COMP 2804 - Introduction.

Today we will do Chapter 1 from the text book.

Purpose is to remind you of COMP 1805 concepts.

Also - introduce problems that we will later solve using probability.

Ramsey's Theory: the idea that structures emerge once problem is big enough.

Problem 1: 6 people are at a party $P_1, P_2, \ldots, P_6$

Any two people must be either friends or strangers.

Example: P_2, P_4 either know each other or do not.

Claim: $\exists$ group of 3 friends or group of 3 strangers in any group of 6 people.

How do we prove this?

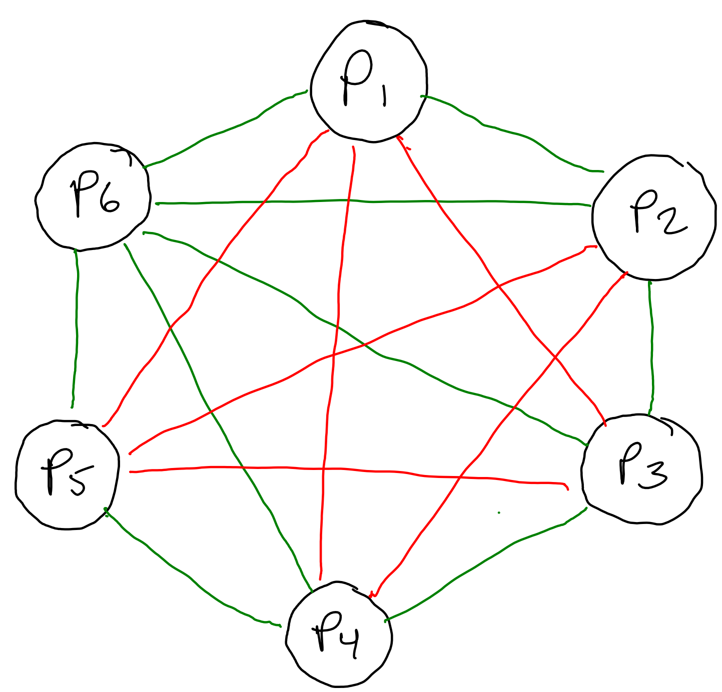
First we must model the problem.

When solving real world problems in math or computer science, we must first model it.

What is a good model? Graph.

Total number of edges?

$5+4+3+2+1$
$=15$



An equivalent claim:

There is a red triangle or a green triangle.

In this example it is true

P_1, P_3, P_5 is a red triangle

P_2, P_3, P_6 is a green triangle

Specifically the complete graph K_6

Red edge - strangers

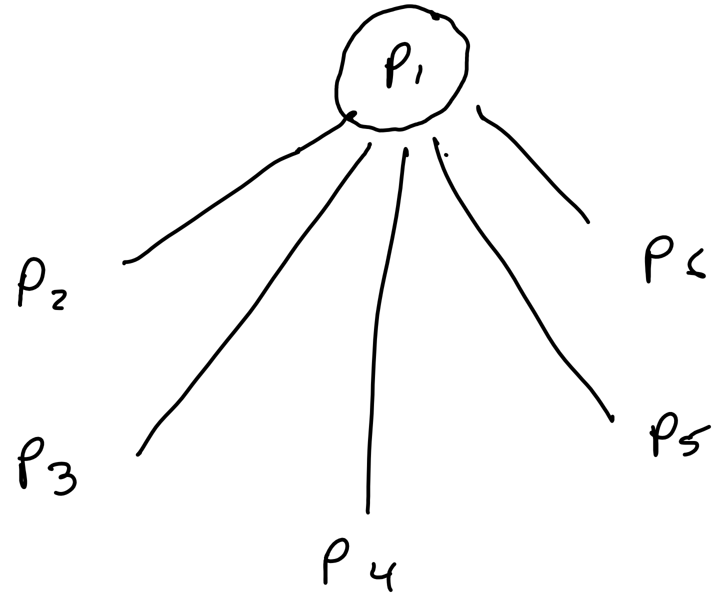
Green edge - friends

In this example both are true
only need one

Example is not a proof.

Consider a single vertex :

P_1 has 5 edges.



K_6

$\boxed{\text{Claim: } P_1 \text{ has at least 3 red or 3 green edges.}}$

Prove (by contradiction):

1. $\neg(P_1 \geq 3$ green edges $\vee$ $P_1 \geq 3$ red edges$)$
2. $\neg(P_1 \geq 3$ green$) \wedge \neg(P_1 \geq 3$ red$)$
3. $P_1 < 3$ green edges $\wedge$ $P_1 < 3$ red edges.
4. P_1 has ≤ 4 edges
5. False

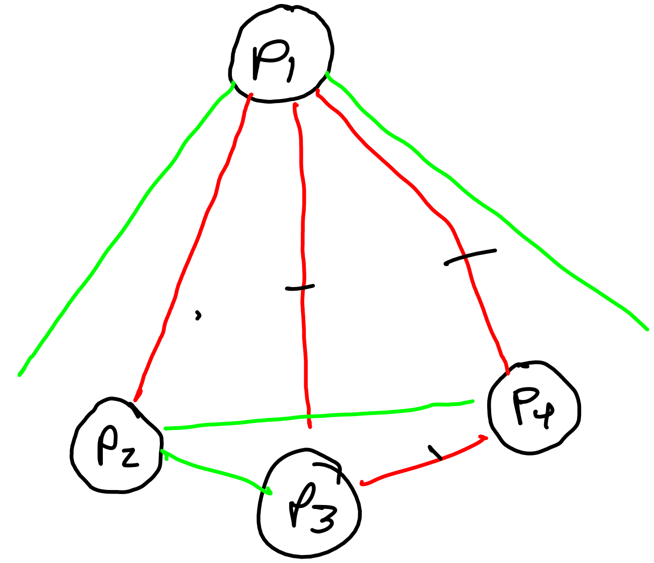
Without loss of generality, assume P_1 has ≥ 3 red edges.

P_2, P_3, P_4

We will examine 2 cases:

① All three edges are green.

② At least one edge is red.



Case 1: All 3 edges are green.

$\rightarrow \exists$ a green triangle.

Case 2: At least 1 red edge

a) if it is P_3P_4, then $P_1P_3P_4$ is a red triangle

b) if it is P_2P_3, then $P_1P_2P_3$ is a red triangle

c) if it is P_2P_4, then $P_1P_2P_4$ is a red triangle.

$\therefore$ at least 1 red triangle.

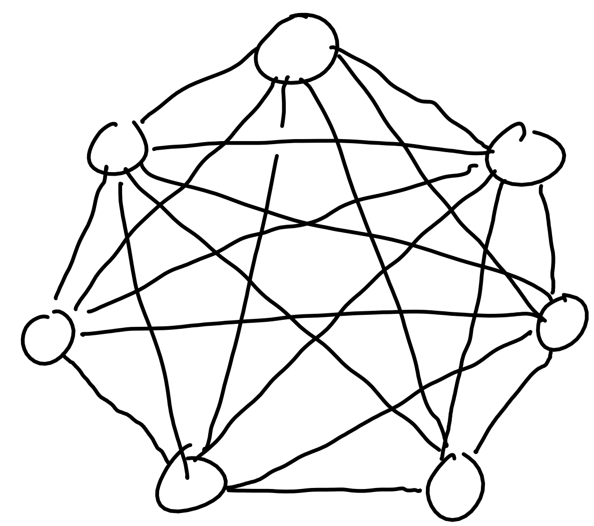
$\therefore$ $\exists$ at least 1 red triangle or 1 green triangle.

$\therefore$ For any group of 6 people, there is a group of 3 strangers or a group of 3 friends.

What about a group of 7 people?

In our model we make it into a graph – K_7

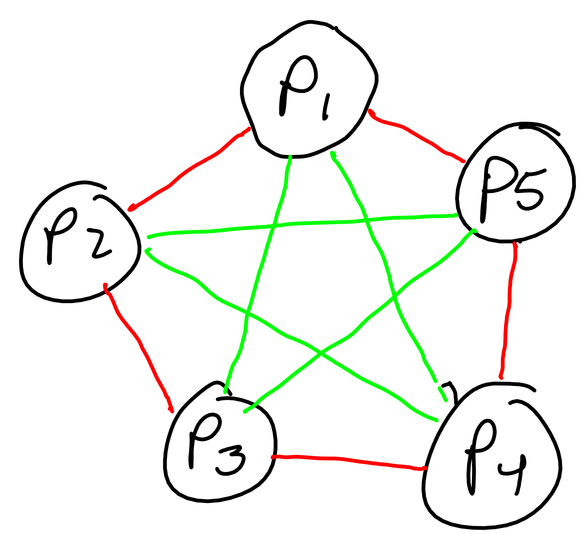
Does this claim hold for K_7 and above?



Yes because K_7 has a copy of K_6.

Is it true for K_5?

How could we prove it is not true?



Every triangle has two inner edges and one outer edge. (could prove that, but won't)

$\forall P(K_5)$

$\neg(\forall P(K_5))$

$\exists \neg P(K_5) \longrightarrow$ show that there exists an edge colouring of K_5 with no green and no red triangles

$\therefore$ not true for a group of 5 people.

More general Claim: there is a group of k friends or k strangers in any group of n or more people, $k \geq 4$

Known for some numbers, but many are still open.
No general proof.

Claim: For any group of $n = 1024$ people
- no group of 20 friends AND
- no group of 20 strangers.

with probability 99.9999999999999158%

For any group of 20 people there are at least 2 friends AND ≥ 2 strangers with high probability

Subsets

Set $S = \{a, b, c, d, e\}$

Take subsets $S_1, S_2, S_3, \ldots, S_m$

How big can m be?

$\forall i \neq j: S_i \not\subseteq S_j$ and $S_j \not\subseteq S_i$.

for $n=5$
$m \geq 10$

(At least one element in S_i not in S_j or
" " " " " S_j not in S_i)

$S_1 = \{a, b\}$ $\{c, d, e\}$
$S_2 = \{a, c\}$ $\{b, d, e\}$
$S_3 = \{a, d\}$ $\{b, c, e\}$
$S_4 = \{a, e\}$
$S_5 = \{b, c\}$
$S_6 = \{b, d\}$
$S_7 = \{b, e\}$
$S_8 = \{c, d\}$
$S_9 = \{c, e\}$
$S_{10} = \{d, e\}$

All subsets of size 2.

If $|S_1| = |S_2| = \ldots |S_m|$ then one is a subset of the other only if they are equal.

If we took all subsets of size 3, we would get 10 subsets also. How do we know?

In general for $n \geq 1$, $|S| = n$
then $m \leq \binom{n}{\lfloor \frac{n}{2} \rfloor}$

$\binom{n}{k}$ "n choose k"

Binomial coefficient – $\binom{n}{k}$ = number of subsets of size k in a set of size n.

$n=5$ $\quad \binom{5}{2} = 10$

$k=2$ $\quad \binom{5}{3} = 10$

we will learn how to compute this in general.

To prove question above, we use random numbers.
Which is strange, because there are no random numbers.

QuickSort (S, n)

S = sequence of n numbers.

Quick Sort Algorithm: (Recursive)

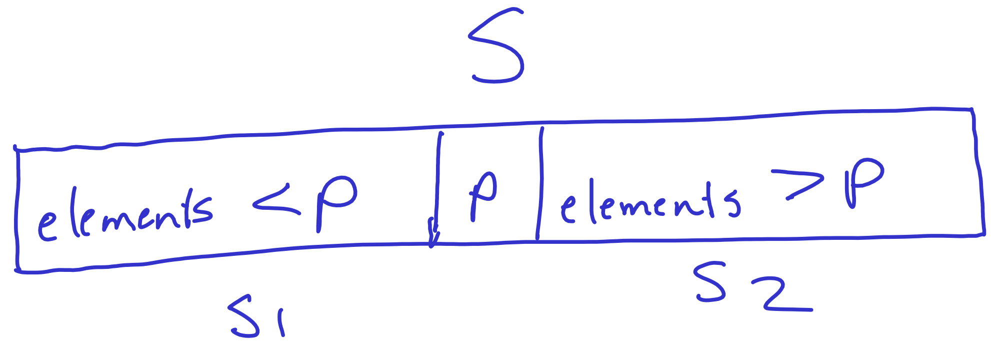
base case: if $n=0$ or $n=1$, list is sorted

if $n \geq 2$:

choose an element p of S as pivot

rearrange S such that:

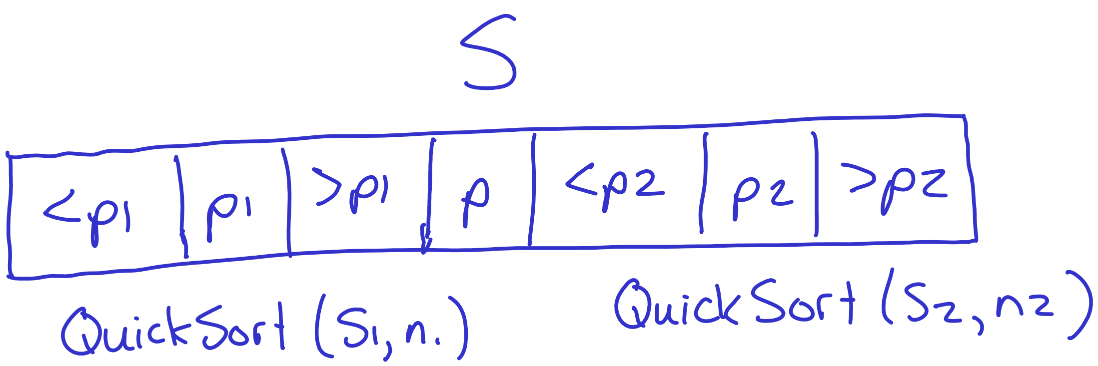
Two things to note:



① Once location of p is found, it does not change

② S_1 and S_2 stay in their positions (positions may change inside S_1 or S_2).

$\Longrightarrow$ We can recursively sort S_1 and S_2 independently in their positions.

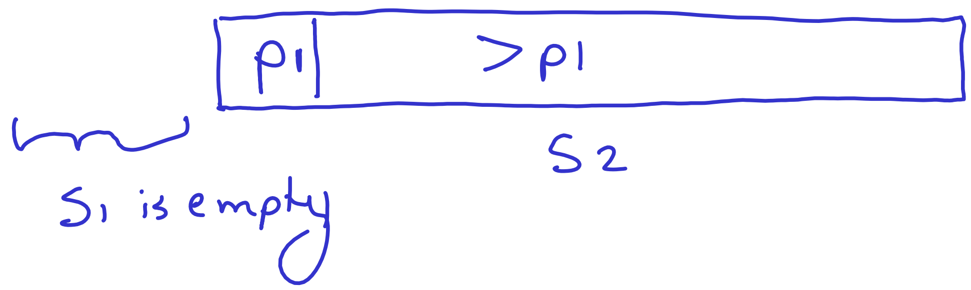


How quick is QuickSort? Depends on location of p.

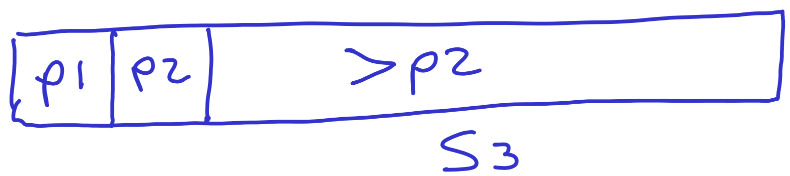
What is a good choice for p?

What is a bad choice for p?

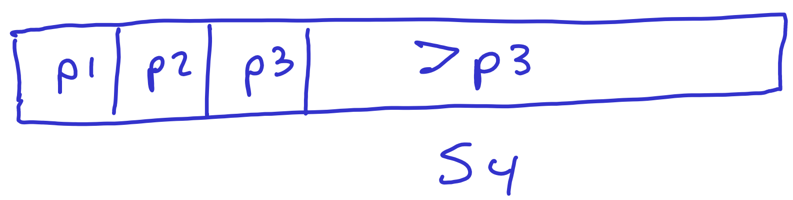
QuickSort (S, n)



What if we choose smallest element as p?

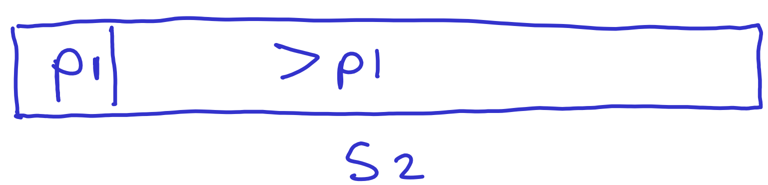


Then next smallest

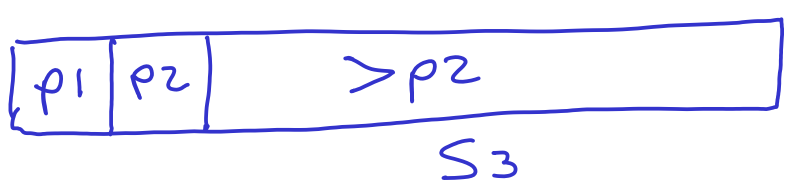


And next smallest

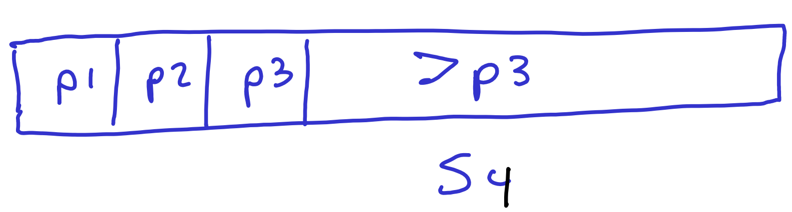
How much work is done?



compare n-1 elements to p1



compare n-2 elements to p2



compare n-3 elements to p3.

Pass 1: rearrange
n elements

Pass 2: $n-1$

Pass 3: $n-2$

$\vdots$

Pass n: 1

Total steps =

$$n+(n-1)+(n-2)+\ldots+1$$

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

$$\begin{array}{r} n+(n-1)+(n-2)+\ldots+2+1 \\ 1+2+3+\ldots(n-1)\ n \\ \hline \underbrace{(n+1)+(n+1)+(n+1)+\ldots+(n+1)+(n+1)}_{n \text{ terms}} \end{array}$$

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

Big-O and Big-Ω

$f(x)=O(g(x))$ if $\exists\ c,k$ such that $\forall x \geq k$, $f(x) \leq c \cdot g(x)$

$f(x)=\Omega(g(x))$ if $\exists\ c,k$ " " $\forall x \geq k$, $f(x) \geq c \cdot g(x)$.

$$\frac{n(n+1)}{2} = \frac{n^2}{2} + \frac{n}{2} \leq n^2 \quad \forall n \geq 1$$

$\therefore$ QuickSort is $O(n^2)$

$$\frac{n^2}{2} + \frac{n}{2} \geq \frac{n^2}{2} \quad \forall n \geq 1$$

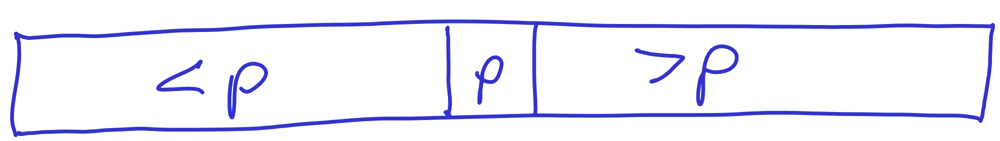
$\therefore$ QuickSort is $\Omega(n^2)$

$\therefore$ QuickSort is $\Theta(n^2)$

MergeSort is $\Theta(n \log n)$

Why called "Quick" Sort?

What happens if we choose a good pivot?



$|S_1| = \frac{n-1}{2}$ $\qquad$ $|S_2| = \frac{n-1}{2}$

Still same work each pass
- less passes

} Problem size is cut in half each pass, then $\log_2 n$ passes.

If we choose a random pivot, expected running time is $O(n \log n)$

Expected = (weighted) average

1	2	3	4	5	6	7	8	9	10

On average how far from the middle is a random pivot?