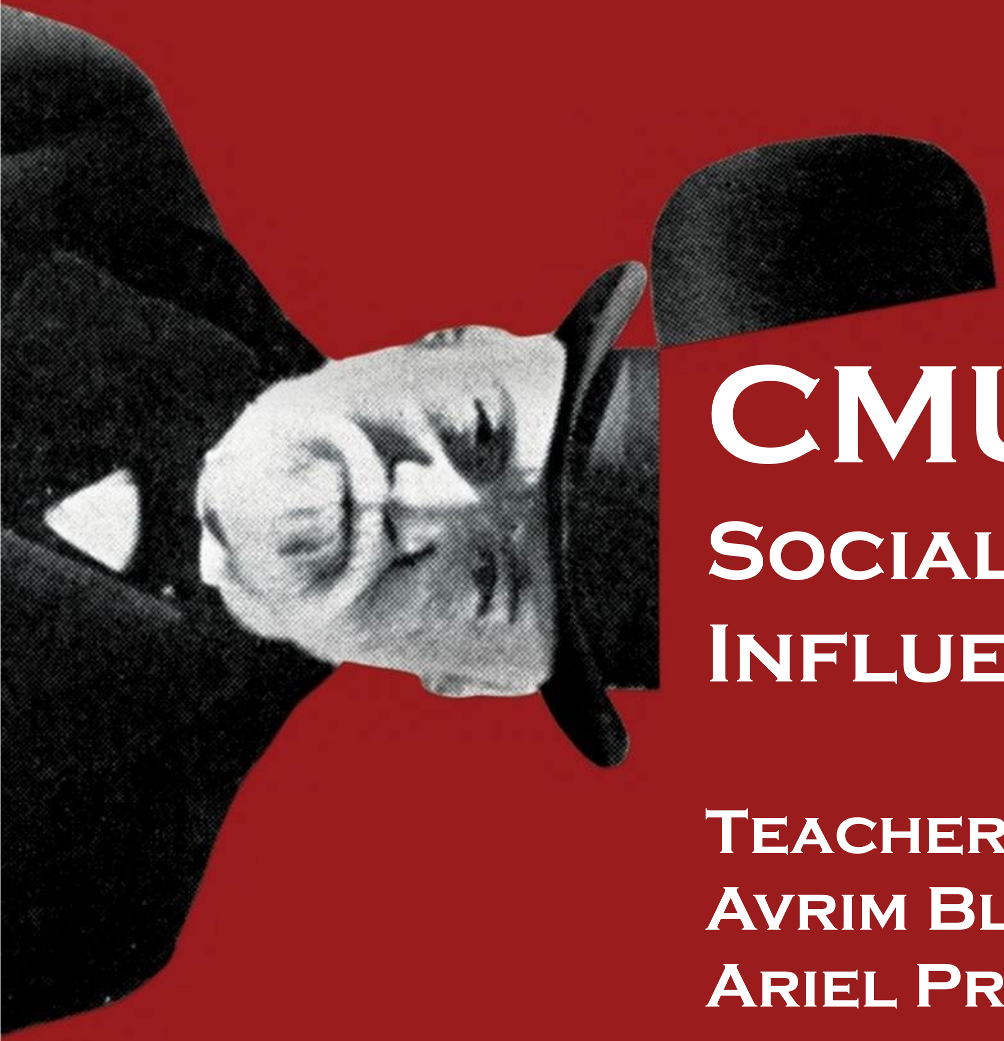


CMU 15-896

Social networks: Influence maximization

Teachers:
Avrim Blum
Ariel Procaccia (this time)

Motivation

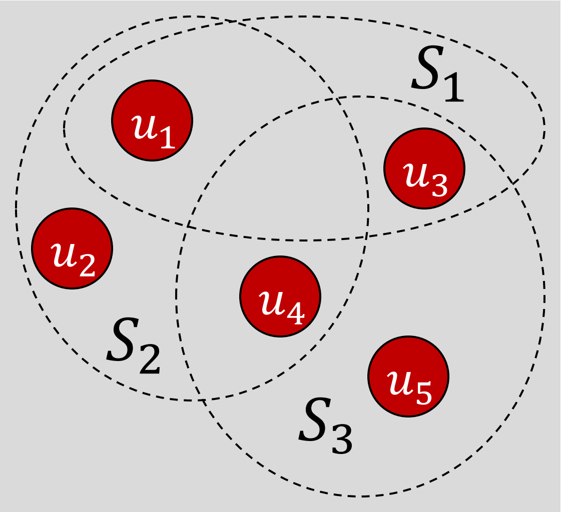
- Firm is marketing a new product
- Collect data on the social network
- Choose set S of early adopters and market to them directly
- Customers in S generate a cascade of adoptions
- Question: How to choose S?

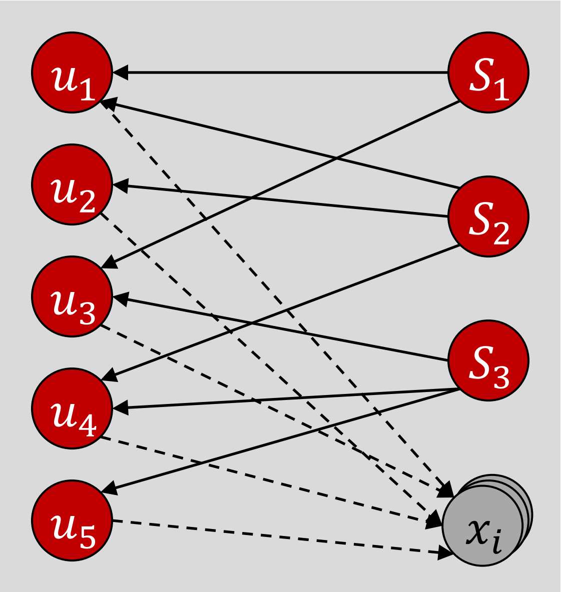
Influence functions

- 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 $S_1, \dots, S_m$ of $U = \{u_1, \dots, u_t\}$; cover of size k?
- Bipartite graph: $u_1, \dots, u_t$ on one side, $S_1, \dots, S_m$ and $x_1, \dots, x_T$ 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 $u_1, \dots, u_t$ are active
- Min set cover of size $k \Rightarrow T + t + k$ covered
- 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

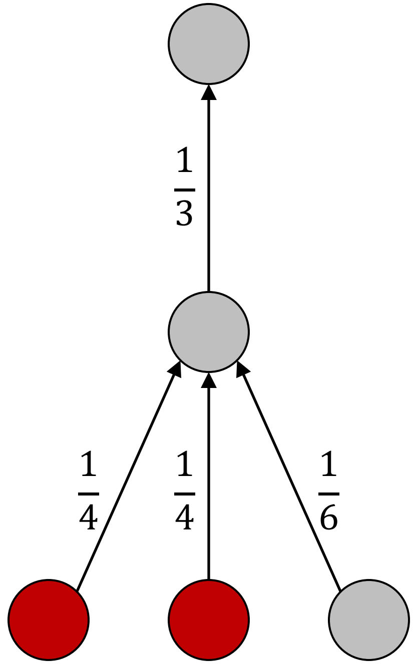
- Reminder of model:
  - For each $(u, v) \in E$ there is a weight p_{uv}
  - When a node u becomes activated it has one chance to activate each neighbor v with probability p_{uv}
- **Theorem [Kempe et al. 2003]:** Under the independent cascade model:
  - Influence maximization is NP-hard
  - The influence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

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 u.a.r. in [0,1]
  - v becomes active if

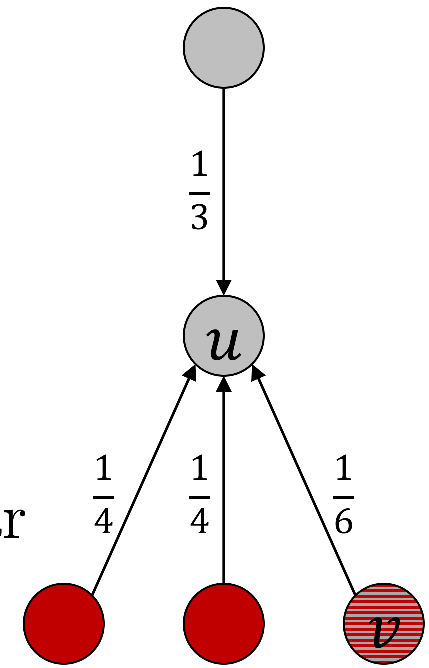
$$\sum_{\text{active } u} w_{uv} \geq \theta_v$$

- Vote: What is $f(S)$?



Linear threshold model

- Vote: Given that u is inactive, prob. that it becomes active when v becomes active
- **Theorem [Kempe et al. 2003]:** Under the linear threshold model:
  - Influence maximization is NP-hard
  - The influence function f is submodular
- We prove the theorem on the board



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