<u>Recap</u> : Utility Function $f: 2^{\{1,\dots,n\}} \to \mathbb{R}$

1. Monotonicity : $f(S) \leq f(S')$ when $S \subseteq S'$

2. Submodularity : $\underbrace{\Delta_e f(S)}_{f(S \cup \{e\}) - f(S)} \geq \Delta_e f(S')$ when $S \subseteq S'$

<u>Submodular function optimization (with size constraint)</u>

<u>Input</u> ; Monotone Submodular Function f, Budget k

<u>Output</u> : $S \subseteq \{1,\dots,n\}$

<u>Constraint</u> : $\|S\| \leq k$

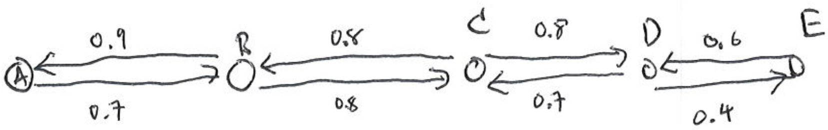
<u>Objective Function</u> : Maximize $f(S)$.

<u>Theorem</u> The greedy algorithm discussed ~~to~~ on last week return SOL

best solution OPT

$$f(SOL) \geq 0.63 \cdot f(OPT)$$

<u>Influence Maximization</u> [Kempe, Kleinberg, and Tardos, ToC, 2015]



vertices = persons in a social network

links = (a,b) if a regularly talk to b

$p((a,b))$: probability that a will talk with b about a specific information.

If we tell an information to C,

- C will know the information with prob. 1
- B " " 0.8
- A " " $0.8 \times 0.9 = 0.72$
- D " " 0.8
- E " " $0.8 \times 0.6 = 0.48$

Expected # persons that know the information

= Influence of C

$= 1 + 0.8 + 0.72 + 0.8 + 0.48$

$= 3.64$

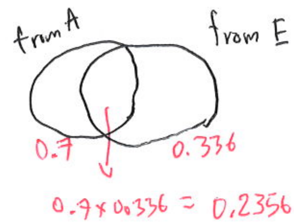
If we tell an information to B,

- B will know the information with prob. 1
- A will 〃 ———————— 〃 0.9
- C 〃 ———————— 〃 0.8.
- D 〃 ———————— 〃 $0.8 \times 0.8 = 0.64$
- E 〃 ———————— 〃 $0.8 \times 0.8 \times 0.4 = 0.256$

Expected # persons
= Influence of B
$= 1 + 0.9 + 0.8 + 0.64 + 0.256$
$= 3.60$

∴ C is more influential than B.

If we tell an information to A and E



- A will know the information with prob. 1
- E 〃 ———————— 〃 1
- B will know the information from A with prob. 0.7

  B 〃 ———————— 〃 from E with prob. $0.6 \cdot 0.7 \cdot 0.8 = 0.336$

  B 〃 ———————— 〃 from A or E with prob. $0.7 + 0.336 - 0.2356 = 0.8$
- C will know the information with prob. 0.7448
- D will know the information with prob. 0.78

Influence of A, E $= 1 + 1 \cancel{+0.7+0.336} + 0.8 + 0.7448 + 0.78 = 4.33$

Problem

Input : A social network (V, E)

For each link $e \in E$, $p(e)$ and budget integer k

Output : A set of persons to tell an information to $S \subseteq V$

Constraint : $|S| \leq k$

Objective Function : Maximize $\boxed{\text{expected \# persons that know the information.}}$ $f(S)$

Bonus question When $k=2$, what is the best solution for our network ?

Restate the problem

Input : function $f : 2^V \to \mathbb{R}$, budget k

$f(S) := $ expected # persons

Output : $S \subseteq V$

Constraint : $|S| \leq k$

Objective Function : Maximize $f(S)$

f is a monotone submodular function $\Rightarrow$ Submodular function maximization with size constraint can use greedy algorithm and have

$$f(SOL) \geq 0.63 \cdot f(OPT)$$

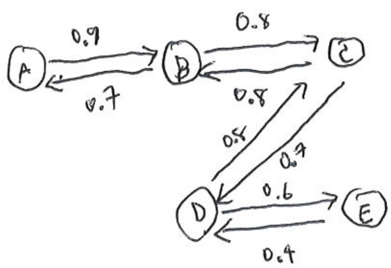
True ~~False~~ because we can reach more persons with a larger set S.

Submodularity

Tell the information to

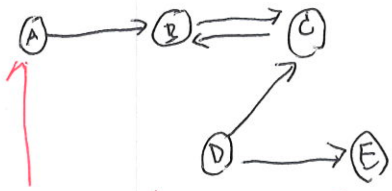
This link is used by B with prob. 0.8

The link exists with prob. 0.8



G

$\Longrightarrow$



Tell information to A and E

- Reach A, B, C, E
- Prob. that we have this network

$$= 0.9 \cdot 0.3 \cdot 0.8 \cdot 0.8 \cdot 0.8 \cdot 0.3 \cdot 0.6 \cdot 0.6$$

We have $A \to B$

We don't have $B \to A$

$$f(S) = \sum_{G' : \text{ possible graphs}} \Pr[G'] \cdot \underbrace{(\#\text{ persons we can reach by } S \text{ in graph } G')}_{f_{G'}(S)}$$

Lemma $f_{G'}$ is a monotone submodular function for all G'

Proof : $P_i :=$ set of nodes that can reach from i in graph G'

$$f_{G'}(S) = \left| \bigcup_{i \in S} P_i \right|$$ ← proved to be a monotone submodular function on last week.

$\blacksquare$

Theorem f is a monotone submodular function.

Proof Assume that $S \subseteq S'$.

$$
\begin{aligned}
\Delta_e f(S) &= f(S \cup \{e\}) - f(S) \\
&= \sum_{G' : \text{possible graphs}} \Pr[G'] \cdot f_{G'}(S \cup \{e\}) - \sum_{G' : \text{possible graphs}} \Pr[G'] \cdot f_{G'}(S) \\
&= \sum_{G' : \text{possible graphs}} \Pr[G'] \left(f_{G'}(S \cup \{e\}) - f_{G'}(S) \right) \\
&= \sum_{G' : \text{possible graphs}} \Pr[G'] \cdot \Delta_e \overset{\text{submodular function}}{f_{G'}}(S) \qquad \geq \Delta_e f_{G'}(S') \\
&\geq \sum_{G' : \text{possible graphs}} \Pr[G'] \cdot \Delta_e f_{G'}(S') \\
&= \sum_{G' : \text{possible graphs}} \Pr[G'] \left(f_{G'}(S' \cup \{e\}) - f_{G'}(S') \right) \\
&= \sum_{G' : \text{possible graphs}} \Pr[G'] f_{G'}(S' \cup \{e\}) - \sum_{G' : \text{possible graphs}} \Pr[G'] f_{G'}(S') \\
&= f(S' \cup \{e\}) - f(S') = \Delta_e f(S')
\end{aligned}
$$

$\therefore \Delta_e f(S) \geq \Delta_e f(S')$ $\blacksquare$

Linear Programming (LP)

can be solved by CPLEX (IBM)

Input: Matrix A, vectors b, c

Output: vector x

Constraint: $Ax \leq b$ $\quad x \geq 0$ → all elements in x are non-negative

Objective Function: Maximize $c^t \cdot x$