)} f_a \\ &= \sum_{a \in \delta^+(s)} f_a - \sum_{a \in \delta^-(s)} f_a \end{aligned}$$

since there are no arcs coming in to s .

$$= \left(\sum_{a \in \delta^+(s)} f_a - \sum_{a \in \delta^-(s)} f_a \right) + \sum_{v \in S-s} \left(\sum_{a \in \delta^+(s)} f_a - \sum_{a \in \delta^-(v)} f_a \right)$$

we can do this because the right hand term is 0. (by the conservation property).

$$= \sum_{v \in S} \left(\sum_{a \in \delta^+(s)} f_a - \sum_{a \in \delta^-(v)} f_a \right)$$

now take any arc. There are four cases:

1.) $(i, j) \in \delta^+(\mathcal{S})$. The value of this arc only appears once in the sum.

2.) $(i, j) \in \delta^-(\mathcal{S})$. The value of this arc only appears once in the sum. But with a negative coefficient.

3.) $i, j \subseteq V - \mathcal{S}$ The value of this arc does not appear in the sum.

4.) $i, j \subseteq \mathcal{S}$ The value of this arc is counted twice, but once positive once negative. Thus 0.

So the only contributing arcs are those leaving or entering the cut.

$$\Rightarrow |f| = \sum_{a \in \delta^+(\mathcal{S})} f_a - \sum_{a \in \delta^-(\mathcal{S})} f_a$$

□

28.5 The Capacity of a Cut

The capacity is defined as the sum of the capacities leaving the cut.

Corollary: $|f| \leq cap(\mathcal{S})$

Proof. By the Lemma, we have:

$$\begin{aligned} |f| &= \sum_{a \in \delta^+(\mathcal{S})} f_a - \sum_{a \in \delta^-(\mathcal{S})} f_a \\ &\leq \sum_{a \in \delta^+(\mathcal{S})} u_a + 0 \end{aligned}$$

by the capacity constraint. □

28.6 Maxflow-Mincut Theorem

By the corollary, the value of the maximum flow is at most the capacity of the minimum cut. They turn out to be equal.

Proof. Let f^* be the output of the Ford-Fulkerson algorithm

Let S^* be the vertices reachable from s in the residual graph G_f .

We know that $|f^*| \leq \text{cap}(S^*)$ So we need to show that $|f^*| \geq \text{cap}(S^*)$

Recall that $\delta_{G_f}^+(S^*) = \emptyset$ otherwise we could have grown S^* .

Take any arc leaving S^* . Then this arc is not in the residual graph. This means they are at full capacity!

$$\Rightarrow f_{ij}^* = u_{ij}$$

similarly, any arc coming into S^* is not in the residual graph.

$$\Rightarrow f_{ij}^* = 0$$

now by the Lemma:

$$\begin{aligned} |f| &= \sum_{a \in \delta^+(S)} f_a - \sum_{a \in \delta^-(S)} f_a \\ &= \sum_{a \in \delta^+(S)} u_a - \sum_{a \in \delta^-(S)} 0 \\ &= \text{cap}(S^*) \end{aligned}$$

□

This actually proves that the algorithm works, since it only terminates once we have a minimum cut, but when we have a minimum cut we have a maximum flow.

We can actually solve a whole new selection of problems related to cuts now since the problems are equivalent.

28.7 Running Time

We find G_f in time $O(m)$, to find a path we run some graph exploration algorithm in $O(m)$. Then we augment in time $O(n)$.

So the total running time is $O(m * \text{number of iterations})$

We always increment the flow value by at least 1 in each iteration, but by the theorem, we know that the maximum flow is the same as the minimum cut, and we know the minimum cut will have capacity at most: nU where U is the largest capacity of all the arcs.

$$\Rightarrow O(mnU)$$

This is pseudo polynomial, since the bits in U can be huge.

29 The Vertex Matching and Cover Problem

Is there an efficient algorithm to find a maximum cardinality matching in a bipartite graph?
Yes! Using maximum flows.

29.1 Auxiliary Network

To construct an auxiliary network:

- Take our graph $G = (X \cup Y, E)$
- Direct each edge from X to Y
- Add a source vertex s with outgoing arc to each vertex in X
- Add a sink vertex t with an incoming arc from each vertex in Y
- Give each arc a capacity of 1

You can now treat this auxiliary network like a flow. It will have value k , since there are k distinct paths from s to t , each with value 1.

Now if you run the max flow algorithm, you'll actually be finding the flow that contains the most possible disjoint edges from X to Y (since we need to satisfy flow conservation). So this is actually a maximum matching, if you remove the source and sink!

29.2 Running Time of Max Matching

How long would it take to find a max matching in a bipartite graph? Recall this depends on the number of iterations, $O(m * \text{iterations})$. So how many iterations are there?

The maximum cardinality of a matching is at most the minimum of the size of X or Y , which will always be less than the number of vertices, n , so $O(mn)$.

29.3 Minimum Cut

If we put infinite capacity on the edges from X to Y auxiliary network we had before, it changes nothing. (Since the edges from s to X still have capacity 1). This means the min cut hasn't changed (by the maxflow-mincut theorem).

Recall, the maximum flow is at most n , so the capacity of the minimum cut S is at most n . In particular, the capacity of S is finite, meaning $\delta^+(S)$ contains no infinite capacity arcs.

29.3.1 Structure of Minimum Cut

$$\begin{aligned} \text{Let } X^* &= X \cap S \\ \Rightarrow \Gamma(X^*) &\subseteq S \cap Y \end{aligned}$$

since otherwise we'd have an infinite capacity arc leaving the cut.

So all the arcs in $\delta^+(S)$ are of the form (s, x_i) and (y_j, t) .

29.4 Vertex Cover Problem

Take an undirected graph $G = (V, E)$

There are only three types of edges in our bipartite graph. We can't have edges that leave the cut. So we can't have edges (X^*) to $(Y - \Gamma(X^*))$. This means that all the other edges are allowed, meaning if we took all the vertices in $X - X^*$ and $\Gamma(X^*)$ we have a vertex cover. Where the size of this vertex cover is the capacity of the cut.

But we said this was a minimum cut, so this is a minimum cover!

Therefore, if C is a minimum vertex cover, and M is a maximum matching, and they have the same size, then we know we have the correct answer.

29.5 Supply/Demand

If the source has a fixed supply b , and the sink has a demand, $-b$. Is there a flow that satisfies the demand constraints:

$$\text{flow} - \text{out}(s) = \text{flow} - \text{in}(t) = b$$

This is true \Leftrightarrow there is a flow of value at least b .

29.6 Multiple Sources and Sinks

Suppose there are $\{s_1, \dots, s_k\}$ sources and $\{t_1, \dots, t_l\}$ sinks.

Source s_i has supply b_i and t_j has demand b_j . Now we just need to satisfy:

$$\text{flow} - \text{out}(v) = \text{flow} - \text{in}(v) = b_v$$

for all $v \in V$, and the $b_v = 0$ for non-source/sink vertices.

The solution to this is to add a source supplying all the sources with supply the sum of the supply of the original sources, and all of the sinks into one vertex the same way.

Now we can just run a maximum flow and get a solution to the original problem.

29.7 Lower Bounds

Suppose we MUST send l_{ij} units on an arc (i, j) . Thus the capacity constraint on (i, j) becomes: $l_{ij} \leq f_{ij} \leq u_{ij}$.

To deal with this, simply force l_{ij} units along the arc, and then edit the supply/demand values:

$$b_i = b_i - l_{ij} \text{ and } b_j = b_j + l_{ij}$$

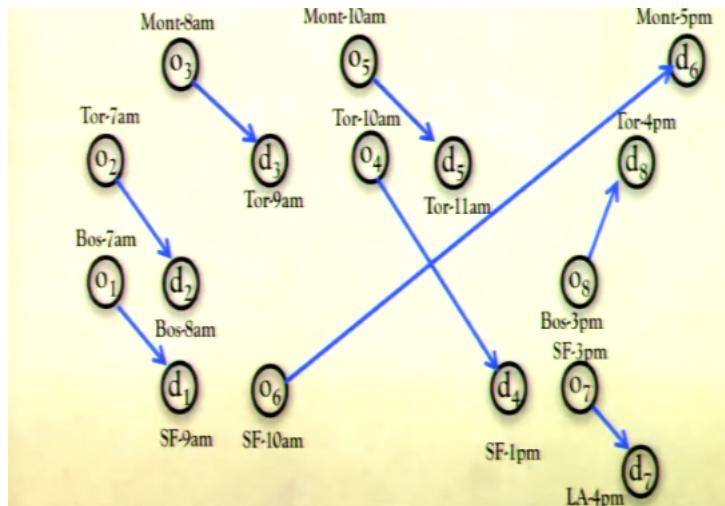
and now again just run the max-flow algorithm.

30 Applications of Max-Flow

30.1 Flight Scheduling

Suppose you have a list of flights with origins, destinations and times. How can we cover all these flights using the least number of planes possible? Given a schedule, can we satisfy it with k planes?

For each flight i we have a vertex for the origin o_i and the destination d_i . We add a source vertex s with a supply k . We add an arc (s, o_i) with capacity 1 for each flight i . We add a sink vertex with demand k . We add an arc (d_i, t) with capacity 1 for every flight. And we add an arc d_i, o_j with capacity 1 if it is feasible for a plane to service flight i and then j . Finally, each arc from origin to destination has a lower bound of 1.



If there's a flow that satisfies the schedule, then there is a schedule that works.

In reality, there's costs involved, crew, safety, flexible schedules, plane capacities. All of these constraints can be accounted for with max flows.

30.2 Open Pit Mining

We have a set V of blocks. Each block i has a profit π_i . We want to maximize the profit of the pit. (The profits can be negative)

We have some topological constraints. The pit cannot be too steep. Meaning to dig up a block i , we must first remove the 3 closest blocks in the layer above.

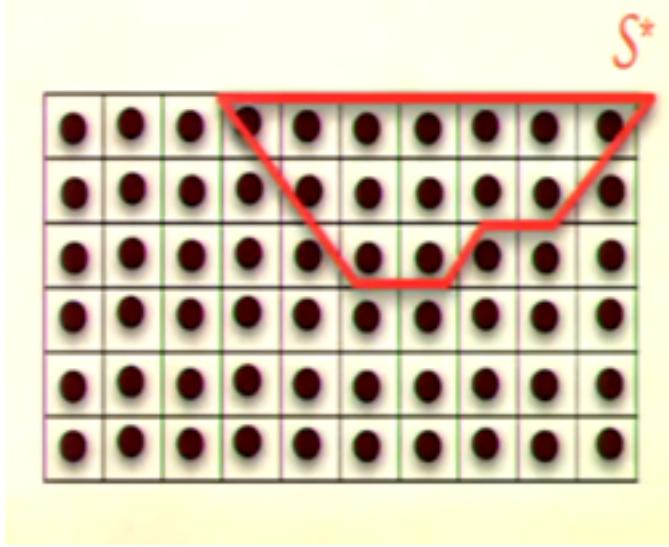
-20	-35	-10	-10	-20	0	-35	-10	-20	-15
10	15	-10	15	0	5	-10	30	90	-5
-5	20	-25	70	50	0	-30	10	-20	-15
0	5	0	0	-60	5	-90	-30	5	10
20	15	-10	15	80	50	-10	30	0	-5
0	10	-25	-10	10	20	-20	15	0	15

There is a vertex for each block. Have an arc (s, i) of capacity $\pi_i \Leftrightarrow \pi_i > 0$. There is an arc (j, t) of capacity $|\pi_j| \Leftrightarrow \pi_j < 0$

There is an arc from i to each of the 3 blocks above it, with capacity ∞ .

Observe that the capacity of s is finite, so that means the capacity of the minimum cut is finite. Which means the maximum flow is finite.

Suppose the block i is in the minimum cut. That means the blocks above it are also in the cut, since they have infinite capacity, and the cut must have finite. (They cannot leave the cut). This means that the minimum cut is a feasible pit!



This is actually the optimal pit, since:

$$\begin{aligned}
 cap(S) &= \sum_{a \in \delta^+(S)} u_a \\
 &= \sum_{i \notin S: \pi_i > 0} \pi_i + \sum_{i \in S: \pi_i < 0} |\pi_i| \\
 &= \left(\sum_{i \in V: \pi_i > 0} \pi_i - \sum_{i \in S: \pi_i > 0} \pi_i \right) + \sum_{i \in S: \pi_i < 0} |\pi_i| \\
 &= \left(\sum_{i \in V: \pi_i > 0} \pi_i - \sum_{i \in S: \pi_i > 0} \pi_i \right) - \sum_{i \in S: \pi_i < 0} \pi_i \\
 &= \left(\sum_{i \in V: \pi_i > 0} \pi_i - \sum_{i \in S} \pi_i \right)
 \end{aligned}$$

from this, since the left sum is constant, independent of S , that to minimize the capacity of S , we need to maximize the profit associated with it!

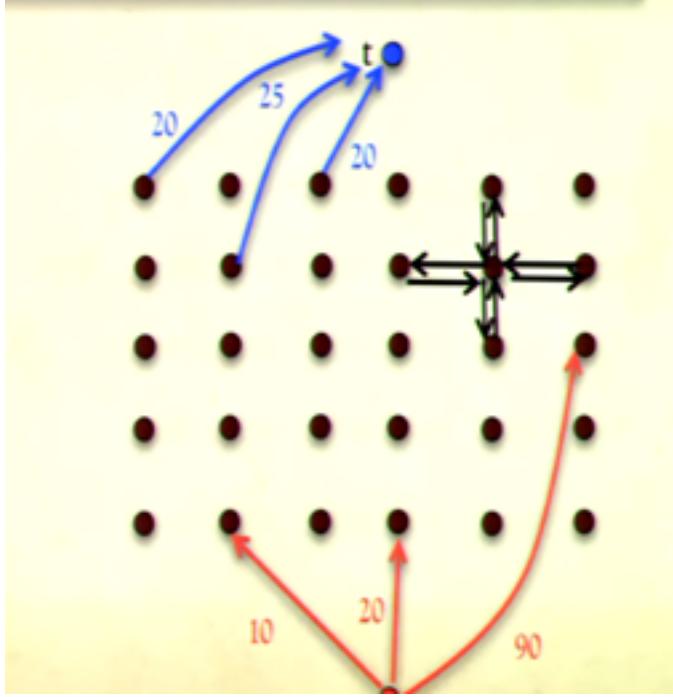
The key strategy in this application was that whenever you have a problem where you need to do something first. "The blocks above needed to be done first", use infinite arcs. Infinite arcs give priorities in flow problems.

30.3 Foreground- Background Segmentation

Let f_i be the likelihood that pixel i is in the foreground. And b_i that it is in the background. In general, if $f_i > b_i$ then we want pixel i in the foreground. But we also want a smooth

boundary between the fore and background.

A penalty ρ_{ij} is for separating pixels that are adjacent. There is a vertex for each pixel, and a source for the foreground, and a sink for the background. There is an arc (s, i) of capacity f_i and (j, t) of b_j . For adjacent pixels, there are arcs (i, j) and (j, i) with capacity ρ_{ij}



The capacity of a cut is:

$$\begin{aligned} \text{cap}(S) &= \sum_{i \notin S} f_i + \sum_{j \in S} b_j + \sum_{(i,j) \in \delta^+(S)} p_{ij} \\ &= (\sum_{i \in V} f_i - \sum_{i \in S} f_i) + (\sum_{j \in V} b_j - \sum_{j \notin S} b_j) + \sum_{(i,j) \in \delta^+(S)} p_{ij} \\ &\quad \sum_{i \in V} (f_i + b_i) - \sum_{i \in S} f_i - \sum_{j \notin S} b_j + \sum_{(i,j) \in \delta^+(S)} p_{ij} \end{aligned}$$

Again, notice that the far left term is independent of the cut, so the value depends only on the right three terms.

So we need to maximize the right three terms as a whole. This is exactly trying to find a minimum cut!

31 Flow Decomposition and Fast Flow Algorithms

Recall that the running time of the Ford-Fulkerson algorithm for finding the max flow, took time $O(m * \#iterations)$ but the number of iterations can be extremely large.

The algorithm does not specify which path to find, but does it matter which paths we choose, in terms of the number of iterations?

31.1 The Flow Decomposition Theorem

Recall that the union of a set of paths and a set of cycles satisfies flow conservation. Also, if the union of paths and cycles satisfy the capacity constraint, then they form a flow.

Theorem 28. Any $s-t$ flow can be decomposed into a collection of $s-t$ paths and directed cycles of cardinality at most m .

To prove this theorem, we'll need this claim:

Any flow with non-zero flow value contains a path or a cycle.

Proof. If f contains a cycle we're done.

So assume there are no directed cycles.

$$\Rightarrow f \text{ has flow value at least 1 since } f \text{ is non zero}$$

So there is at least one arc leaving the source with positive value. But then by flow conservation, the vertex receiving this arc must output the same flow. But since there are no cycles, it can never return to where it came from. So eventually it would have to reach the sink. The graph is finite so the process must terminate. \square

Now for the main theorem"

Proof. By induction on the number of arcs in the path.

If $m = 0$ then the flow is empty and has no paths or cycles. If $m = 1$ then f is just the arc (s, t) , and so this is one path.

Assume any flow with k arcs, $k < m$ arcs can be decomposed into a collection of at most k cycles and $s-t$ paths.

Take any flow with m arcs. By the claim, it contains at least one path or cycle. Take the case where it contains a cycle. Look at the arc with minimum flow in that cycle, and remove that flow from each arc in the cycle. We then lost the smallest arc from the cycle in the flow.

By the induction hypothesis, this new flow can be decomposed into at most $m - 1$ paths and cycles.

Now suppose we have a path, do the same thing we did for the path. We lose an arc, and by induction we can decompose into $m - 1$ paths and cycles. \square

The flow decomposition theorem tells us that the maximum flow can be decomposed into m paths and cycles. But since using a cycle doesn't increase the flow, don't use any cycles. So then our maximum flow must decompose into m paths.

But then if the algorithm selects the paths from the decomposition, then we'd have $O(m^2)$. But how do we choose the correct paths?

31.2 Finding the correct paths

We could try the greedy approach, picking the path that would increase the flow the most at the current step.

This actually gives a polynomial time algorithm for max flow.

First, observe that if f^* is a maximum flow, then there is a path with capacity at least $\frac{1}{m}|f^*|$ since there are at most m paths.

Next, if f is a flow, and f^* is the max flow, then in the residual graph G_f there is a path P with bottleneck capacity:

$$\geq \frac{|f^*| - |f|}{m}$$

since, $(f^* - f)$ satisfies flow conservation, so it can be decomposed into at most m paths.

So at least one of these paths carries one $m - th$ the difference in the flow values.

Theorem 29. The maximum capacity augmenting path algorithm terminates in at most $m(\ln(n) + \ln(U))$ iterations where U is the maximum capacity arc.

Proof. Let the algorithm find the paths $\{P_1, P_2, \dots, P_T\}$ we want to show that $T \leq m(\ln(n) + \ln(U))$

Let f_t be the flow found after t iterations. By the previous observation, the next step will add at least $\geq \frac{|f^*| - |f_t|}{m}$ to the flow value.

So the amount that we can add afterwards, $\Delta_t + 1$:

$$\begin{aligned} &\leq \Delta_t - \frac{1}{m}\Delta_t \\ &= (1 - \frac{1}{m})\Delta_t \\ &\leq (1 - \frac{1}{m})(1 - \frac{1}{m})\Delta_{t-1} \end{aligned}$$

$$\leq \left(1 - \frac{1}{m}\right)^t \Delta_1$$

But Δ_1 is the maximum flow.

Now recall that $(1 - x) < e^{-x}$

$$\begin{aligned} \Rightarrow \Delta_{t+1} &\leq \left(1 - \frac{1}{m}\right)^t |f^*| \\ &< \left(e^{-\frac{1}{m}}\right)^t |f^*| \\ &= e^{\frac{-t}{m}} |f^*| \end{aligned}$$

now setting $t = m \ln |f^*|$ gives:

$$\begin{aligned} \Delta_{t+1} &< e^{-\frac{m \ln |f^*|}{m}} |f^*| \\ &= 1 \end{aligned}$$

So, this means we're done, since the quantity of flow remaining to be found is less than one.
Now the maximum flow has value at most nU so:

$$\ln |f^*| \leq \ln(nU) = \ln(n) + \ln(U)$$

So the number of iterations is at most $m(\ln(n) + \ln(U))$ □

However, the time per iteration has now changed! Since we now need to find the capacity of ALL paths to choose the fastest one.

Theorem 30. The maximum capacity augmenting path algorithm takes $O(m^2)$ time per iteration.

Proof. Label the arcs $\{1, 2, \dots, 2m\}$ in the residual graph in decreasing order of residual capacity.

We can test if there is an s-t path using only arcs in $\{1, 2, \dots, k\}$ in $O(m)$.

We can do this for all k in time $O(m^2)$. The maximum capacity augmenting path is the path we find using the smallest k for which $s - t$ path exists using only arcs $\{1, 2, \dots, k\}$ (Since they were in decreasing order) □

We can actually speed that up using binary search on k , and computing paths with k arcs based on if the previous value of k was able to find a path or not.

This algorithm is weakly polynomial, since $O(m^3(\ln(n) + \ln(U)))$, it depends on U , which may be very large, and is not part of the input size.

Is it possible to find a strongly polynomial time algorithm?

31.3 Strongly Polynomial

If we choose the shortest length path in each step, (using BFS or shortest paths), which actually runs in $O(mn^2)$. (Not proved here)

Part VI

Data Structures

We've already looked at many data structures in COMP250, this is just a continuation and adding detail to things.

32 Heaps

Heaps are used to implement a priority queue. Each element of the set has a priority.

It supports these operations:

- Insert($S.x$) add new element x to S
- Min(S): Find highest priority element
- ExtractMin(S): Find highest priority element and delete it

We could use a sorted linked list or array to implement this, in which case we could $\text{Min}(S)$ in $O(1)$, but then $\text{Insert}(S.x)$ is $O(n)$. This is bad!

A heap will allow us to perform all these operations in $O(\log n)$.

A heap is a complete binary tree which satisfies: $\text{key}(x) \geq \text{key}(\text{parent}(x))$ for all x .

The depth of a heap is logarithmic to the number of nodes! (Measure of the running time of the operations).

Notice, if we number the nodes of this tree, we can actually represent this as an array! For any node l , $\text{parent}(l) = \text{floor}(\frac{l}{2})$, and $\text{leftchild}(l) = 2l$, and $\text{rightchild}(l) = 2l + 1$.

Notice that in a heap, the key of a vertex is less than or equal to every vertex in the subtree of that vertex. So the minimum key must be at the root.

32.1 Insertion and UpHeap

To insert a new key, we need to add it at the bottom left to keep the tree complete. We then need to UpHeap it to the correct position.

While(key(x) < key(parent(x)))
 Swap x with parent.

We have $O(\log(n))$ swaps to do.

32.2 Extract Min and DownHeap

To extract the root, swap it with the last element in the array. Then just remove the last element. Next, we need to keep swapping it with the smallest of its children until it satisfies the heap property. Again this is done in $O(\log(n))$ swaps.

32.3 Building a Heap

How quickly can we build a heap? We could just apply the insert operation n times, which would take time $O(n \log(n))$

We can actually take two heaps, and merge them together by using a new element as the new root. Then the only place it can go wrong is at this new root, which we can just downheap.

So what we'll do is just add all the elements to the heap in no particular order. Then, we can turn the leaves into heaps (trivially) and keep merging sub-heaps together until we have a full heap!

32.3.1 Building A Heap is Linear time

The number of vertices at depth d is at most 2^d . A vertex at depth d , is at distance at most $\log(n) - d$ from the leaves.

When we apply Heapify down, we may compare it to $2(\log(n) - d)$ since we compare to each child.

So the total number of comparisons is:

$$\sum_{d=0}^{\log(n)} 2^d * 2(\log(n) - d)$$

now let $d = \log n - k$

$$\begin{aligned} &= \sum_{k=0}^{\log(n)} 2^{\log n - k} 2k \\ &= \sum_{k=0}^{\log(n)} n \frac{k}{2^{k-1}} \end{aligned}$$

$$\begin{aligned} &\leq n \sum_{k=0}^{\infty} \frac{k}{2^{k-1}} \\ &= n * \frac{1}{1 - \frac{1}{2}}^2 \end{aligned}$$

(by taking the derivative of the geometric series where $x = 1/2$)

33 Applications of Heaps

33.1 HeapSort

We can sort n numbers into increasing order using a heap.

The steps are:

- Build Heap (using the method outlined previously)
- Repeatedly remove minimum (correcting using DownHeap)

This takes time $O(n\log n)$. (n for building the heap, then $n\log n$ for applying DownHeap ($O(\log n)$), n times.)

33.2 Shortest Path Problem

Recall Dijkstra's algorithm for shortest paths (see section 12 for a refresher). That algorithm takes time $O(n^2)$, but using heaps, we could actually get $O(m\log n)$!

If we make the keys of the heap be the distance label of the arcs, then in the algorithm, when choosing the minimum label, we can simply extract the minimum from the heap! Whenever we update the labels, we can simply update the key in the heap.

Updating the key of some vertex in the heap to a lower value, is just a matter of changing the key, and performing UpHeap. But, if we change the position in the heap, we should keep track of its original location.

We build the heap once, extract the minimum n times, and update m times, so $O(n + n\log n + m\log n) = O(m\log n)$.

33.3 Minimum Spanning Tree Problem

Recall Prims algorithm for finding minimum spanning tree. (See 14.2). Recall that this algorithm takes time $O(mn)$, however, using heaps we can do this in $O(m\log n)$.

The keys of the vertices are the minimum cost of that vertex to the tree. At each iteration, we add to T the minimum vertex from the heap. After adding this vertex, we may

need to decrease the keys of the vertices adjacent to v . So all in all $O(n \log n + m \log n) = O(m \log n)$

33.4 Data Compression Problem

Recall the Huffman coding algorithm. (See section 13) It took $O(n^2)$ but again we can do it in $O(n \log n)$. The heap here is that the keys are the frequencies, and at every iteration we take the two smallest keys, sum them, and add the result to the heap. There are $n - 2$ iterations, so $O(n \log n)$.

34 Dictionaries and Hash Tables

For a more thorough overview of hash tables, see my COMP250 guide. This gives the main points.

34.1 Dictionaries

A dictionary is a data structure that is good for:

- Insert(D, x)
- Lookup(D, x)
- Delete(D, x)

for an element x . Recall that the priority queue can only search and delete the smallest element. Here it's ANY element.

34.2 Hash Table

A hash table implements a dictionary, and can do the above operations in typically $O(1)$ time.

There is a very large universe U of possible keys, but we only expect to encounter a smaller number n of the keys. We store the keys in a table T of cardinality m . Usually n is roughly m .

34.2.1 Hash Function

To do this we have a hash function $h : U \rightarrow T$ that assigns each key k to a slot in the table. $h(k) \in \{0, 1, \dots, m - 1\}$

Since the universe is much larger than m , there will be collisions (by Pigeonhole Principle). This means a slot in the table is assigned to multiple keys. To handle this, we use a linked list in each slot.

So the running time is proportional to the length of the corresponding linked list. So we want to minimize the length of these lists.

The expected length of a chain is $\frac{n}{m}$ and since m is roughly n , this ratio is constant. However, this only applies if the hash function is well chosen, (minimizes collision) or the data doesn't fit the hash function. The worst case chain length can be huge $\Omega(n)$.

34.2.2 Universal Hash Functions

A function has the universal hash function property if the probability that it causes a collision is $\frac{1}{m}$.

34.2.3 Random Hash Functions

For any key, chose a slot $\{0, 1, \dots, m - 1\}$ then the probability that they are assigned to the same slot is $\frac{1}{m}$.

But with a random hash function how do we keep track of which function was kept with each key? Then we'd need to store the function, but then we'd need a good storage system to store our good storage system! (Doesn't work!)

34.2.4 Deterministic Hash Functions

But for a deterministic hash function there is no probability! So there will always be collisions. DoS attacks usually exploit this.

We'll define a family of deterministic functions.

We'll choose a hash function from a class of deterministic functions!

Represent the keys U in $[m]$ (numbers up to m), by $x = x_1x_2x_3\dots x_l$ where x_i is an integer in $[m]$. Then, we'll define the function:

$$h(x) = (\sum_{i=1}^l a_i * x_i) \bmod p$$

where p is a prime number, and a_i is randomly chosen.

34.2.5 Proof That It Works

Theorem 31. For any $x \neq y$ the probability that $h(x) = h(y) = \frac{1}{m}$

Proof. As $x \neq y$ we must have that $x_\tau \neq y_\tau$ for some index $1 \leq \tau \leq l$

Now, to see if there's a collision, then we need that:

$$\left(\sum_{i=1}^l a_i * x_i \right) \text{mod } p = \left(\sum_{i=1}^l a_i * y_i \right) \text{mod } p$$

which means that:

$$a_\tau x_\tau + \left(\sum_{i \neq \tau} a_i * x_i \right) \text{mod } p = a_\tau y_\tau + \left(\sum_{i \neq \tau} a_i * y_i \right) \text{mod } p$$

which means that:

$$a_\tau (x_\tau - y_\tau) \text{mod } p = \left(\sum_{i \neq \tau} a_i * (y_i - x_i) \right) \text{mod } p$$

If we pick a_τ last, then we've already picked all the other a_i , x and y are fixed, so the sum must be fixed. Call it β .

Therefore they collide if: $a_\tau (x_\tau - y_\tau) \bmod p = \beta$. We now want to show that only one possible a_τ will cause a collision.

Suppose there were two, then:

$$(a_\tau (x_\tau - y_\tau)) \text{mod } p \equiv a'_\tau (x_\tau - y_\tau) \text{mod } p$$

so:

$$(a_\tau - a'_\tau)(x_\tau - y_\tau) = 0 \bmod p$$

This means that p divides the product, which means that p divides either factor.

But all these numbers are between 0 and $m - 1$, so $|a_\tau - a'_\tau| < m$ and $|(x_\tau - y_\tau)| < m$.

So the only way this can work is that $|a_\tau - a'_\tau| = 0$, since $x_\tau \neq y_\tau$.

This means there's only one choice of a_τ , which has probability $\frac{1}{m}$

□

34.3 String Searching

Does a string $x = \{x_1 x_2 \dots x_n\}$ contain a substring $s = \{s_1 s_2 \dots s_l\}$ where $l \ll n$?

The easiest way to do this is check the first l letters, then $x_2 x_3 \dots x_{l+1}$ and so on.

We'll use a hash function, and check to see if the substrings have the same hash value,

if they do, then we'll actually check string equivalence. This saves having to iterate through each substring to check for equality over and over.

But this takes $O(ln)$ time, how can we do it faster?

We can represent each string by a number with base equal to a large prime. For example: "hi" $[base101] = 104 * 101^1 + 105 * 101^0 = 10690$ where the 104 and 105 are the ASCII codes for h and i.

In general:

$$x_i x_{i+1} \dots x_{i+l-1} [basep] = x_i p^{l-1} + x_{i+1} p^{l-2} \dots x_{i+l-1} p^0$$

and to find the next shift over:

$$x_{i+1} x_{i+2} \dots x_{i+l} = (x_i x_{i+1} \dots x_{i+l-1} - x_i p^{l-1}) p + x_{i+1} p^0$$

To see why this is useful, consider the case of finding the substring "hello" in the string "adfshellouaf".

Suppose we've computed the section of 5 characters (the length of hello): "shell". Then to find the next section of 5 characters, we don't need to process all l of them, since we can just do:

$$hello[base101] = ("shell" - "s" * 101^4) * 101 + "o" * 1$$

This lets us check the hash functions in $O(1)$ time. So $O(l + n) = O(n)$.

So we then need to verify each substring in $O(l)$ time, but we assume our hash function is good, so we only do this a constant amount of times.

34.3.1 Searching for Multiple Strings

What if we want to find k different substrings s in a string x ?

We do the same thing, but we compare our hash value for x to all hash values of s_k . So $O(kn)$

35 Binary Search Trees

These are used to implement both a priority queue and dictionary at the same time! It supports the operations:

- Insert
- Search

- Delete
- FindMin
- FindMax
- predecessor
- successor

We'll see why the last two are important a little later.

35.1 BST Property

A tree is a BST if for any vertex x , it's left subtree L_x has the property that all of it's elements are less than x , and that it's right subtree R_x has all of it's elements greater than x .

35.2 Operations and Observations

35.2.1 Search(r, k)

To look for an element k , start at the root r . If $\text{key}(r) > k$, search the left subtree, else if $\text{key}(r) < k$, search the right subtree.

This takes time $O(d)$ where d is the depth of the tree. (Might be large if tree is unbalanced! More on that later.)

35.2.2 Insert(x)

Inserting is a matter of copying the search operation until we find a leaf, and then inserting the element there. This will ensure that the element is placed at the correct spot.

35.2.3 Building a BST

We could just repeatedly insert, but could be built in many different shapes, depending on the order in which we insert.

35.2.4 Find-Min and Find-Max

To do this, we start at the root, and keep going left until we cannot possibly go left anymore, that is, the left child is empty. Similarly, to find the max we start at the root and keep going right until we can't anymore.

35.2.5 Sorting Using a BST

If we build a BST, we can sort its elements using in-order DFS. If you remember, we can model DFS by the caterpillar crawl around the tree. (See section 10.4)

We know we've searched a subtree using this walk when we return to a vertex we've previously visited. So, to sort the tree, simply output the key once you've searched the leftmost subtree!

35.2.6 Successor and Predecessor

We define $\text{pred}(k)$ to be the predecessor of k when $\text{pred}(k)$ is directly preceding k in the sorted list of elements. So if $\{1, 2, 3, 4, 5\}$ is the sorted list of elements, 3 precedes 4. Similarly, the $\text{succ}(k)$ is the element directly succeeding k .

Finding the predecessor of the root is easy. Simply find the max of L_r , since nothing in R_r can possibly be smaller than r , we don't even look at it. Similar for the successor of the root, find min on R_r .

Observation: Either $\text{pred}(v)$ is a descendant of v or v is a descendant of $\text{pred}(v)$. To see this, suppose not. Then they would have a common ancestor, x . And without loss of generality: $\text{pred}(v) \in L_x$, $v \in R_x$. But then this means that $\text{key}(\text{pred}(v)) < \text{key}(x) \leq \text{key}(v)$ which is impossible, since then x would be the $\text{pred}(v)$.

Claim: If v has a left child then $\text{pred}(v)$ is a descendant of v .

Proof. By contradiction:

Let x be an ancestor of v with $\text{key}(x) < \text{key}(v)$. This means that $v \in R_x$. Let y be the max of L_v (exists since v has a left child). y is a descendant of v , v is a descendant of x .

This means that $y \in R_x$, but then

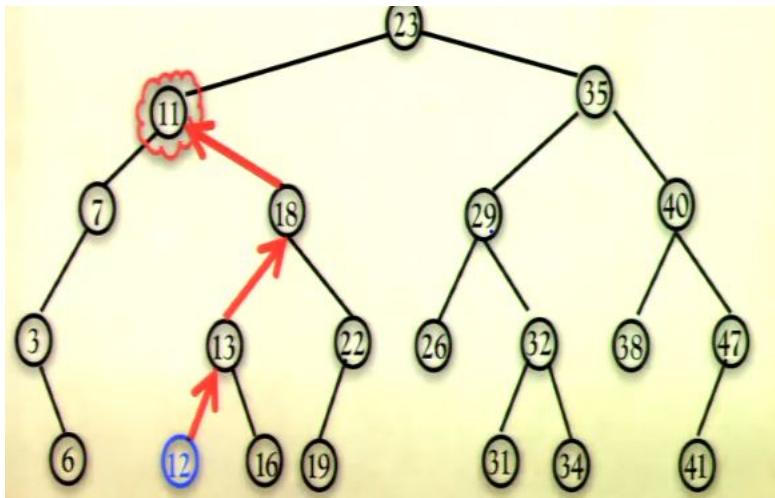
$$\text{key}(x) \leq \text{key}(y) \Rightarrow \text{key}(x) \leq \text{key}(y) < \text{key}(v)$$

so y is the predecessor of v , which is a contradiction. \square

So now to find the predecessor of a non-root vertex, we just have a few cases:

Case 1: v has a left child: Then $\text{pred}(v)$ is the vertex with the maximum key in L_v .

Case 2: v has no left child: Then $\text{pred}(v)$ must be an ancestor of v . (by the observation). But since v must be just greater than $\text{pred}(v)$ this means that v has the minimum key in $R_{\text{pred}(v)}$. So we can walk up-wards to the right from v until we must go left. And then we've found $\text{pred}(v)$.



35.2.7 Delete

It's easy to delete a vertex if the vertex we're deleting is a leaf, since removing it won't affect the BST property.

If there is one child, we can simply connect the child(v) to the parent(v). This is known as "splicing". This maintains the BST property.

If there are two children, then we need to find $\text{pred}(v)$. Replace v by $\text{pred}(v)$. If $\text{pred}(v)$ has no children, we need not make any adjustments. If it has one child, then we can splice it out to move it.

The $\text{pred}(v)$ will never have more than one child since it is the maximum in L_v (since v has two children). This means $\text{pred}(v)$ definitely doesn't have a right child. So it can be spliced.

Further, replacing v by $\text{pred}(v)$ is safe since in the sorted list they compare the same way to every other element.

35.3 Balanced BST's

To maintain the speed of operations on our BST, we need to do some extra work in maintaining the structure. We want as many vertices to have two children as possible, so that the depth of the tree is as close to $\log(n)$ as possible. This is done using Red-Black trees, AVL trees, or Splay Trees. (Not covered in this course)...

36 Data Structures for Disjoint Sets

In Kruskal's algorithm for a minimum spanning tree, the runtime was $O(mn)$. But we can improve on this!

Initially every vertex of the input graph is in its own component. When edges are added, we're taking disjoint components of the tree, and joining them together.

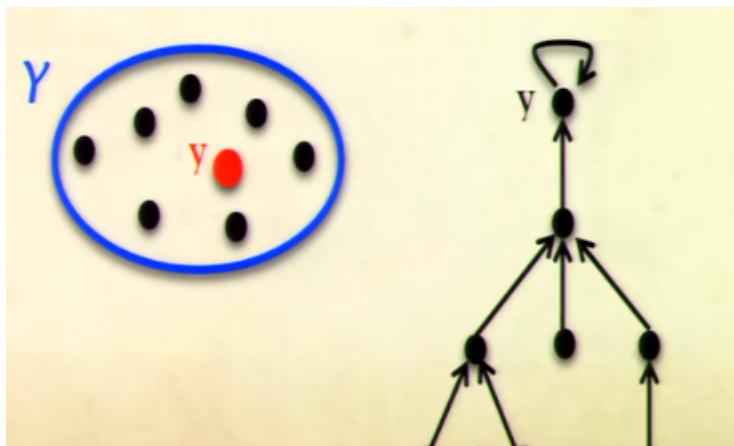
The data structure for disjoint sets supports:

- $\text{Make-Set}(v)$: creates a new set with a single element v .
- $\text{Find-Set}(v)$: find the set containing v .
- $\text{Merge}(u,v)$: merge the two sets which contain u and v .

36.1 Set-Representatives + Storage

To keep track of the sets we will identify each set by a unique representative.

To store the sets, we use directed trees. (Root nodes point to themselves and are the set representatives.) For example:



36.2 Operations

36.2.1 Make-Set(v)

Simply add a new node v with a pointer to itself.

36.2.2 Find-Set(v)

Since each element is a member of a tree for a set. Then to find the set we are in is to find the root of the tree. Simply follow the path from v to the root.

36.2.3 Merge(u,v)

Suppose we have two sets U containing u and V containing v . To implement this, we must first Find-Set(v) and Find-Set(u). We then make the parent of the root of U to be the root of V . (or the other way around).

36.3 Runtime

Find-Set(v) is the bottle-neck, since make-set takes constant time, and merge uses find-set and then has some constant operation. The running time of find-set depends on the length of the path from the vertex to the root. This is just the depth of the tree.

We must then make the depth of the tree small. It's possible that after several merge operations the tree will have a large depth.

36.4 Union By Rank

For the merge operation, the root of the tree with the bigger depth becomes the parent of the root of the tree with the smaller depth.

We must then store the depth of each root vertex. It's easy to keep track of this, simply initialize during the make-set operations, and update during the merge.

36.5 Kruskals Revised

Make-Set for all $v \in V$

Sort the edges $\{e_1, \dots, e_m\}$ by cost, $c_1 < c_2 \dots < c_n$.

Set $T = \emptyset$

For $i = \{1, \dots, m\}$

let $e_i = (u, v)$ if(Find-Set(u) \neq FindSet(v)) then Merge(u, v) and set $T \leftarrow T \cup e_i$.

36.6 Upper-Bound on Tree Depths

Theorem 32. A tree with n nodes has depth $O(\log(n))$ if we apply Union by Rank.

Proof. By induction on n . Base case: A tree with $n=1$ nodes has $d = 0$. Induction hypothesis: Assume tree $2^d \leq n$ for any tree with n nodes.

Take any tree T with $n + 1$ nodes. Assume that T was created by merging the two trees T_1 and T_2 with cardinalities n_1 and n_2 and depth d_1 and d_2 . Then by the induction hypothesis we have $2^{d_1} \leq n_1$ and $2^{d_2} \leq n_2$. Then without loss of generality we have two cases:

1) $d_1 > d_2$ then $d = d_1$, so:

$$2^d = 2^{d_1} \leq n_1 \leq n + 1$$

2) $d_1 = d_2$ then just pick either one as the new root, so $d = d_1 + 1 = d_2 + 1$ so:

$$2^d = 2^{d_1+1} = 2 * 2^{d_1} = 2^{d_1} + 2^{d_2} \leq n + 1$$

□

36.7 New Running Time for Kruskals

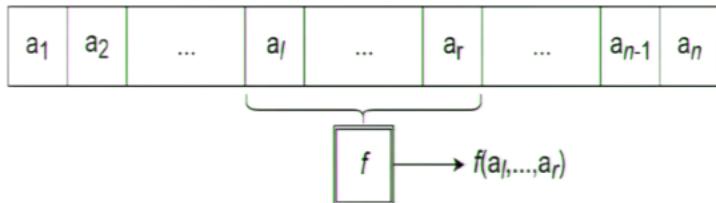
So the new running time for Kruskals algorithm is: $O(m\log(n))$.

36.8 Path Compression

After traversing a path P during a find-set operation, let each node on P now have a parent pointer to the root. Turns out to be quite a bit faster than our other method. Can be proven that it is $O(m\log^*(n))$ to find-set. So finding the minimum spanning tree is easier than sorting!

37 Data Structures for Range Queries

Given a list of integers $A = \{a_1, \dots, a_n\}$ and a list of queries: $\{\{l_1, r_1\}, \dots, \{l_q, r_q\}\}$. Each query is a pair of indices that defines an interval (range) on the list of integers. We want to efficiently compute some function $f(a_l, \dots, a_r)$ for each query.



We want to basically apply the function to each element of the list within the range of the query.

37.1 Range Min/Max Problem

Here f is finding the *min* or *max* of A within the range (l, r)

With brute force, we could just iterate through each element a_i to a_l to find the min/-max. But this takes time $O(n)$ for each query, so for q queries takes $O(nq)$.

37.1.1 Pre-Computation

If we could store the answer to every possible query beforehand, then accessing would be very fast.

There are $O(n^2)$ pairs of indices that can define ranges. So to compute the max/min by brute force on $O(n^2)$ would take $O(n^3)$.

It's actually possible to pre-compute all $O(n^2)$ values in $O(n^2)$ using dynamic programming!

$$\min(a_l, \dots, a_r) = \min(\min(a_l, \dots, a_{r-1}), a_r)$$

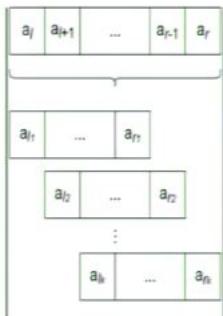
$$\min \left\{ \min \left[\begin{array}{|c|c|c|c|} \hline a_l & a_{l+1} & \dots & a_{r-1} \\ \hline \end{array} \right], a_r \right\}$$

So our computation algorithm is a matter of iterating through all possible pairs, and for each pair, store: $\min(\min_{prev}, a_j)$ where prev is the result of the previous iteration. And j is the current end-index.

If we have many many queries, this saves us time later. But if $q \ll n$ then this does worse!

37.1.2 Overlapping Ranges

We can improve on this, since many overlap!



So if you somehow know which pieces overlap, and that the union of the pieces make up a whole interval (l, r) then if we've computed the min for the sub-ranges, we can be faster!

$$\min(a_l, \dots, a_r) = \min(\min(a_{l_i}, \dots, a_{r_i}))$$

for $i = 1, \dots, k$.

So we can find a small collection S of ranges such that every range can be broken up into these small set of ranges.

We then precompute the min/max for ranges of S . Then, for every query range, (l, r) , find the set of ranges T in S who's union forms (l, r) . Find the minimum among the pre-computed values of T .

37.2 Sparse Table

Construct a table of $\log(n)$ rows and n columns. Number the rows, and on the i th row, store the minimum value in all ranges of length 2^i .

Suppose $A = \{5, 2, 3, 8, 1, 4, 7, 6, 9\}$.

0	5	2	3	8	1	4	7	6	9
1	2	2	3	1	1	4	6	6	
2	2	1	1	1	1	4			
3	1	1							

Notice that the first row $i = 0$ is just A . Then, to fill the row $i = 1$, for each value, simply look at the values above you to the right, in a range of 2^i .

More precisely:

$$S(i, j) = \min(S(i - 1, j), S(i - 1, j + 2^{i-1}))$$

We can precompute this table in $O(n \log n)$ time.

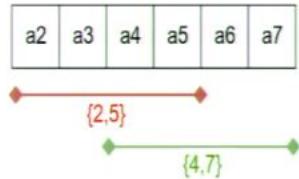
Every range $\{l, r\}$ can be broken up into two (possibly overlapping) ranges whos length is the largest power of 2 smaller than the length of the range.

The length of any range is $(r - l + 1)$, so let $p = \lfloor \log(r - l) \rfloor$. The ranges $\{l, l + 2^p - 1\}$ and $\{r - 2^p + 1, r\}$ are in the p th row of the table and will cover the range (l, r) .

So we can just lookup the minimum values of these two ranges and take the minimum of the two.

For example:

- Suppose the query range is $\{2, 7\}$, so we want to find $\min(a_2, a_3, a_4, a_5, a_6, a_7)$.
- The length of this range is 6, so $p = \lfloor \log_2 5 \rfloor = 2$.
- The ranges $\{2, 5\}$ and $\{4, 7\}$ have length 4 (the largest power of 2 less than 6), and cover the range $\{2, 7\}$.



Each query takes constant time, and so the total running time for precomputation and q queries is $O(n \log n + q)$.

37.3 Range Sum Query Problem

We need to compute:

$$\sum_{i=l}^r a_i$$

The sparse table can be build in a similar way, but we can't really query it. If the sub-ranges overlap then the sum of the query range is not the sum of the sub-ranges.

37.4 Segment Trees

These are actually better for both the sum and the min/max problem. Can do operations in $O(\log n)$. Operations include: adding an integer into a range, and changing every integer.

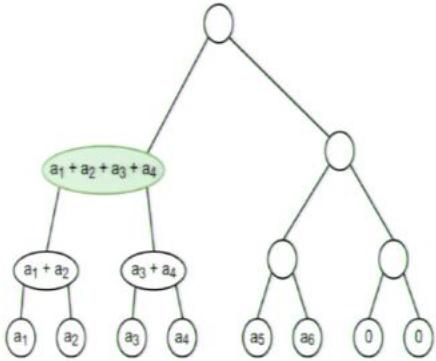
It does these operations while simultaneously querying the tree between these updates

We won't look at how these operations are performed in this course.

To make things easier, pad the input list with zero's, up to the nearest power of two.

To build one, construct a complete binary tree of height $\log_2(n)$. Then there are exactly n leaves, and $2n - 1$ nodes. In the i th leaf node, counting left to right, store the value of

$A[i]$ which is a_i or zero since we padded. For each non-leaf node, store the sum of the values of the leaves in the subtree rooted at that node.



As an exercise, write an algorithm (recursive) to construct a segment tree, which takes $O(n)$ time.

To query the tree, start at the root, and form a path down to a_l . To see how it works, suppose n is size 16, and the range is from 4 to 12. then we know that a_l is somewhere in the left subtree, since $4 < 8$. (8 is half the input size, half the leaves will be on either side). We then divide the input size by two, (now 8). We see that 4 must be in the left subtree. We divide the input size again, (now 4), and we see that it must be in the right subtree. And so on.

Mark this path red.

We do the same for a_r . Mark it green.

Now for each element of the red path, if it has an unmarked right-child, mark this child blue. Symmetrically, check the elements of the green path and mark the unmarked left-children blue.

If we now sum the blue nodes, with the value of a_l and a_r .

The total running time is now $O(n + q \log(n))$.

Write a recursive $O(\log(n))$ to perform a query on a segment tree as an exercise.

38 Learning

38.1 Multi-Armed Bandits

You visit a casino with multiple slot machines. Let's assume each slot machine pays out at its own rate. You want to gamble on the machine with the highest payoff rate.

Which machine do you use?

Could try doing the experiment many times to learn which machine is the best, but we'll lose money using bad machines.

Assume the machines payoff randomly.

38.2 Expert Advisors

There are n expert advisors advising Alice on her retirement plan. Every year she asks an advisor what to do.

How can we figure out which advisor will give the best advice? (Exactly the same as the previous problem)

38.2.1 The Learning Model

There is an agent with a set A of n actions. There are T time periods with an unknown cost function c^t at time t .

We want to minimize costs.

At each period t the agent picks an action $a^t \in A$. The cost is then $\sum_{t=1}^T c^t(a^t)$