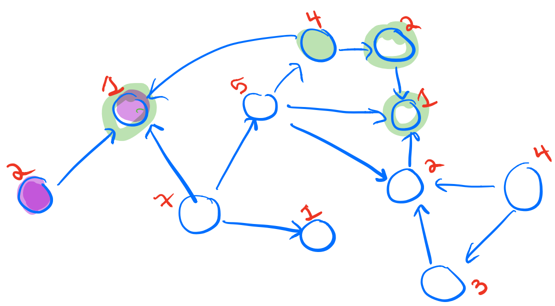
A cool application: reachability

n # vertices
m # edges.

Input: directed graph $G=(V,E)$
Output: $\forall v \in V$, $|S(v)|$ where $S(v)$: set of nodes reachable from v

Today
- Announcements
  - Stirling's Approx
  - hw 2 - google form
- Reachability
- intro to LP
- randomized rounding of LPs

$\Rightarrow n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$



BFS/DFS $O(mn)$

Randomized alg $O(m \log n)$ approximately correct

$\forall v$ let $R(v) \sim \exp(1)$

$$\boxed{s(v) = \min_{w \in S(v)} R(w)}$$

little s

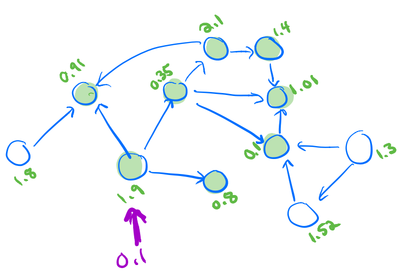
set of vertices reachable from v

$X \sim \exp(\lambda)$

$f(x) = \lambda e^{-\lambda x}$ $x \geq 0$

$\Rightarrow \boxed{\Pr(X \geq x) = e^{-\lambda x}}$

$\Rightarrow \boxed{E(X) = \frac{1}{\lambda}}$



distn of $s(v)$ $S(v)=\{w_1, w_2, \ldots, w_k\}$

$s(v) = \min(\underbrace{R(w_1), R(w_2) \ldots R(w_k)}_{\text{indep } \exp(1)})$

$\Pr(s(v) \geq x) = \Pr(R(w_1) \geq x) \cdots \Pr(R(w_k) \geq x)$

$= e^{-x} \cdot e^{-x} \cdots e^{-x}$

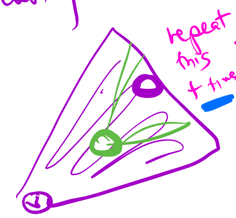
$= e^{-kx} \Leftarrow \exp(k)$

$\Rightarrow \boxed{E(s(v)) = \frac{1}{k}}$

Suggests.
1. $\forall v$ sample $R(v) \sim \exp(1)$
2. compute $\forall v$ $s(v)$
   output as estimate for $|S(v)|$ $\frac{1}{s(v)}$

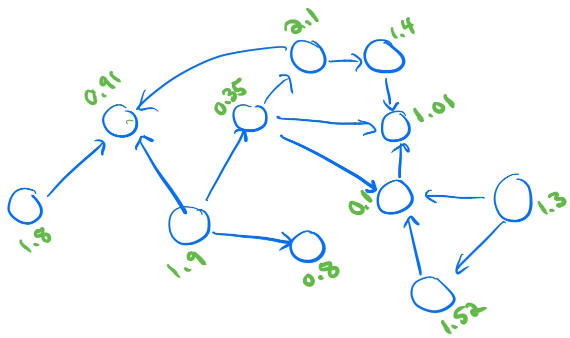
repeat this t times

$O(m+n)$



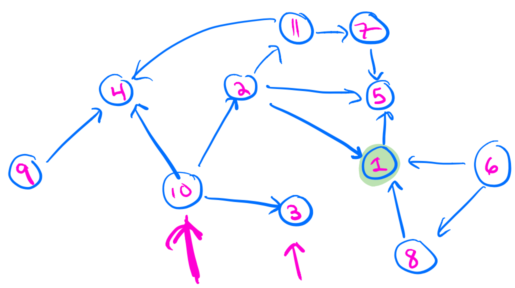
Q1: how to implement step ① efficiently

Q2:

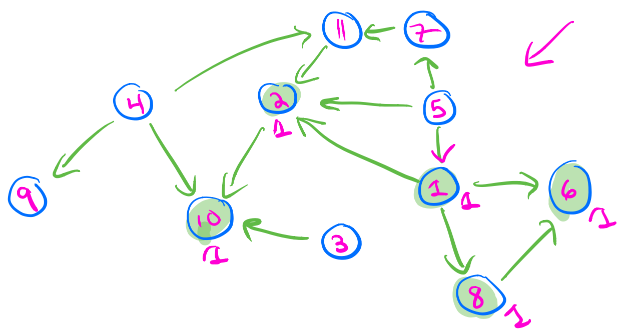


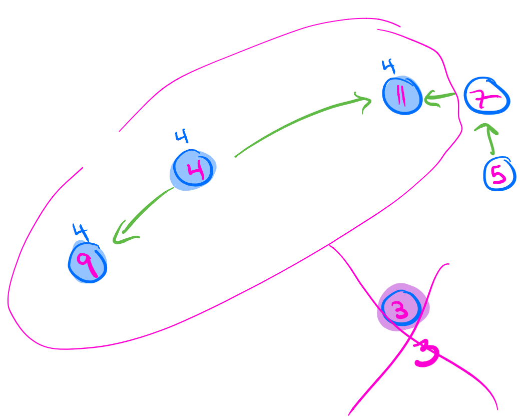
$R(v)$	rank(v)
0.1	1
0.35	2
0.8	3
0.91	4
⋮	

instead of asking $\forall v$ what is $s(v)$.
(or rank of $s(v)$)



reverse directions of edges





$$s_1(v) = \min_{w \in S(v)} \big(R(w)\big)$$

Reverse edge directions.
While $V \neq \emptyset$
let $i :=$ min rank of nodes in V
Perform search to find V_i, set of nodes reachable from node of rank i
$\forall v \in V_i$, let $s(v) := R(v_i)$
$V := V \setminus V_i$
Remove from E all edges incident to nodes in V_i

Repeat this whole thing t times
$\forall v \longrightarrow \underbrace{s_1(v), \ldots, s_t(v)}$

$\boxed{s_1(v) \geq s_2(v) \geq}$ $\geq s_t(v)$

$E(s_i(v)) = \frac{1}{|S(v)|}$

want to show if $\boxed{|S(v)| = k}$

then $\hat{s}(v) \approx k$

showing $\exists$ consts c_1 & c_2 s.t. > 1

$$\boxed{\frac{k}{c_1} \le \frac{1}{\hat{s}(v)}} \le c_2 k$$

$$\Downarrow$$

$$\hat{s}(v) \le \frac{c_1}{k}$$

$\boxed{\left\lfloor \frac{t}{e} \right\rfloor^{th} \text{ highest of these} \quad \text{call that } \hat{s}(v) \quad \text{output as my estimate } \frac{1}{\hat{s}(v)}}$

$s_1(v) \ge s_2(v) \ge \dots \ge s_{\frac{t}{e}}(v)$ $\qquad \ge s_t(v)$

trouble $> \frac{c_1}{k}$ $\qquad \boxed{\le \frac{c_1}{k}}$

if $s_i(v) \le \frac{c_1}{k} \implies$ say $s_i(v)$ good

o.w. its bad.

\# bad ones $\qquad$ $\#s_i$'s $> \frac{c_1}{k}$ $\qquad$ w/ high prob $< \frac{t}{e}$

$N_i = \mathbb{1}_{s_i(v) > \frac{c_1}{k}}$ bad.

$$E(N_i) = Pr(\underbrace{s_i(v)}_{\exp(k)} > \frac{c_1}{k}) = e^{-k\frac{c_1}{k}} = e^{-c_1}$$

choose c_1 so that $Pr\left(\sum_{i=1}^{t} N_i \ge \frac{t}{e}\right)$ very small.

$$\boxed{E\left[\sum_{i=1}^{t} N_i\right] = t \cdot e^{-c_1}}$$

want to choose c_1 so that $\frac{t}{e} \ge 2E\left(\sum_{i=1}^{t} N_i\right)$

$e^{-c_1} = \frac{1}{2e}$

$c_1 = 1 + \ln 2 \approx 1.69$

choose c_1 so that:

$$\boxed{E\left(\sum_{i=1}^{t} N_i\right) = \frac{t}{2e}}$$

$$\boxed{\begin{array}{ll} Pr\left(X \le (1-\delta)\mu\right) \le e^{-\frac{\delta^2 \mu}{2}} & 0 \le \delta \le 1 \\ Pr\left(X \ge (1+\delta)\mu\right) \le e^{-\mu\left[(1+\delta)\ln(1+\delta) - \delta\right]} \le \begin{cases} e^{-\frac{\delta^2 \mu}{2+\delta}} & \delta > 0 \\ e^{-\frac{\delta^2 \mu}{3}} & 0 < \delta \le 1 \end{cases} & \end{array}}$$

$\Uparrow$ $\delta = 1$

$$Pr(X > 2\mu) \le e^{-\frac{\mu}{3}}$$

$\Downarrow$ bad event $\quad s_{\frac{t}{e}}(v) > \frac{c_1}{k}$

occurs when $\sum_{i=1}^{t} N_i > 2E\left(\sum_{i=1}^{t} N_i\right)$

$$Pr\left(s_{\frac{t}{e}}(v) > \frac{c_1}{k}\right) = Pr\left(\sum_{i=1}^{t} N_i > 2E\left(\sum_{i=1}^{t} N_i\right)\right) \le e^{-\frac{E(\sum N_i)}{3}}$$

$t\_$

$$= e^{-\frac{}{6e}}$$

$\forall v \quad \hat{s}(v) > \frac{c_1}{k}$

$\equiv \quad \frac{k}{c_1} > \frac{1}{\hat{s}(v)}$ with prob $\leq e^{-\frac{t}{6e}}$

$$\Pr(\underbrace{\exists v \text{ s.t. } \hat{s}(v) > \frac{c_1}{k}}_{\text{bad event.}}) \leq \sum_v \Pr\left(\hat{s}(v) > \frac{c_1}{k}\right)$$

$$= n \cdot e^{-t/6e}$$

want

$$= o(1)$$

what should t be?

$c \ln n$

$$n \cdot n^{-c/6e}$$

$c > 6e$

$$e^{-t/6e} = e^{-\frac{c \ln n}{6e}} = n^{-\frac{c}{6e}}$$

$o\left(\frac{1}{n}\right)$

$\frac{c}{6e} > 1+\varepsilon$

w.h.p. $\forall v \quad \frac{k}{\cancel{\ldots}} \leq \frac{1}{\hat{s}(v)}$

$$(1-\varepsilon)k \leq \frac{1}{\hat{s}(v)} \leq (1+\varepsilon)k$$

Similarly, we want to show $\exists$ constant c_2 s.t.

$$\frac{k}{1.69} < \boxed{\frac{1}{\hat{s}(v)} \le c_2 k} \quad \text{w.h.p.}$$

$$\Uparrow$$

consider single $s(v)$

$$\Pr\left(s(v) \ge \frac{1}{c_2 k}\right) = e^{-k \cdot \frac{1}{c_2 k}} = e^{-\frac{1}{c_2}}$$

$$N_i' = \mathbb{1}_{s_i(v) \ge \frac{1}{c_2 k}} \qquad E\left(\sum_{i=1}^{t} N_i'\right) = t e^{-\frac{1}{c_2}}$$

choose c_2 so $\Pr\left(\sum_{i=1}^{t} N_i' < \frac{t}{e}\right)$ small

$\underbrace{s_1(v) \ge \dots \, s_{\frac{t}{e}}(v) \ge \dots}_{\text{want all } \ge \frac{1}{c_2 k}}$

$\Uparrow$ want this to be $(1-\delta)\, t e^{-\frac{1}{c_2}}$

For example, if we take $c_2 = 2 \Rightarrow$

$$e^{-\frac{1}{c_2}} = e^{-1/2} \approx 0.606$$

whereas $\frac{1}{e} \approx 0.3678$

$$\Rightarrow \quad \frac{t}{e} \simeq 0.6 \; t e^{-1/2}$$

$$\Rightarrow \delta \approx 0.4$$

$$\Rightarrow \Pr\left(\sum_{i=1}^{t} N_i' < \frac{t}{e}\right) \le e^{-\frac{0.4^2}{2} \cdot 0.6 t}$$

$$\boxed{\begin{array}{l} 0 \le \delta \le 1 \\ \Pr\left(X \le (1-\delta)\mu\right) \le e^{-\frac{\delta^2 \mu}{2}} \end{array}}$$

So again if $t = \Omega(\log n)$

this is $o\left(\frac{1}{n}\right)$

$\Rightarrow$ by a union bound with prob $1 - o(1)$

$$\forall v \quad \frac{1}{\hat{s}(v)} \le 2k$$

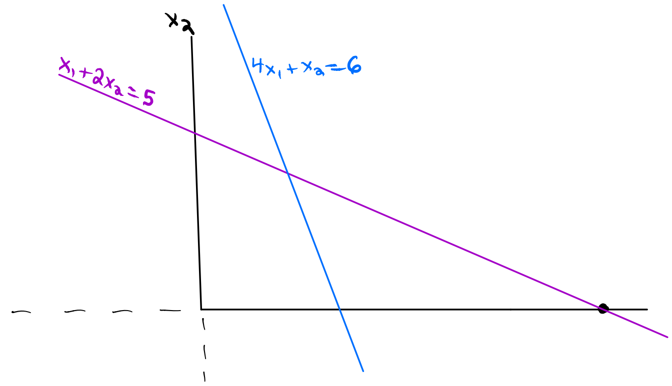
Optimization

min/max $f(x_1, \dots, x_n)$
subject to constraints

When can we solve efficiently?

Linear programming L

$f(x_1, \dots, x_n)$ linear fn.
constraints linear



Example: Diet Problem

Athlete wants to max protein consumption
subject to ≤ 5 units of fat/day
$\leq \$6$ / day

	protein/lb	fat/lb	\$/lb
steak	2	1	4
peanut butter (PB)	1	2	1

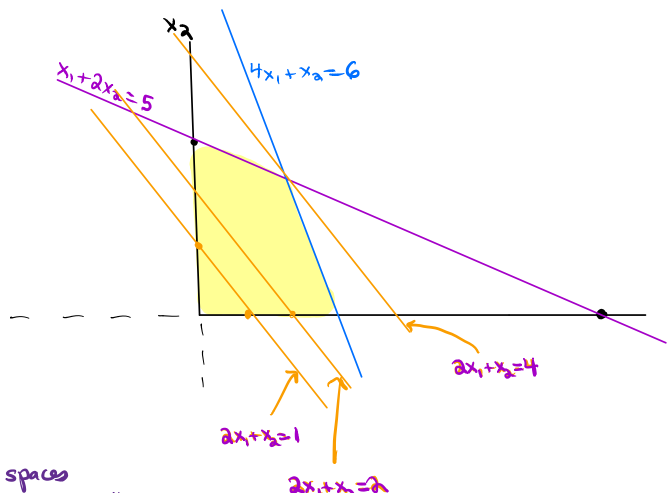
x_1 #lbs of steak/day
x_2 #lbs of PB/day

$$\max \quad \boxed{2x_1 + x_2} \quad \leftarrow \text{objective fn}$$

subject to

$$\boxed{\begin{array}{c} 4x_1 + x_2 \leq 6 \\ x_1 + 2x_2 \leq 5 \\ x_1, x_2 \geq 0 \end{array}} \quad \leftarrow \text{feasible set / feasible region}$$

feasible pt with max objective fn value is "optimal solution"



feasible set is "polyhedron": intersection of half spaces
if also bounded & nonempty $\Rightarrow$ "polytope"

feasible set convex

linear cost fns define family of parallel hyperplanes
optimal feasible pt must occur at corner, a.k.a. vertex
[can't be expressed as convex comb of feasible pts]

Unfortunately too many vertices to enumerate $\approx m^{n/2}$
m constraints, n variables

Ex: $0 \leq x_i \leq 1 \quad 1 \leq i \leq n$
n dimensional hypercube

Input to LP problem

$$\boxed{\begin{array}{l}\vec{c} \in \mathbb{R}^n \\ A \in \mathbb{R}^{m \times n} \\ \vec{b} \in \mathbb{R}^m\end{array}}$$

find $x \in \mathbb{R}^n$ to
max $c \cdot x$
subject to $Ax \le b$
$x_i \ge 0 \ \forall i$

max $\boxed{2x_1 + x_2}$
subject to
$$\boxed{\begin{array}{l} 4x_1 + x_2 \le 6 \\ x_1 + 2x_2 \le 5 \\ x_1, x_2 \ge 0 \end{array}}$$

$c = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$

$A = \begin{pmatrix} 4 & 1 \\ 1 & 2 \end{pmatrix}$

$b = \begin{pmatrix} 6 \\ 5 \end{pmatrix}$

$$\begin{aligned} \equiv \max \ & c_1x_1 + c_2x_2 + \dots + c_nx_n \\ \text{s.t.} \ & a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \le b_1 \\ & a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \le b_2 \\ & \vdots \\ & a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n \le b_m \\ & x_1, x_2, \dots, x_n \ge 0 \end{aligned}$$

$a \cdot x \le b$
$\Rightarrow -a \cdot x \ge -b$

$ax = b$ ...
$ax \ge b$
$ax \le b$

There are efficient algs for LP.

- simplex } exp worst case, "smoothed" poly time
- ellipsoid
- interior pt methods } polynomial time

$\boxed{\text{smoothed analysis of algs:}}$
measures perf of alg under slight random perturbation of input

$\boxed{\text{ellipsoid}}$: sometimes even exponential-sized LPs can be solved in poly time

poly time w/ "separation oracle" - poly time alg
given x → yes x feasible
→ no separating hyperplane

$\boxed{\text{Bottom line:}}$ If you can formulate your problem as LP, it can be solved in time $\mathrm{poly}(m, n, \log(\max_{i,j}(a_{ij}, b_j, c_i)))$

$\boxed{\text{Using "linear-programming relaxations" to get approx algs for NP-hard problems \& then apply randomized rounding}}$

c- approx alg.
poly time alg that always returns soln within factor of c of optimal.

Example: MAX 2-SAT

Given a Boolean formula in CNF (and of ors) where each clause has 2 literals, find T/F assignment to variables that maximizes # satisfied clauses

max-version NP-hard

0.94.

Ex: $\underset{T}{(x_1 \vee x_3)} \wedge \underset{T}{(x_1 \vee \bar{x}_3)} \wedge (\underset{F}{\bar{x}_1} \vee \underset{T}{x_4}) \wedge \underset{T}{(\bar{x}_4 \vee x_1)} \wedge \underset{T}{(x_1 \vee x_4)} \wedge (\bar{x}_1 \vee \bar{x}_4)$

2-SAT: $\boxed{\text{is the 2-SAT formula satisfiable in P.}}$

$\boxed{\text{max-2SAT}}$
⇓
NP-hard.

$\boxed{\text{Step 1:}}$ Formulate problem as <u>integer</u> linear program

2SAT formula with n vars, m clauses

for ILP

$x_1, \dots, x_n$ $\forall$ var in formula

$z_1, \dots, z_m$ $\forall$ clause.

$$x_i = \begin{cases} 1 & \text{if var } i \to T \\ 0 & \text{o.w.} \end{cases}$$

$$z_j = \begin{cases} 1 & \text{if clause } j \text{ is satis} \\ 0 & \text{o.w.} \end{cases}$$

Integer linear prog
NP-hard.

Example.

$\Rightarrow (x_1 \vee x_2) \wedge (x_1 \vee \bar{x}_2) \wedge (\bar{x}_1 \vee x_2) \wedge (\bar{x}_1 \vee \bar{x}_2) \wedge (x_1 \vee \bar{x}_3)$

max $z_1 + z_2 + z_3 + z_4 + z_5$

s.t.

$$x_1 + x_2 \geq z_1$$
$$x_1 + (1 - x_2) \geq z_2$$
$$(1 - x_1) + x_2 \geq z_3$$
$$(1 - x_1) + (1 - x_2) \geq z_4$$
$$x_1 + (1 - x_3) \geq z_5$$

$$\begin{cases} x_1, x_2, x_3 \in \{0,1\} \\ z_1, z_2, z_3, z_4, z_5 \in \{0,1\} \end{cases}$$