# Section 20.3: Influence Maximization

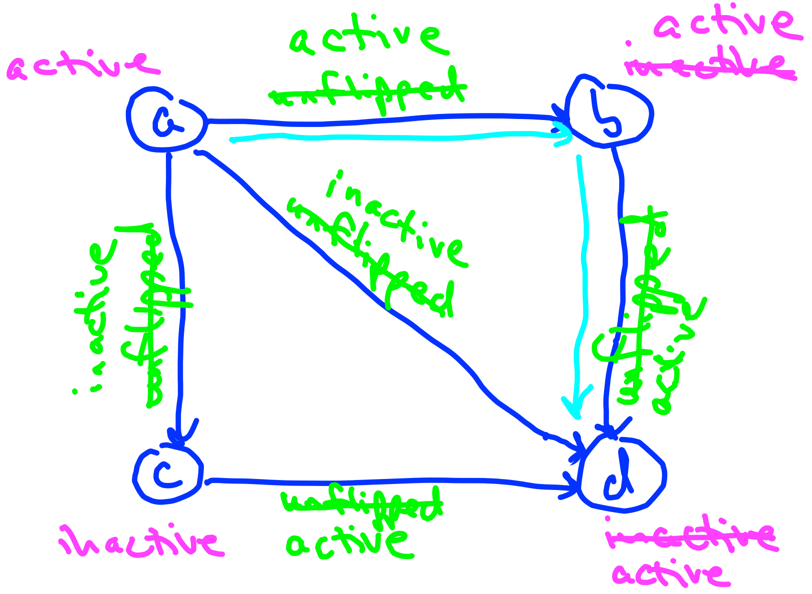
# Cascades in Social Networks

**Social network**: directed graph $G=(V,E)$.
- $V$ = people   $(v,w) \in E \Leftrightarrow v$ "influences" $w$

**A Simple Cascade Model**: [$p$ = "activation probability", $S$ = "seed" vertices]
- seed vertices initially "active"; others "inactive"; all edges "unflipped"
- while there is an active vertex $v$, unflipped edge $(v,w)$:
  - flip biased coin, comes up "heads" w/ probability $p$
  - "heads" => make $(v,w)$ "active", $w$ "active" (if not active already)
  - "tails" => make $(v,w)$ "inactive"

# Example

[a = initial seed vertex]



# Problem Definition

(Kempe-Kleinberg-Tardos)

**Input:** directed graph $G=(V,E)$, probability $p\in[0,1]$, positive integer $k$.

**Notation:** $A(S)$ = vertices eventually activated with seed vertices $S$ (random)

$f_{inf}(S) = E[|A(S)|]$ (expectation over coin flips in cascade model)

**Output:** set $S$ of $k$ vertices that maximizes $f_{inf}(S)$.

# A Greedy Algorithm

KKT

- $S := \emptyset$ [chosen vertices]
- for $j = 1$ to $k$: [greedily increase influence]
  - $v^* = \operatorname{argmax}_{v \in V} [f_{inf}(S \cup \{v\}) - f_{inf}(S)]$
  - $S := S \cup \{v^*\}$
- return $S$

# Quiz #1

What is the running time of a straightforward implementation of the KKT algorithm? [n = # of vertices, m = # of edges]

(a) $O(knm)$

(b) $O(knm^2)$

(c) $O(kn2^m)$

(d) Unclear

$$f_{inf}(S) := E[|A(S)|]$$

eventually activated vertices

# Approximate Correctness

**Guarantee:**

influence of KKT's solution $\geq \left(1-\left(1-\frac{1}{k}\right)^k\right) \cdot$ max-possible influence.

[best-case scenario!]

**Intuition:** influence = weighted average of coverage/event attendance functions. (on subgraph of activated edges)

[need to check that previous proof extends]

# Influence and Coverage

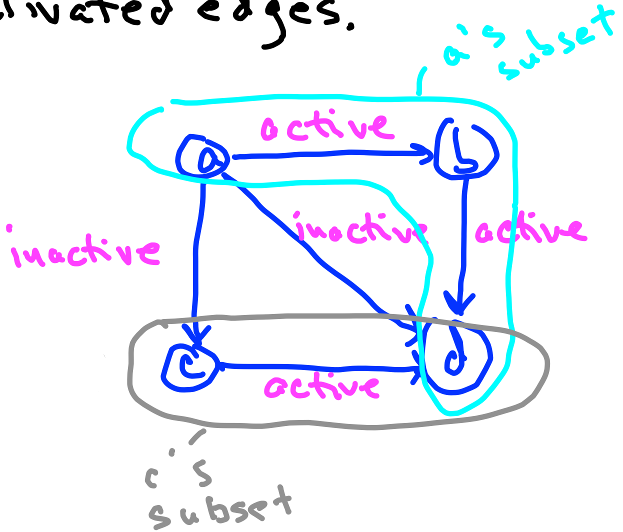
**Fix:** $G=(V,E)$, $p\in[0,1]$, $k$. (+include postprocessing step)

**Recall:** v eventually activated $\Leftrightarrow$ exist directed path of activated edges from seed vertex to v.

**Imagine:** flip all coins at beginning. H = activated edges.

$\Rightarrow$ influence is just a coverage function!

(ground set = vertices, subset for v = vertices reachable from v using edges of H)




**Upshot:** $f_{inf}$ has form $\sum_{H\subseteq E} p_H \cdot f_H(S)$

$p_H = p^{|H|}(1-p)^{|E|-|H|}$

$f_H$: coverage function

# Proof of Approximate Correctness (Part 1)

**Key Lemma:** Let $I^*$ = max-possible influence.
Each vertex chosen by KKT increases the influence by at least $\frac{1}{k}(I^* - \text{current influence})$.

**Proof:** Let $S^*$ = some optimal solution (w/ influence $I^*$). Fix $H \subseteq E$.

- max coverage analysis implies:

$$\sum_{v \in S^*} \Big[ \overbrace{f_H(S_{j-1} \cup \{v\}) - f_H(S_{j-1})}^{\text{coverage increase from } v \text{ (given } H)} \Big] \geq \underbrace{f_H(S^*) - f_H(S_{j-1})}_{\text{coverage deficiency (given } H)}$$

# Proof of Approximate Correctness (Part 2)

$$\sum_{H \subseteq E} p_H \left( \sum_{v \in S^*} \left[ f_H(S_{j-1} \cup \{v\}) - f_H(S_{j-1}) \right] \right) \geq \sum_{H \subseteq E} p_H \left( f_H(S^*) - f_H(S_{j-1}) \right)$$

[ $2^m$ such inequalities, one per $H \subseteq E$ ($m$ = # of edges) ]

Rearranging....

$$\sum_{v \in S^*} \Big[ \underbrace{\sum_{H \subseteq E} p_H f_H(S_{j-1} \cup \{v\})}_{f_{inf}(S_{j-1} \cup \{v\})} - \underbrace{\sum_{H \subseteq E} p_H f_H(S_{j-1})}_{f_{inf}(S_{j-1})} \Big]$$

$$\geq \underbrace{\sum_{H \subseteq E} p_H f_H(S^*)}_{f_{inf}(S^*)} - \underbrace{\sum_{H \subseteq E} p_H f_H(S_{j-1})}_{f_{inf}(S_{j-1})}$$

Recall: $\forall S$, $f_{inf}(S) = \sum_{H \subseteq E} p_H f_H(S)$

$\Rightarrow$ KKT influence increases by $\geq \frac{1}{k}(f_{inf}(S^*) - f_{inf}(S_{j-1}))$