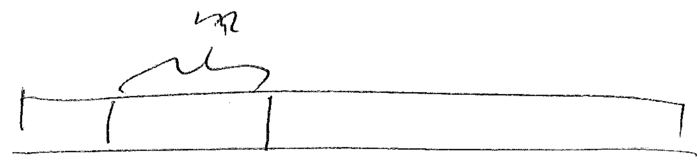
April 21

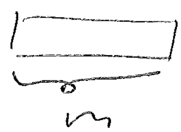
Office Hour Today (April 21) 2–2:40pm

String Matching Problem

Given a string $S[1..n]$ and a pattern $p[1..m]$ for Σ

Determine if p appears in S? The number of occurrences of p in S

Rabin Karp Algorithm S 

$hp = hash(p[1..m])$ p 

For $j = 1$ to $n-m+1$,

  $hs = hash(S[j..j+m-1])$

  if $hs = hp$

    return yes.

$hp = hash(p[1..m])$ p is from Σ

$p \in \underbrace{\Sigma \times \Sigma \times \cdots \times \Sigma}_{m}$.

We will view p as an m-digit, base $|\Sigma|$ number.

we can convert p to a decimal number

Example $\Sigma = \{\overset{0}{a}, \overset{1}{b}, \overset{2}{c}, \ldots, \overset{25}{z}\}$.

$p =$ "algorithm"

$|\Sigma| = 26$

we will view p as a base 26, 12-digit number

What's the decimal representation of p?

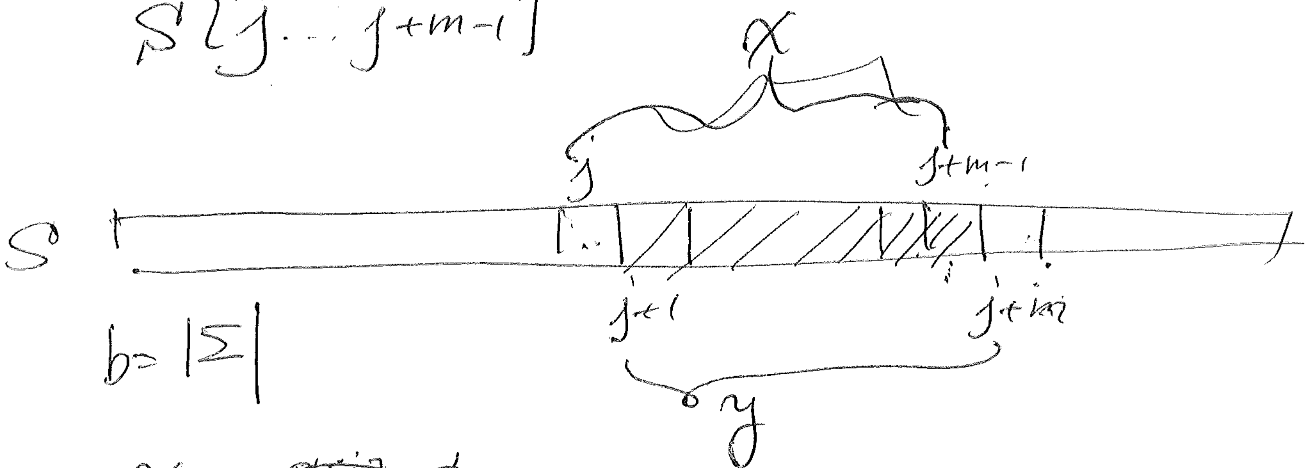
$a \times 26^8 + l \times 26^7 + g \times 26^6 + \cdots + h \times 26^1 + m \times 26^0$

$= 0 \times 26^8 + \cdots$

$+$

Question: can we speed up the hashing of

$S[j \dots j+m-1]$



$b = |\Sigma|$

$$x = S[j]\cdot b^{m-1} + \underbrace{S[j+1]\cdot b^{m-2} + \cdots + S[j+m-2]\cdot b + S[j+m-1]}_{z}$$

$$y = \underbrace{S[j+1]\cdot b^{m-1} + S[j+2]\cdot b^{m-2} + \cdots + S[j+m-1]\cdot b}_{b\cdot z} + S[j+m]$$

$$x = S[j]\cdot b^{m-1} + z$$

$$y = b\cdot z + S[j+m]$$

$$\begin{cases} h(x) = x \bmod p = (S[j]\cdot b^{m-1} + z)\,\%\,p \\ h(y) = y \bmod p = (bz + S[j+m])\,\%\,p \end{cases}$$

$$\begin{cases} h(x) = \left(S[j]\%p \cdot b^{m-1}\%p + z\%p \right)\%p \\ h(y) = \left(b\%p \cdot z\%p + S[j+m]\%p \right)\%p \end{cases}$$

$$h(y) = \left(b\%p \left(h(x) - S[j]\%p \cdot b^{m-1}\%p \right) + S[j+m]\%p \right)\%p$$

Given $h(x)$, how long it takes to calculate $h(y)$?

$$\left.\begin{array}{ll} h(x): & \text{given} \\ b\%p: & \text{precalculated.} \\ S[j]\%p & O(1)\text{ time} \\ b^{m-1}\%p & \text{precalculated.} \\ S[j+m]\%p & O(1)\text{ time} \end{array}\right\} \begin{array}{l} O(1)\text{ time to} \\ \text{calculate } h(y) \\ \text{given } h(x) \end{array}$$

$a^{16}\%p$

$a^{8}\%p$ $\quad$ $a^{8}\%p$

$a^{4}\%p$ $\quad$ $a^{4}\%p$

Final Running Time $O(n+m)$

Recall Universal Hashing.

For every pair of distinct keys the odds of collision is $\leq \frac{1}{n}$, where n is the size of the hash table.

Maximum Flow. Problem.

Given a digraph $G(V,E)$, each edge $e \in E$ is associated with a nonnegative capacity. $u(e)$

$\uparrow$ utility.

A flow f from a source vertex $s \in V$ to a sink vertex $t \in V$ is a function $f: E \to [0, +\infty)$ such that:

(1) for any $e \in E$, $\quad 0 \le f(e) \le u(e)$

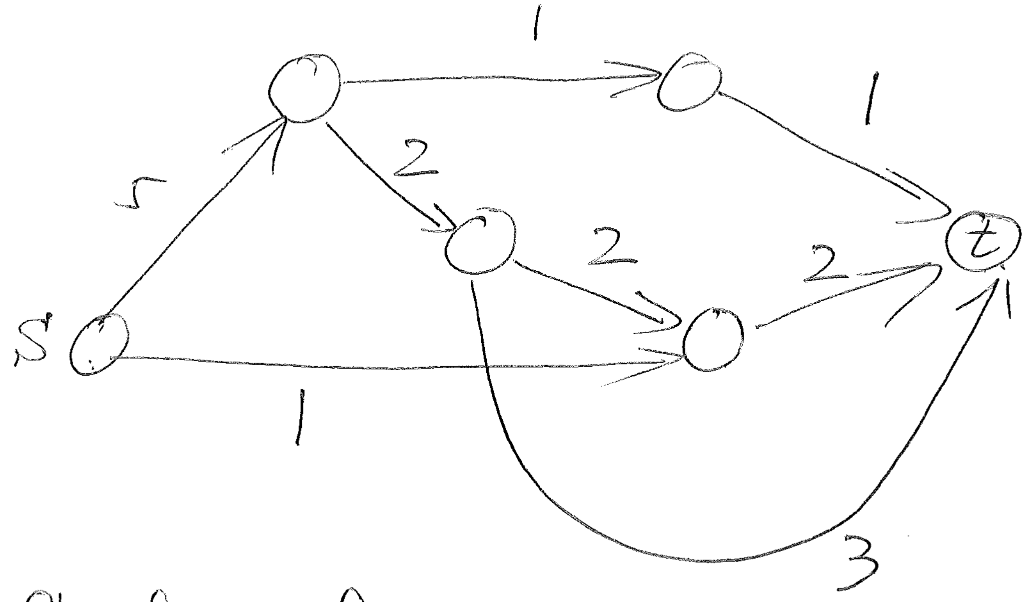
(2) For every $v \in V$ and $v \neq s, t$,

$$\sum_{\forall e(w,v)} f(e) = \sum_{\forall e(v,w)} f(e)$$

(3) The value of f is

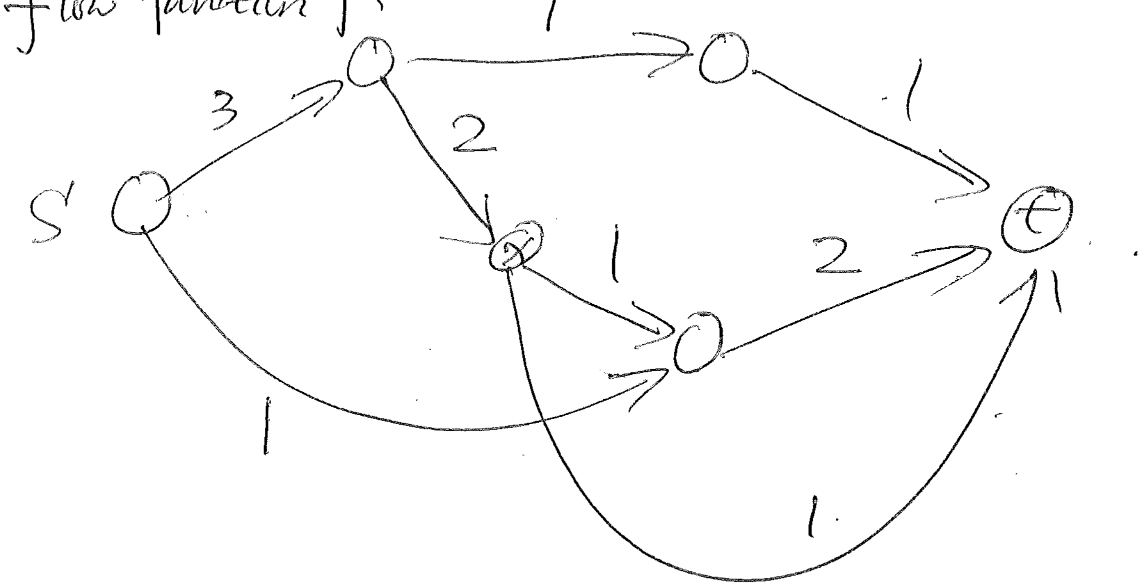
$$|f| = \sum_{\forall e(s,w)} f(e) - \sum_{\forall e(w,s)} f(e)$$

The maximum flow problem: $\max |f|$

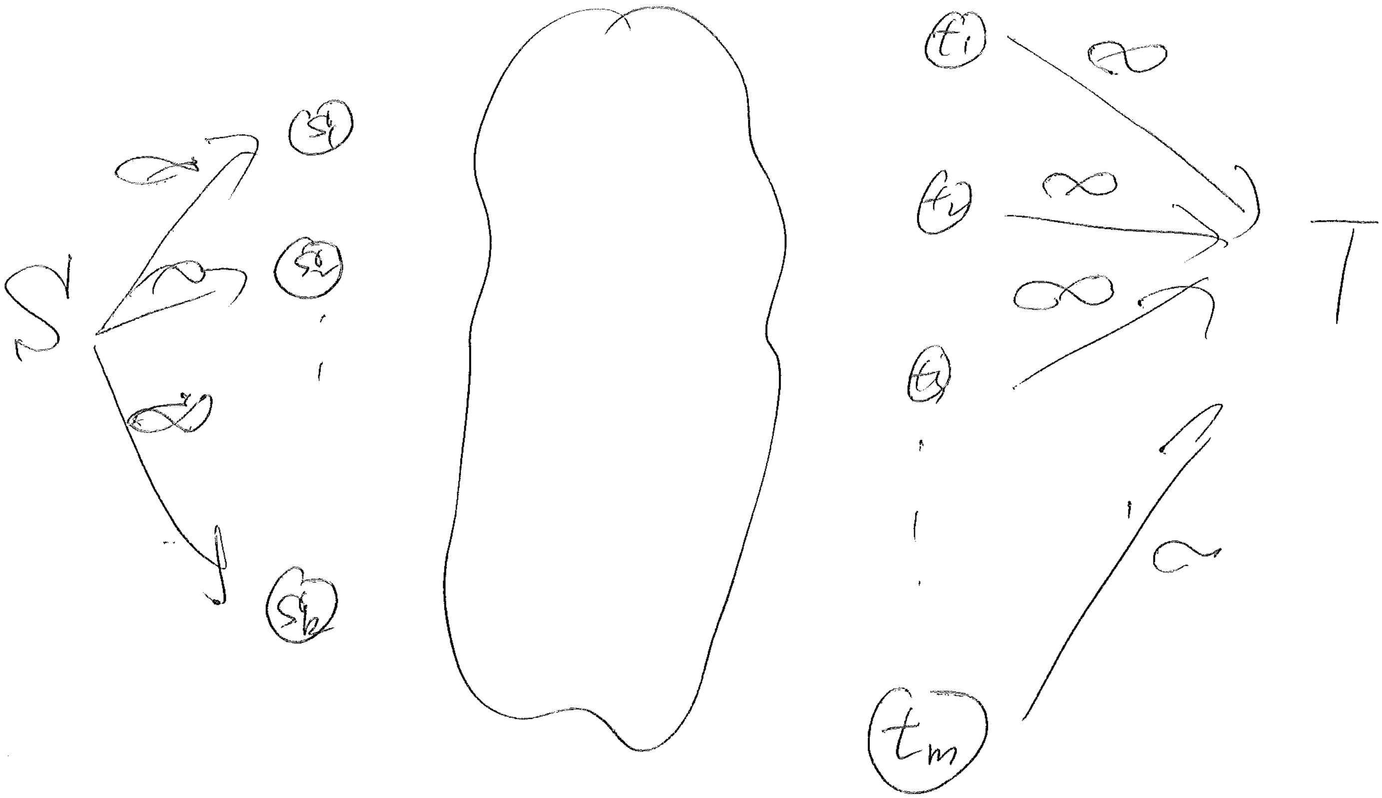


a sample flow network

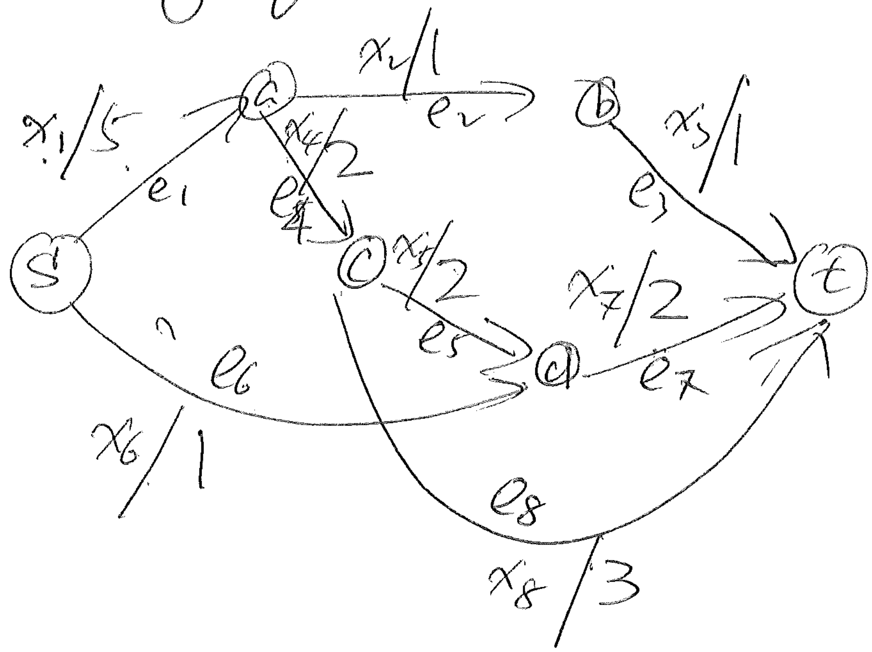
flow funcrin f:



$|f| = 4$


S
s1
s2
sk
∞
∞
∞
t1
t2
t3
tm
∞
∞
∞
∞
T

LP modeling of Maximum Flow.



max $\quad x_1 + x_6$

s.t.

e_1: $\quad 0 \le x_1 \le 5$

e_2: $\quad 0 \le x_2 \le 1$

e_3: $\quad 0 \le x_3 \le 1$

e_4 $\quad 0 \le x_4 \le 2$

e_5: $\quad 0 \le x_5 \le 2$

e_6: $\quad 0 \le x_6 \le 1$

e_7: $\quad 0 \le x_7 \le 2$

e_8: $\quad 0 \le x_8 \le 3$

For a. $x_1 = x_2 + x_4$

b $x_2 = x_3$

c: $x_4 = x_5 + x_8$

d: $x_5 + x_6 = x_7$

Minimum Cost Flow (MCF)

Recall: the maximum flow asks what is the maximum amount of flow one can send from a source to a sink through a network

The MCF asks what's the <u>optimal way</u> to send a given amount of flow through a network?

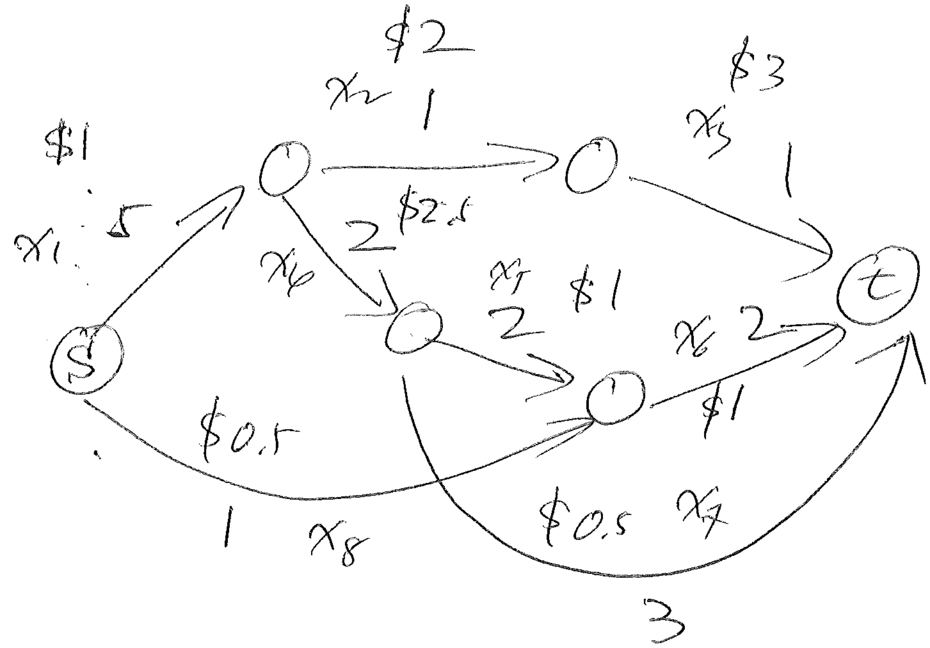
Given a digraph $G(V,E)$, each $e \in E$, $u(e)$, $c(e)$

$s, t \in V$, s: source, t: sink

per unit capacity.

$|f| = b \leftarrow$ the amount of flow

The goal is to minimize the total cost of f.



$b = 3$

min $\$1 \cdot x_1 + \$2 \cdot x_2 + \$3 \cdot x_3 + \$2.5\, x_4$

$+ \$1 \cdot x_5 + \$1 \cdot x_6 + \$0.5 \cdot x_8 + \$0.5\, x_7$

s.t. $x_1 + x_8 = 3$

capacity constraint: $0 \le x_1 \le 5$

$\vdots$

flow balance constraint: $x_1 = x_4 + x_6$

$\vdots$

MCF

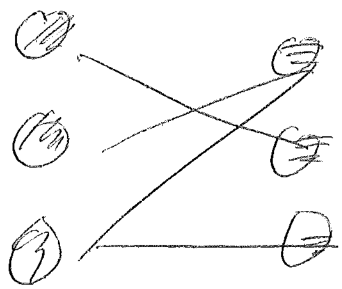
$$\min \sum_{j=1}^{n} c_j x_j$$

s.t. $0 \le x_j \le u_j$ for $j=1 \ldots n$

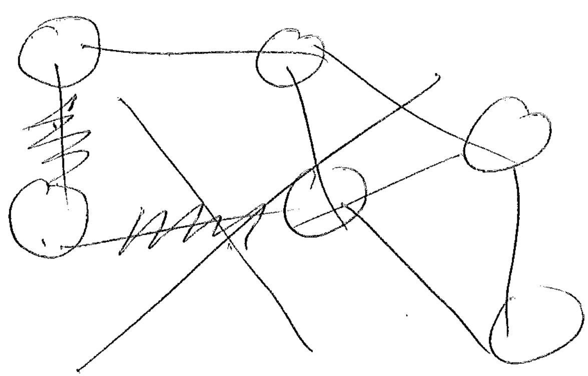
$$\sum_{\forall w,\, e(w,v)} x_e = \sum_{\forall u\, e(v,u)} x_e$$ for $\cancel{j}$ $v=1 \ldots m$

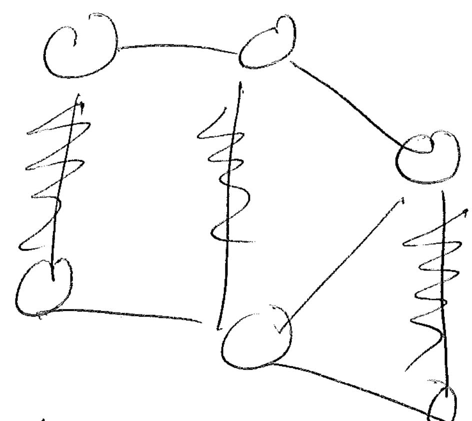
$$\sum_{\forall w\; e(s,w)} x_e = b = \sum_{\forall u\, e(u,t)} x_e$$

Recall an undirected graph $G(V,E)$ is bipartite if V can be partitioned into U and W $\left(\begin{array}{l} U \cap W = \phi \\ U \cup W = V \end{array}\right)$ such that all edges in E connects a vertex in U to a vertex in W.



A matching M in a undirected graph $G(V,E)$ is a subset of edges such that no two edges share a common vertex.



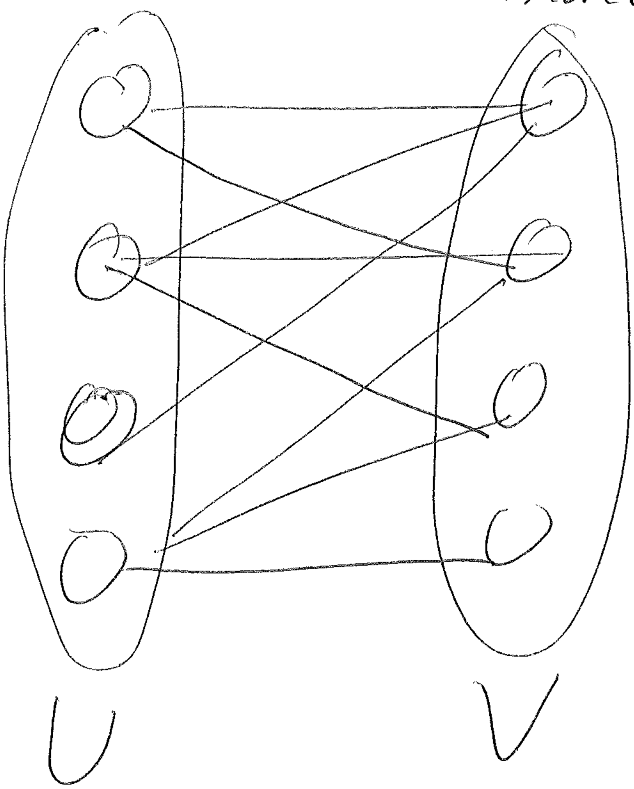


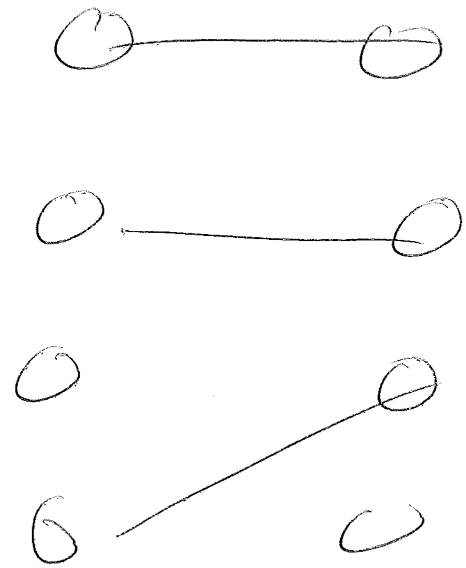
interesting problem: maximum matching

Biparite Maximum Matching

Given a bipartite graph $G(U, V, E)$

what's the maximum matching M?





maximum matching