CMU 15-896

Social networks 2: Influence Maximization

Teacher:
Ariel Procaccia

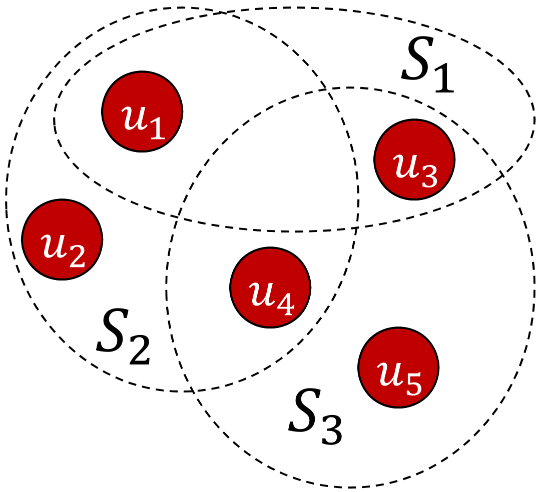
Motivation

- Firm is marketing a new product
- Collect data on the social network
- Choose set S of early adopters and market to them directly
- Customers in S generate a cascade of adoptions
- Question: How to choose S?

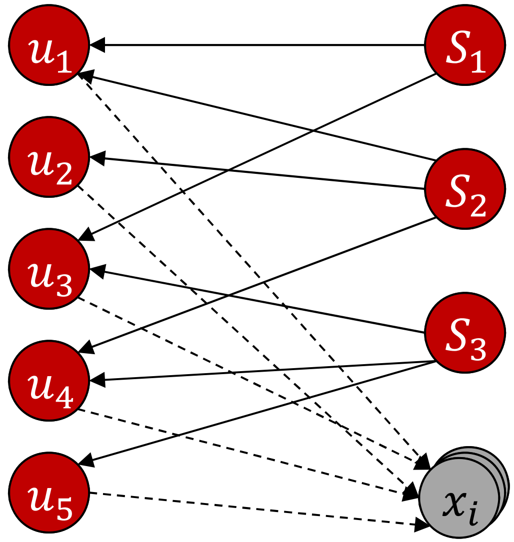
Influence functions

- Assume: finite graph, progressive process
- Fixing a cascade model, define influence function
- $f(S)$ = expected #active nodes at the end of the process starting with S
- Maximize $f(S)$ over sets S of size k
- Theorem [Kempe et al. 2003]: Under the general cascade model, influence maximization is NP-hard to approximate to a factor of $n^{1-\epsilon}$ for any $\epsilon > 0$

Proof of theorem



- SET COVER: subsets $S_1, \ldots, S_m$ of $U = \{u_1, \ldots, u_t\}$; cover of size k?
- Bipartite graph: $u_1, \ldots, u_t$ on one side, $S_1, \ldots, S_m$ and $x_1, \ldots, x_T$ for $\mathrm{T} = t^c$ on the other
- u_i becomes active if $S_j \ni u_i$ is active
- x_j becomes active if $u_1, \ldots, u_t$ are active
- Min set cover of size $k \Rightarrow T + t + k$ active
- Min set cover of size $> k \Rightarrow < t + k$ active ■



Submodularity for approximation

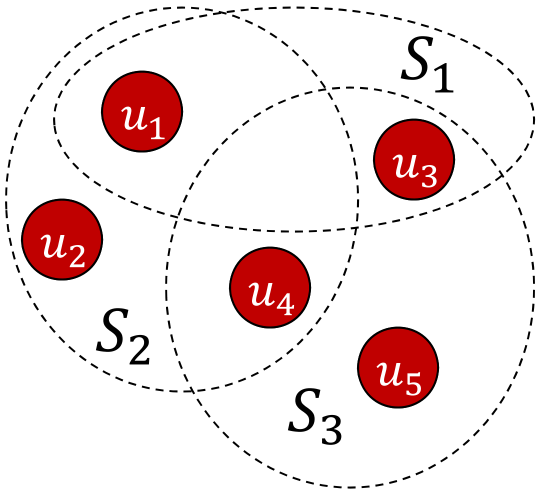
- Try to identify broad subclasses where good approx is possible
- f is submodular if for $X \subseteq Y, v \notin Y$,
$$f(X \cup \{v\}) - f(X) \geq f(Y \cup \{v\}) - f(Y)$$
- f is monotone if for $X \subseteq Y, f(X) \leq f(Y)$
- Reduction gives f that is not submodular
- Theorem [Nemhauser et al. 1978]: f monotone and submodular, S^* optimal k-element subset, S obtained by greedily adding k elements that maximize marginal increase; then
$$f(S) \geq \left(1 - \frac{1}{e}\right) f(S^*)$$

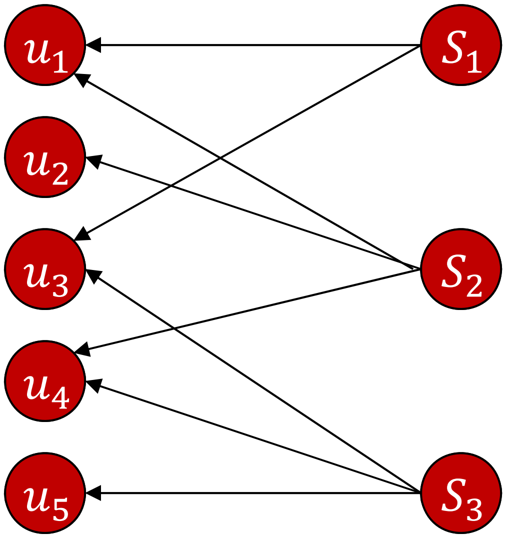
INDEPENDENT CASCADE MODEL

- Reminder of model:
  - For each $(u, v) \in E$ there is a weight p_{uv}
  - When a node u becomes activated it has one chance to activate each neighbor v with probability p_{uv}
- Theorem [Kempe et al. 2003]: Under the independent cascade model:
  - Influence maximization is NP-hard
  - The influence function f is submodular

Proof of NP-hardness

- Almost the same proof as before
- SET COVER: subsets $S_1, \dots, S_m$ of $U = \{u_1, \dots, u_t\}$; cover of size k?
- Bipartite graph: $u_1, \dots, u_t$ on one side, $S_1, \dots, S_m$ on the other
- If $u_i \in S_j$ then there is an edge (S_j, u_i) with weight 1
- Min SC of size $k \Rightarrow t + k$
- Min SC of size $> k \Rightarrow < t + k$ active ■





Proof of submodularity

- Lemma: If $f_1, \dots, f_r$ are submodular functions, $c_1, \dots, c_r \geq 0$, then $f = \sum_{i=1}^{r} c_i f_i$ is a submodular function
- Proof: Let $X \subseteq Y$ and $v \notin Y$, then

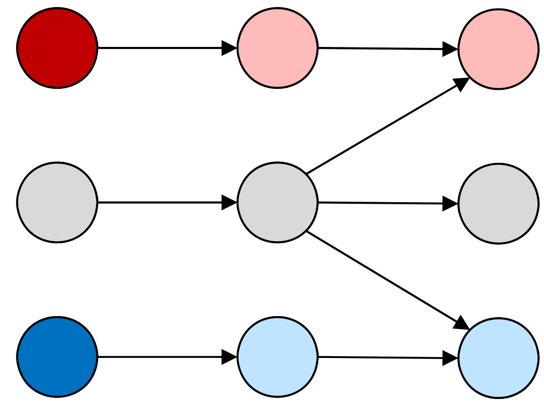
$$f(X \cup \{v\}) - f(X) - (f(Y \cup \{v\}) - f(Y))$$
$$= \sum_{i=1}^{r} c_i [f_i(X \cup \{v\}) - f_i(X) - (f_i(Y \cup \{v\}) - f_i(Y))] \geq 0$$

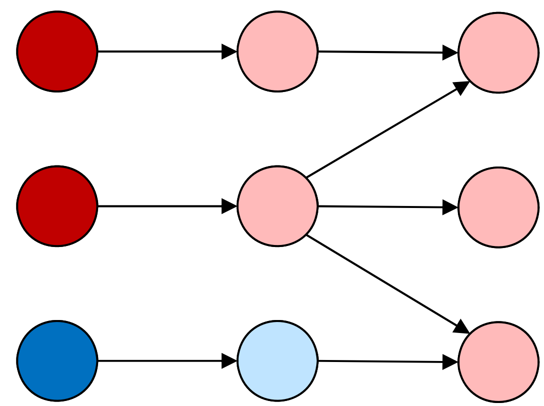
Proof of submodularity

- Key idea: for each (u, v) we flip a coin of bias p_{uv} in advance
- Let α denote a particular one of the $2^{|E|}$ possible coin flip combinations
- $f_\alpha(S) =$ activated nodes with S as seed nodes and α coin flips
- $v \in f_\alpha(S)$ iff v is reachable from S via live edges

Proof of submodularity

- f_α is submodular
- $f(S) = \sum_\alpha \Pr[\alpha] \cdot f_\alpha(S)$, that is, f is a nonnegative weighted sum of submodular functions
- By the lemma, f is submodular ■





Linear threshold model

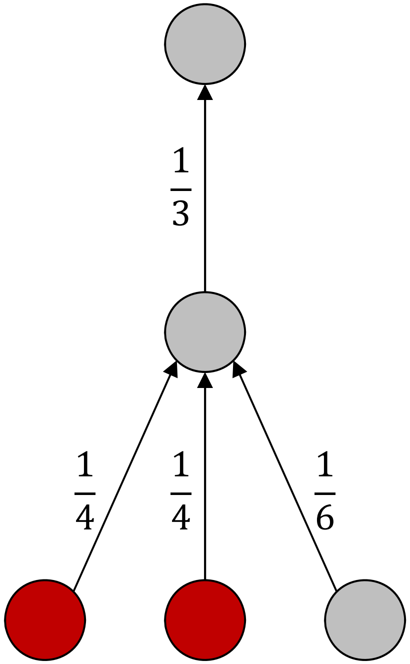
- Reminder of model:
  - Nonnegative weight w_{uv} for each edge $(u,v) \in E$; $w_{uv} = 0$ otherwise
  - Assume $\forall v \in V$, $\sum_u w_{uv} \leq 1$
  - Each $v \in V$ has threshold θ_v chosen uniformly at random in [0,1]
  - v becomes active if

$$\sum_{\text{active}\, u} w_{uv} \geq \theta_v$$

Linear threshold model

Poll 1: What is $f(S)$?

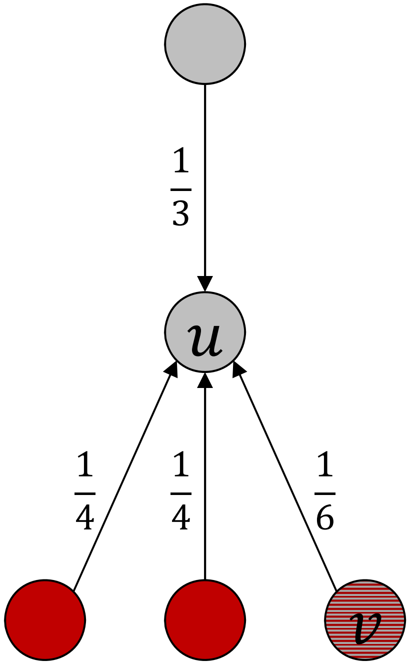
1. $\frac{2}{3}$
2. $\frac{5}{2}$
3. $\frac{8}{3}$
4. $\frac{13}{4}$



Linear threshold model

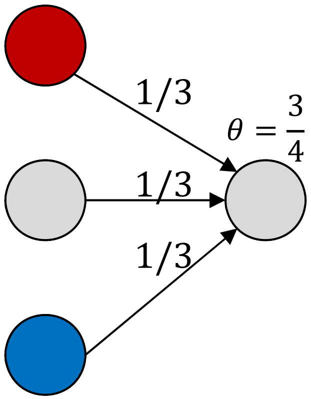
Poll 2: Given that u is inactive, prob. it becomes active after v becomes active

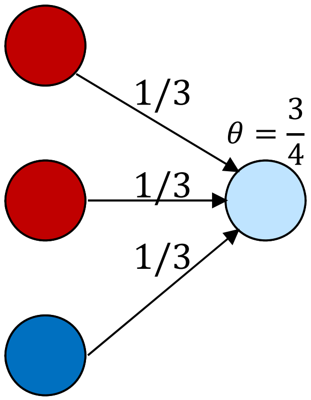
1. 1/6
2. 1/3
3. 1/2
4. 2/3



Linear threshold model

- Theorem [Kempe et al. 2003]: Under the linear threshold model:
  - Influence maximization is NP-hard
  - The influence function f is submodular
- Difficulty: fixing the coin flips α, f_α is not submodular





Proof of submodularity

- Each v chooses at most one of its incoming edges at random; (u, v) selected with prob. w_{uv}, and none with prob. $1 - \sum_u w_{uv}$
- If we can show that these choices of live edges induce the same influence function as the linear threshold model, then the theorem follows from the same arguments as before

Proof of submodularity

- We sketch the equivalence of the two models
- Linear threshold:
  - $A_t =$ active nodes at end of iteration t
  - $\Pr[v \in A_{t+1} \mid v \notin A_t] = \frac{\sum_{u \in A_t \setminus A_{t-1}} w_{uv}}{1 - \sum_{u \in A_{t-1}} w_{uv}}$
- Live edges:
  - At every times step, determine whether v's live edge comes from current active set
  - If not, the source of the live edge remains unknown, subject to being outside the active set
  - Same probability as before ■

Progressive vs. nonprogressive

- Nonprogressive threshold model is identical except that at each round v chooses θ_v^t u.a.r. in [0,1]
- Suppose process runs for T steps
- At each step $t \leq T$, can target v for activation; k interventions overall
- Goal: $\sum_v$ #rounds v was active
- Reduces to progressive case