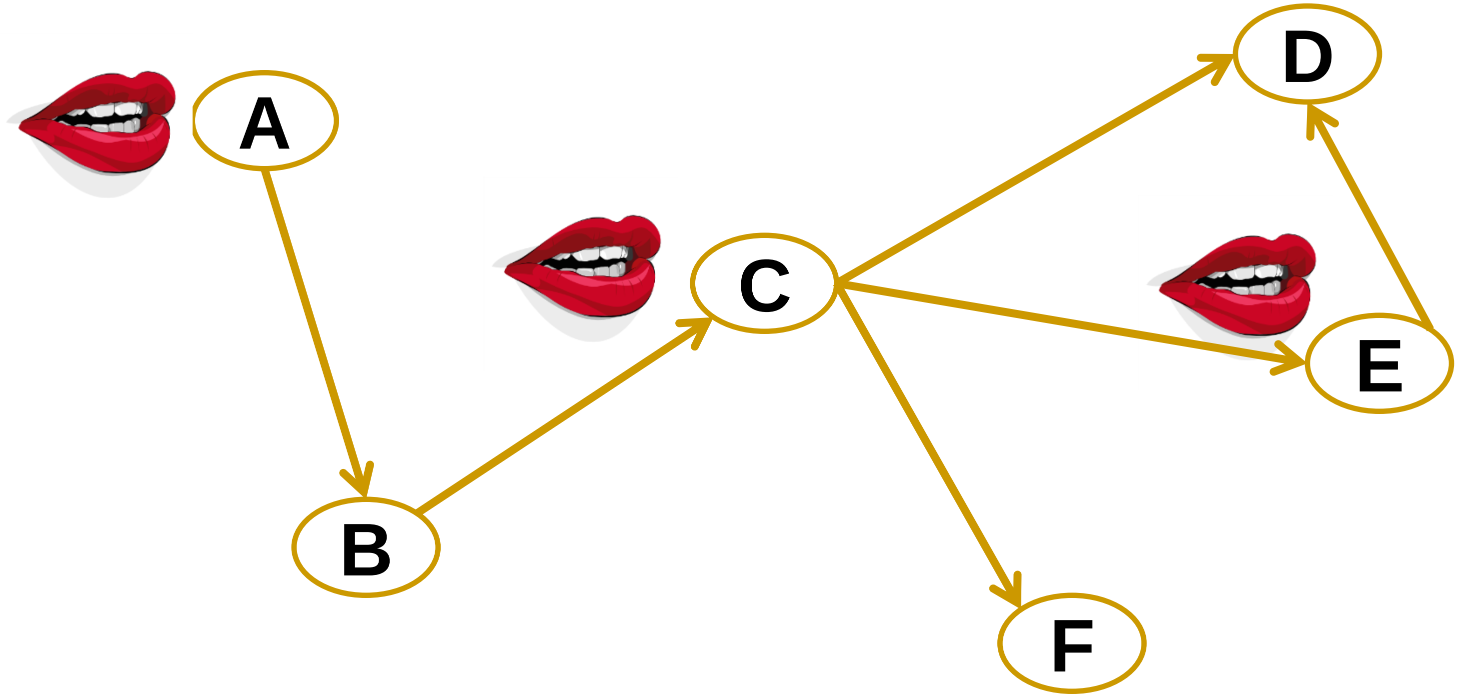
Maximizing the Spread of Influence through a Social Network

Han Wang
Department of Computer Science
ETH Zürich

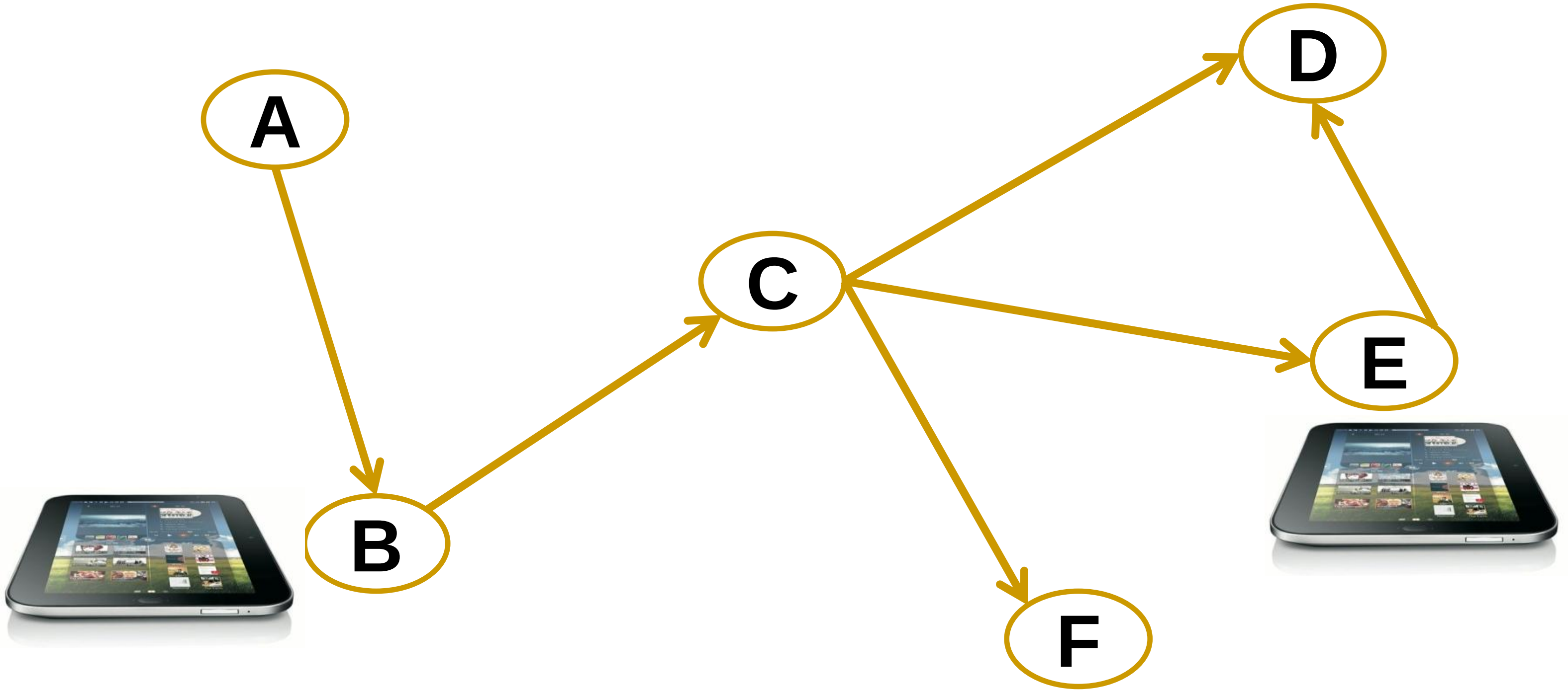
Problem Example 1: Spread of Rumor

- 2012 = end!



Problem Example 2: Viral Marketing

- ezPad 1 beats iPad 3



Problem Definition

- G: a social network (n nodes)
- Model: spread process
- S: initially active subset (k **seeds**)
- $\boldsymbol{\sigma(S)}$**:** #final active nodes (**achievement**)

- Task: Choose S^*
- Goal: ~~$\sigma(S^*) = \max \sigma(S)$~~ NP-Hard

Realistic Goal:

Approximate the maximum with a guarantee

Choose S: $\sigma(S) \geq r \cdot \sigma(S^*)$

Contents in This Talk

- G: a social network (n nodes)
- Model: spread process — Two Models
- S: initially active subset (k **seeds**)
- $\boldsymbol{\sigma(S)}$: #final active nodes (**achievement**)

- Task: Choose S^*
- Goal: ~~$\sigma(S^*) = \max \sigma(S)$~~ — Prove: NP-Hard

Realistic Goal:

Approximate the maximum with a guarantee — Prove: guarantee

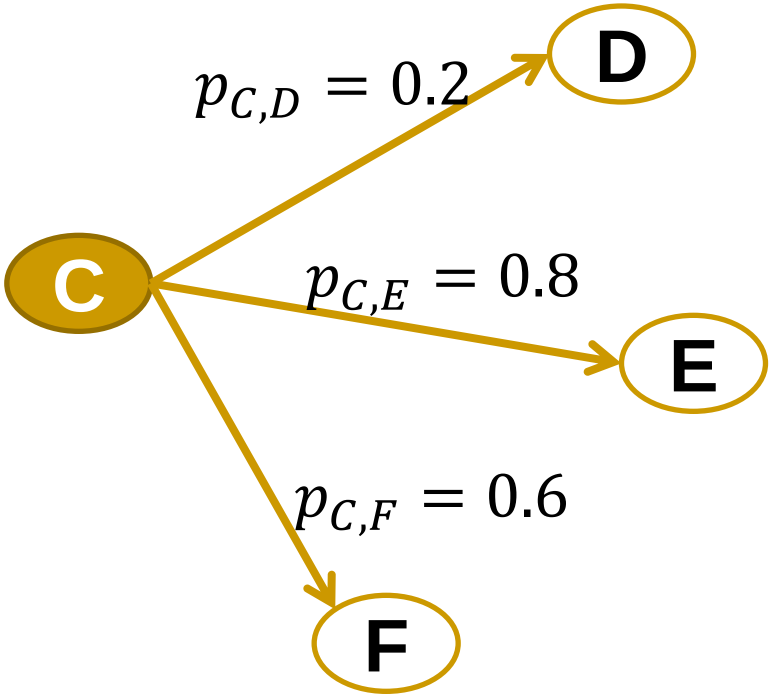
Choose S: $\sigma(S) \geq r \cdot \sigma(S^*)$

Model 1:
Independent Cascade Model

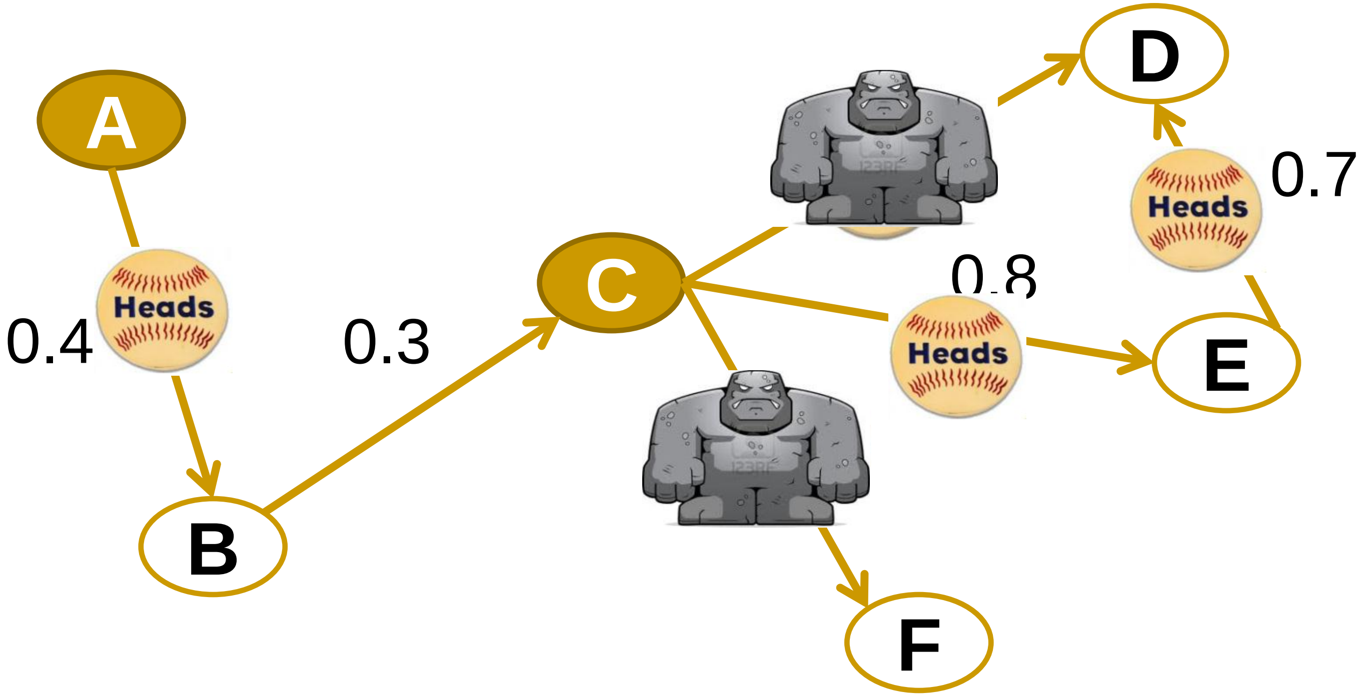
Model 1: Cascade Model

- Each active node try to activate his neighbors

  - $p_{u,v}$ $\quad 1 - p_{u,v}$
  - Only a single chance

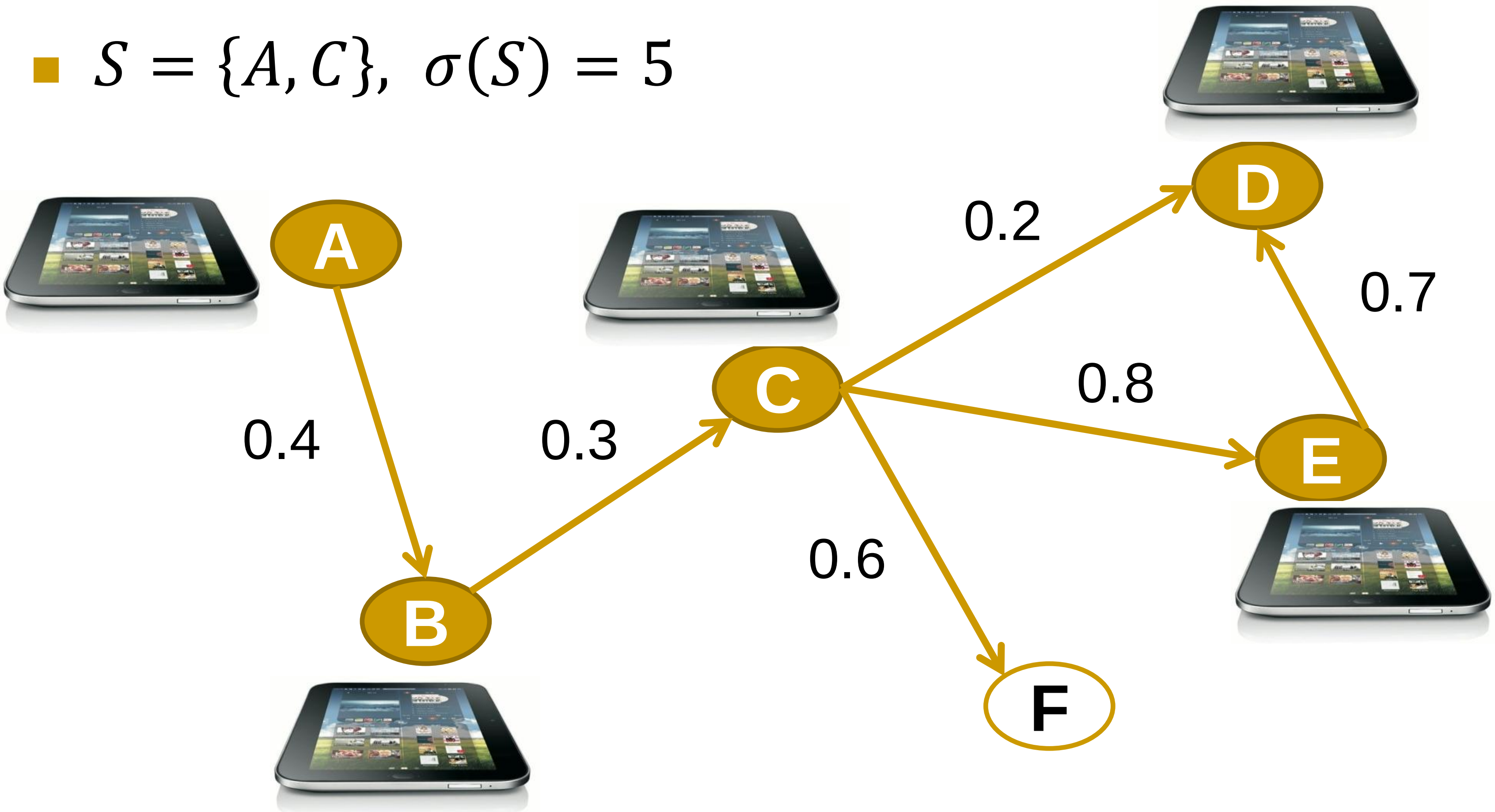


Model 1: Cascade Model



Model 1: Cascade Model

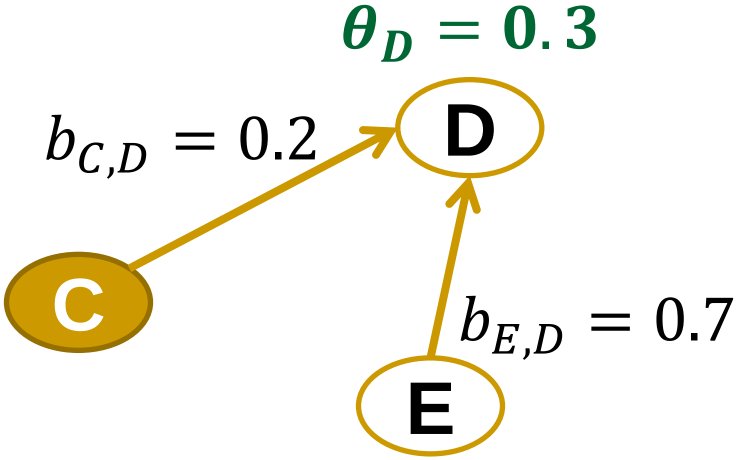
- $S = \{A, C\}, \ \sigma(S) = 5$



Model 2:
Linear Threshold Model

Model 2: Threshold Model

- Each inactive node picks a random $\theta_v \in [0,1]$
  - Active condition: $\sum_{u:\, active\ neighbor\ of\ v} b_{u,v} \geq \theta_v$



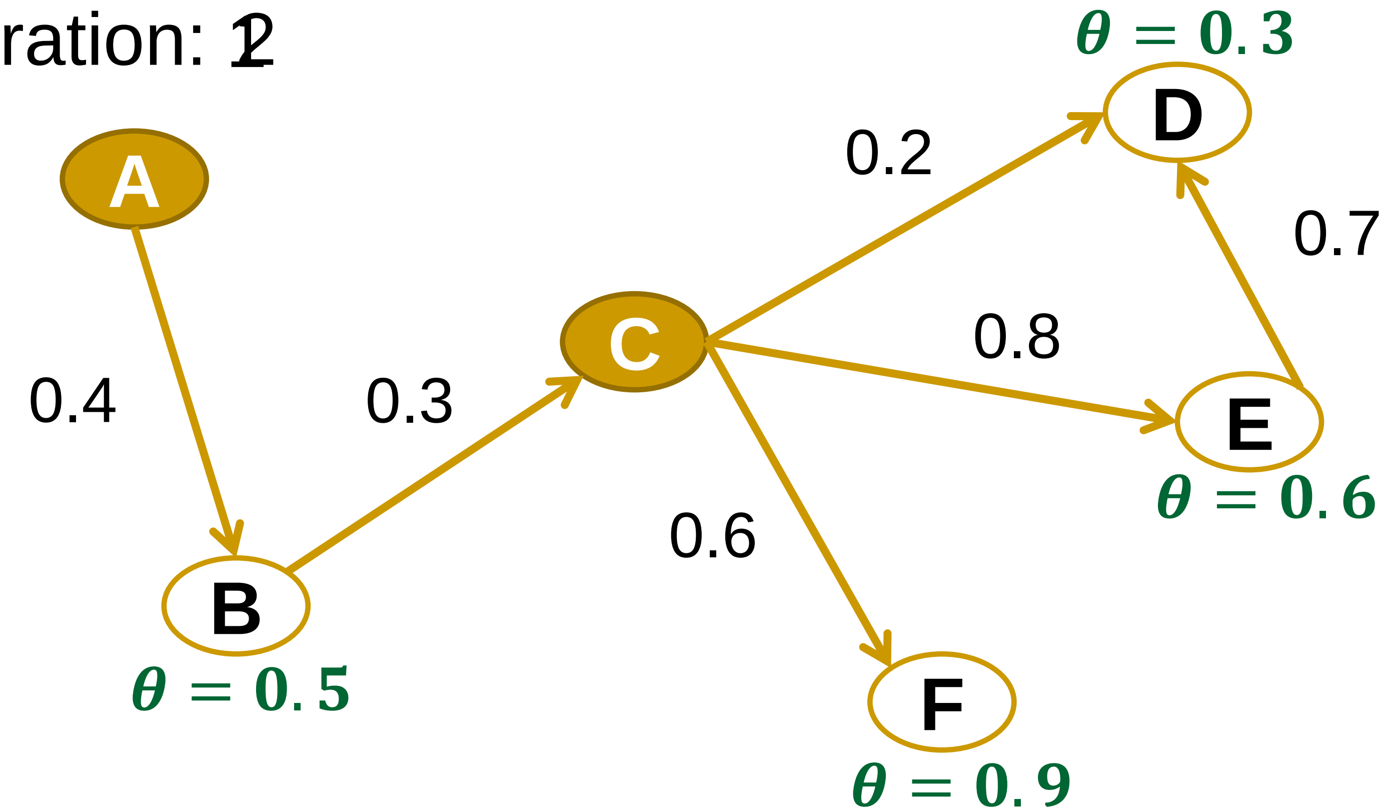
Iteration 2: 0.2 < 0.3

Iteration 4: E → active

Iteration 5: 0.2+0.7 > 0.3
D → active

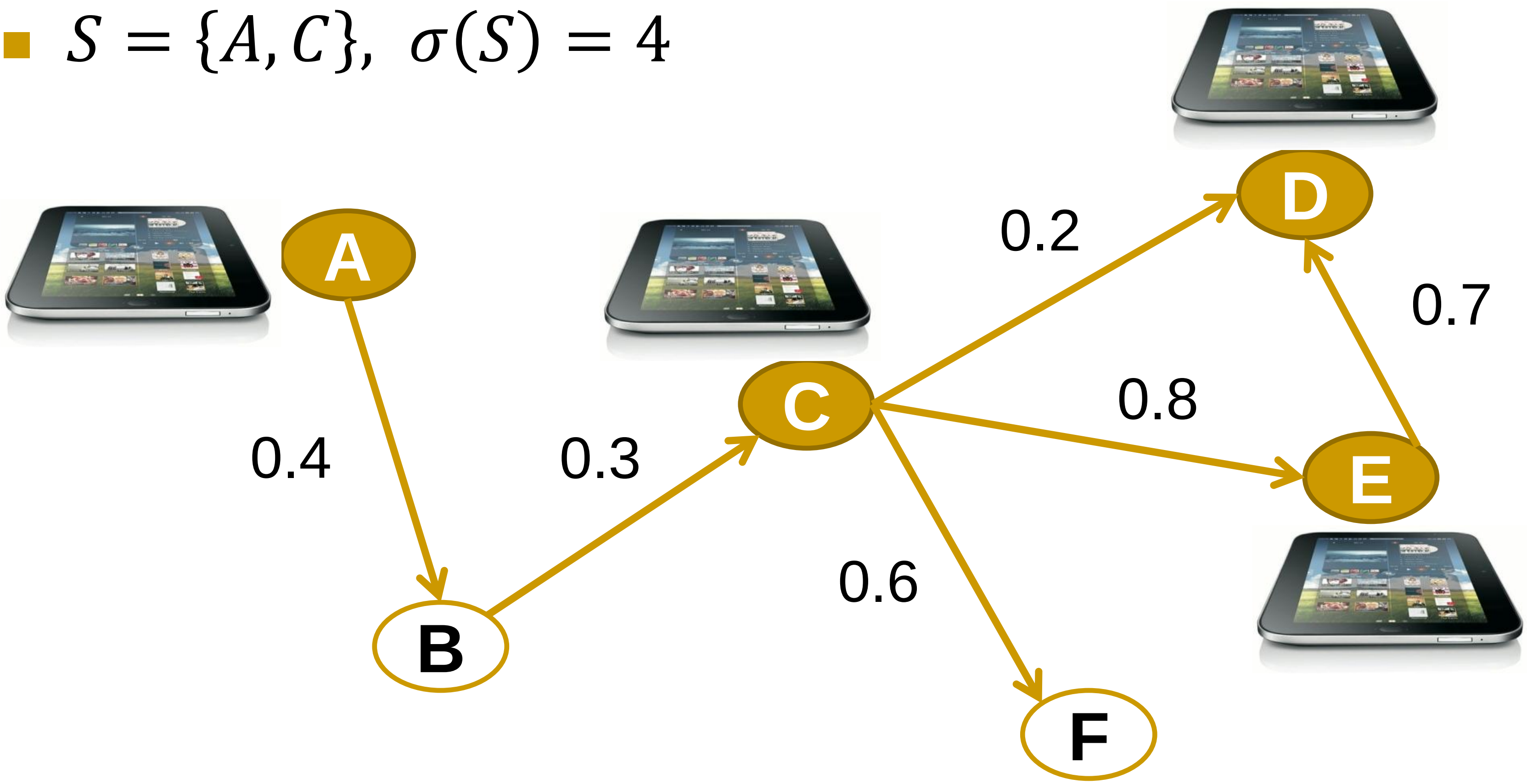
Model 2: Threshold Model

Iteration: 12

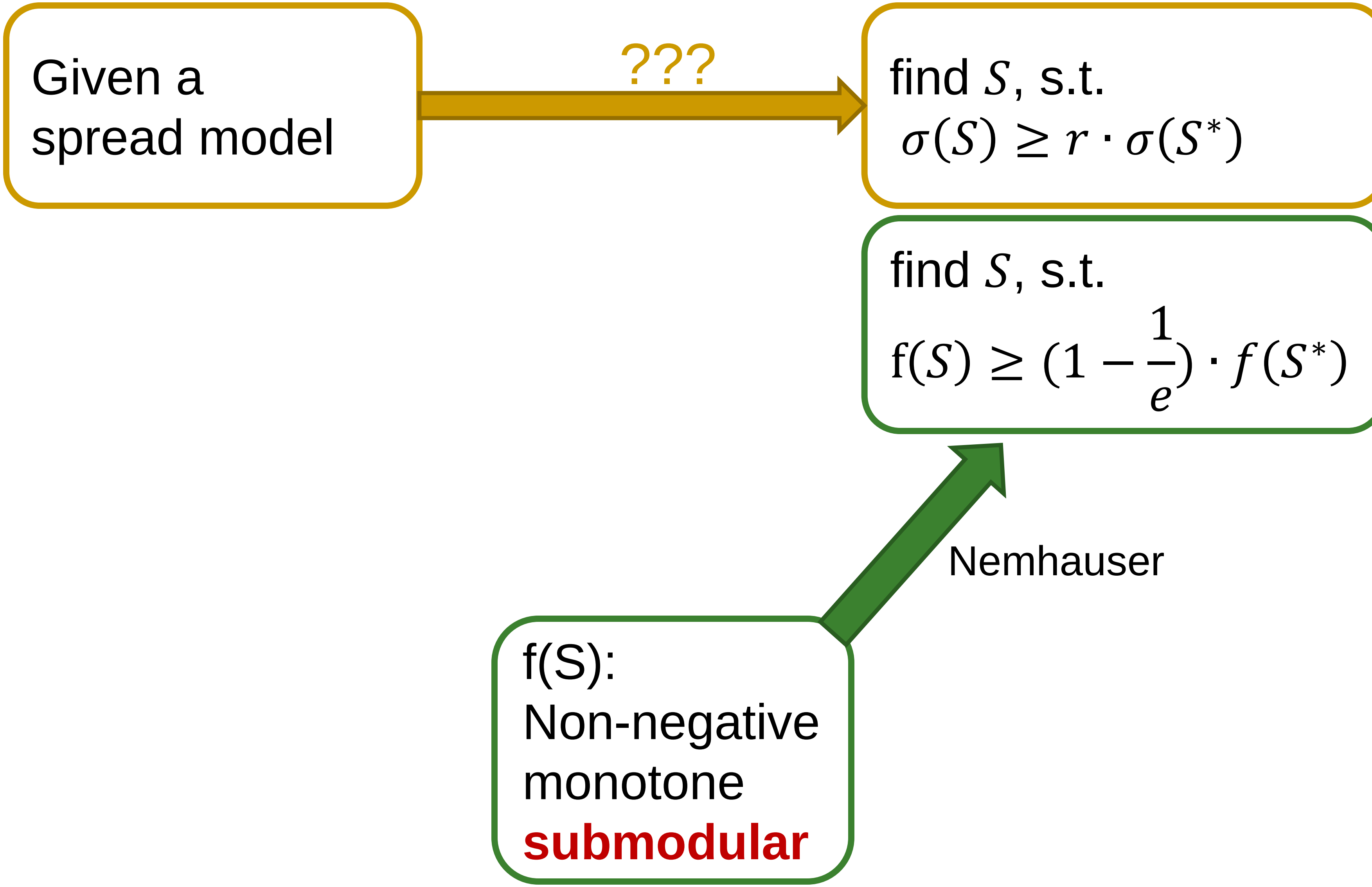


Model 2: Threshold Model

- $S = \{A, C\}, \ \sigma(S) = 4$



How to Prove the Guarantee?



Submodularity

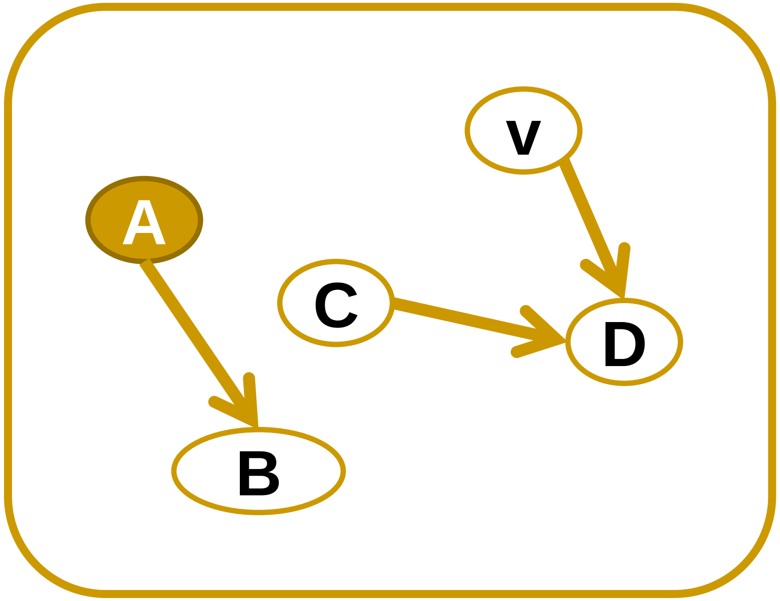
- U: a finite ground set
- $P(U)$: power set of U
- $f(\cdot)\colon P(U) \to R^*$

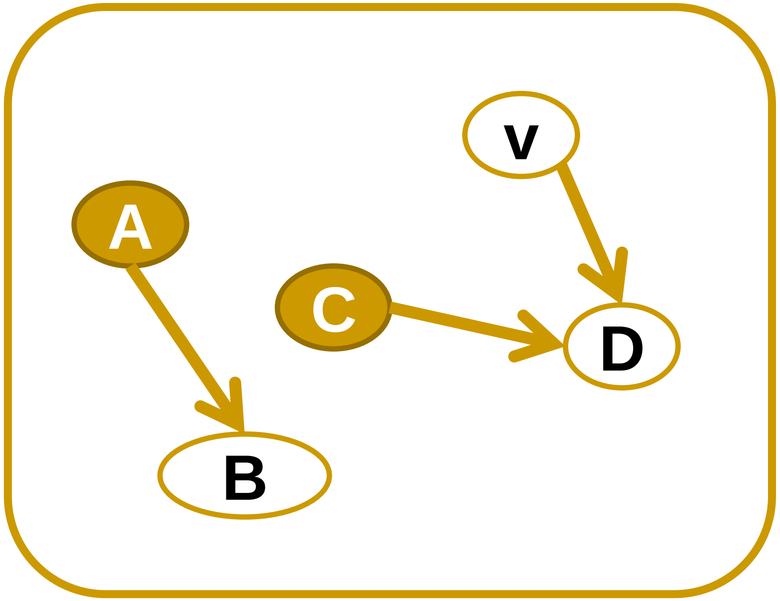
- Submodularity: $\forall\, node\ v, \forall S \subseteq T$

$$\boldsymbol{f}(\boldsymbol{S} \cup \{\boldsymbol{v}\}) - \boldsymbol{f}(\boldsymbol{S}) \geq \boldsymbol{f}(\boldsymbol{T} \cup \{\boldsymbol{v}\}) - \boldsymbol{f}(\boldsymbol{T})$$

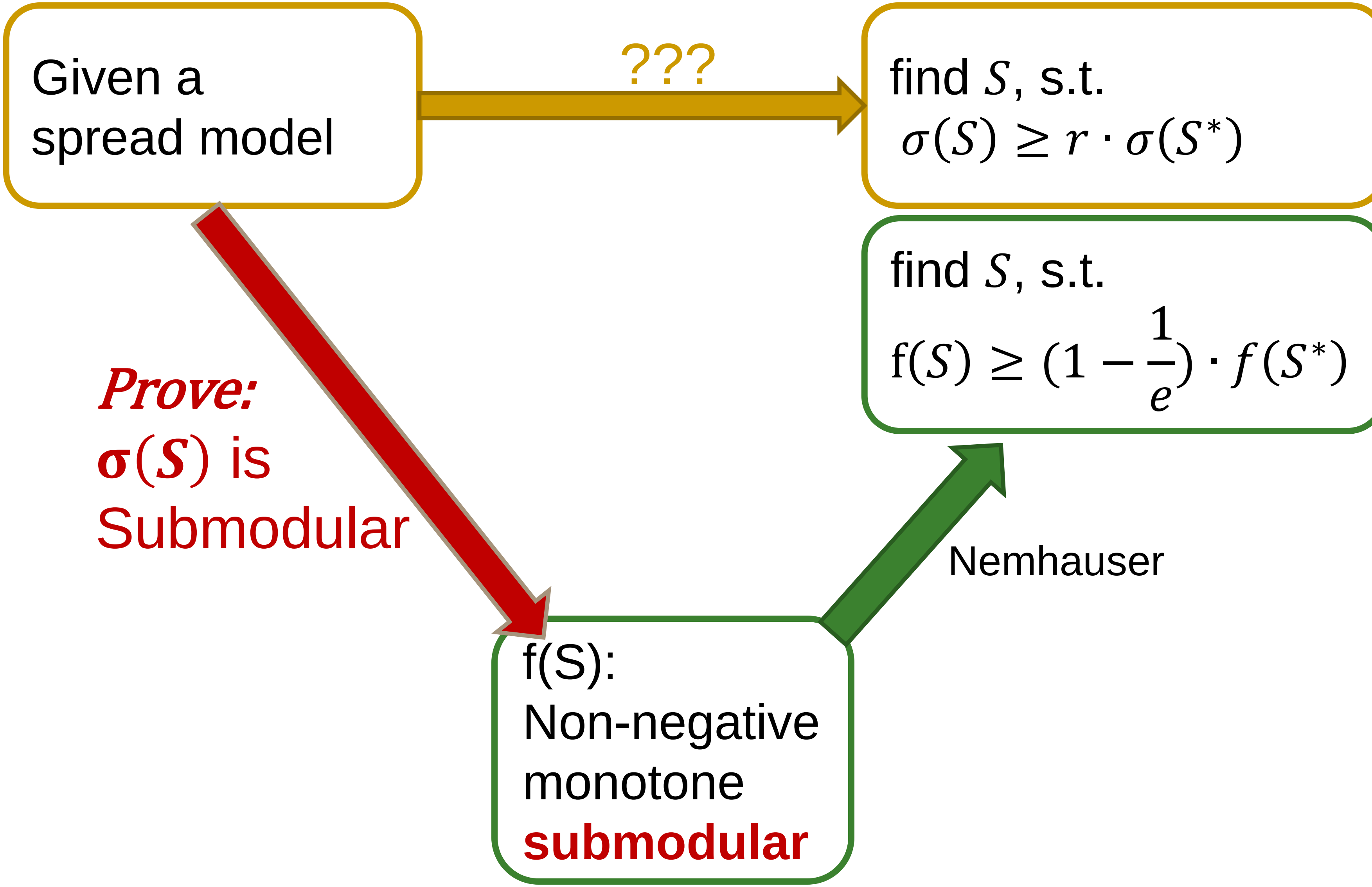
Example: Submodularity

- $f(S)$: number of vertexes reachable from vertexes in S





How to Prove the Guarantee?



We Want to Prove…

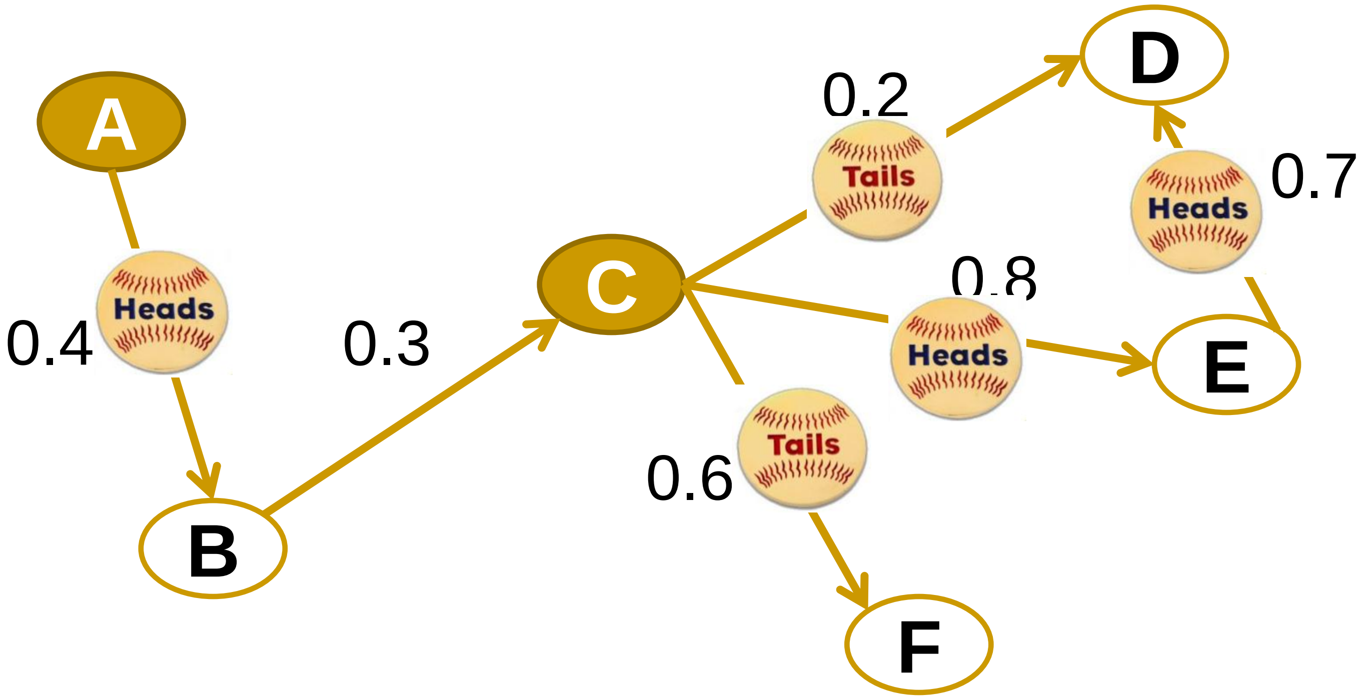
Model	$\boldsymbol{\sigma}(S)$ is Submodular	NP-hard
Independent Cascade	✓	✓
Linear Threshold	✓	✓

Prove: Submodularity

Cascade Model

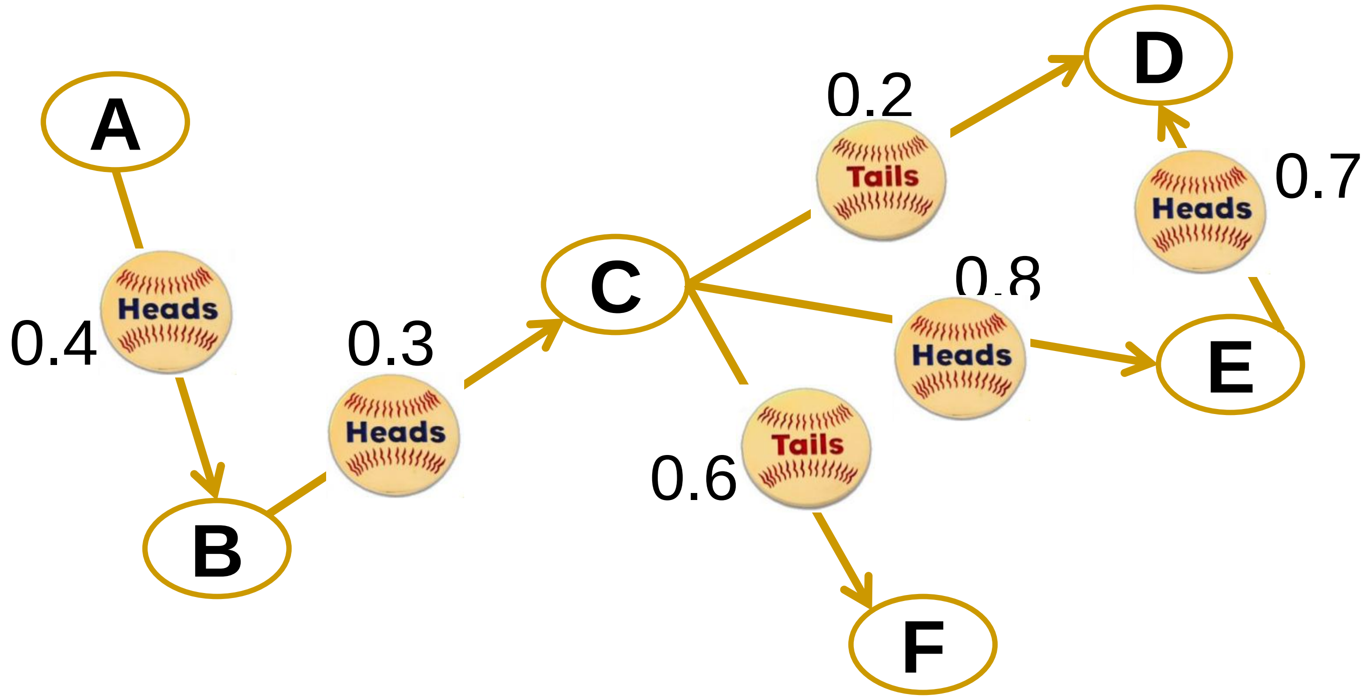
Submodularity (Cascade Model)

- Recall: flip coin



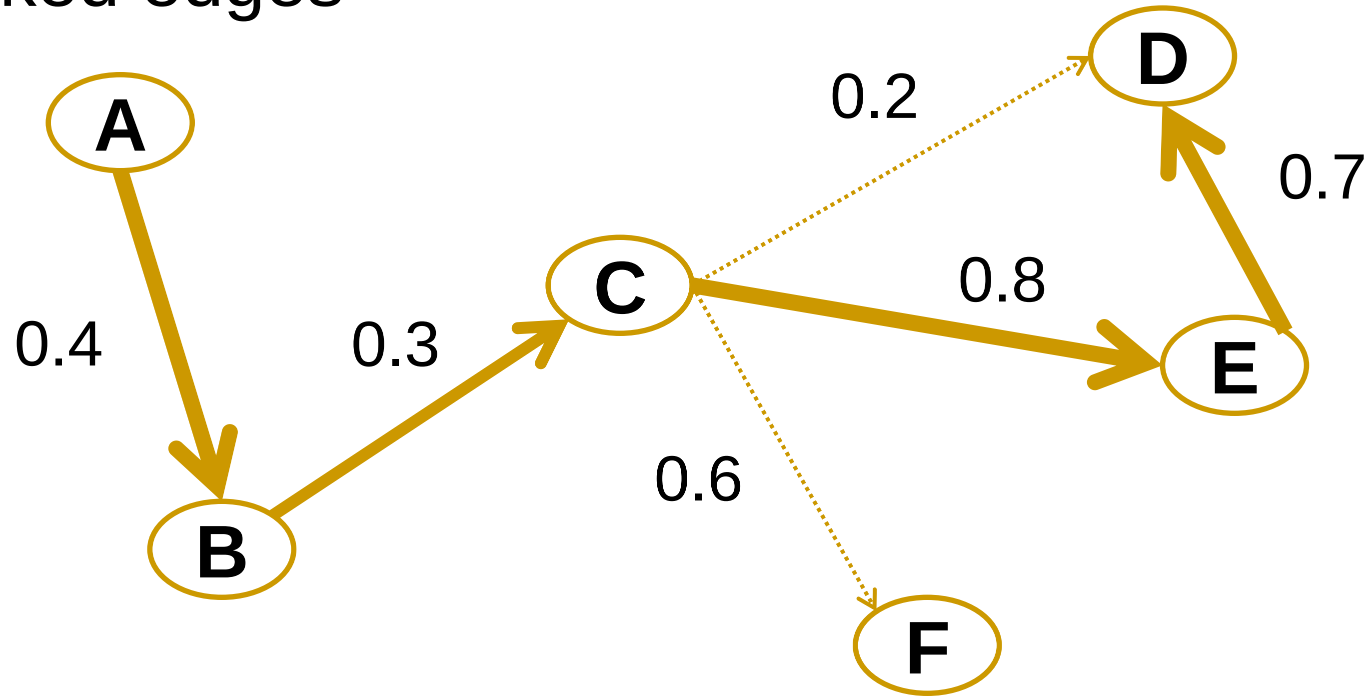
Submodularity (Cascade Model)

- Why not flip all the coins in the begining?



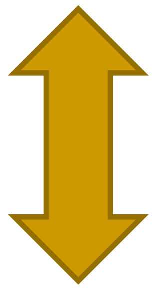
Submodularity (Cascade Model)

- Live edges → live paths
- blocked edges



Simplify Cascade Model

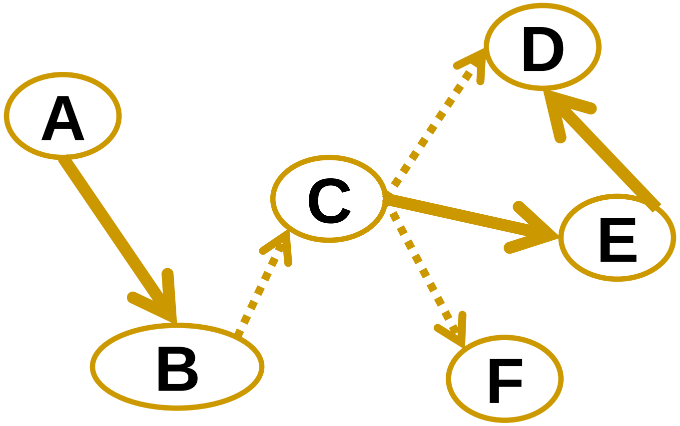
Node v ends up active



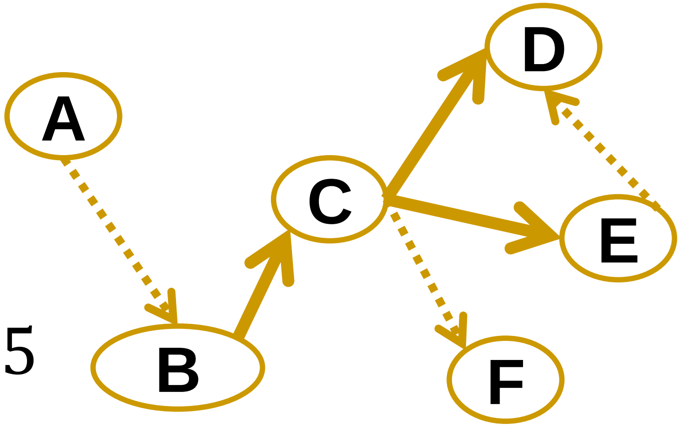
A live path: some seed → v

Achievement(Simplified Model)

- X: coin flipping outcome
  - e.g. X1, X2



- $R_X(v)$
  - $R_{X1}(A) = \{A, B\}$
  - $R_{X1}(C) = \{C, D, E\}$



- $\sigma_X(S) = |\cup_{v \in S} R_X(v)|$
  - $\sigma_{X1}(\{A, C\}) = |\{A, B, C, D, E\}| = 5$

Submodularity (Cascade Model)

- Fix x, $\sigma_X(S)$ is submodular 

- Linear combination of submodular functions is still submodular

$$\sigma(S) = \sum_X Prob[X] \cdot \sigma_X(S)$$

Summary of the proof

Active = Has a live path

$\sigma_X(S)$ is submodular

$\sigma(S)$ is submodular

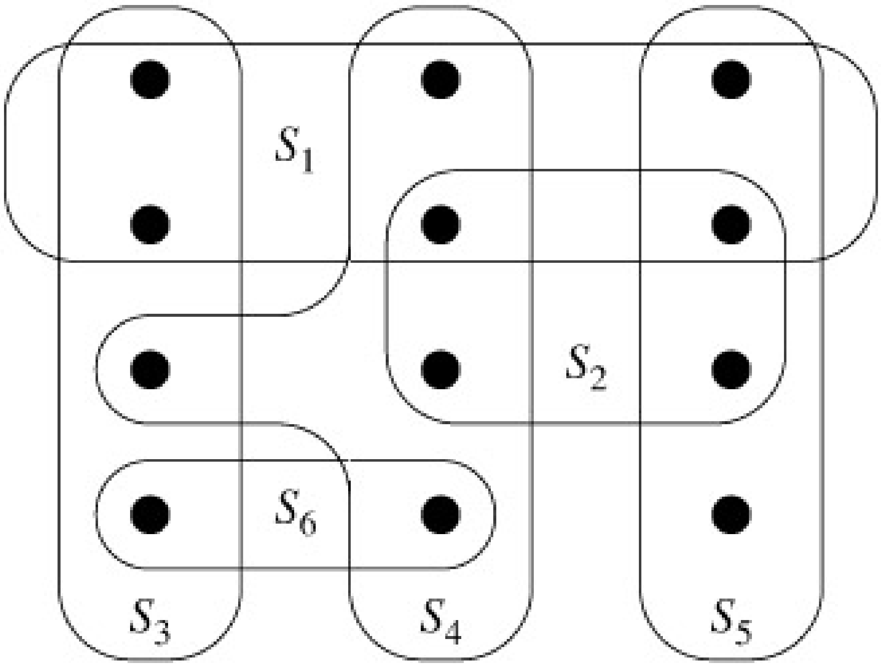
Prove:
NP-hard

Simplified Cascade Model

NP-Hard (Cascade Model)

- Set Cover Problem: k subsets cover all?

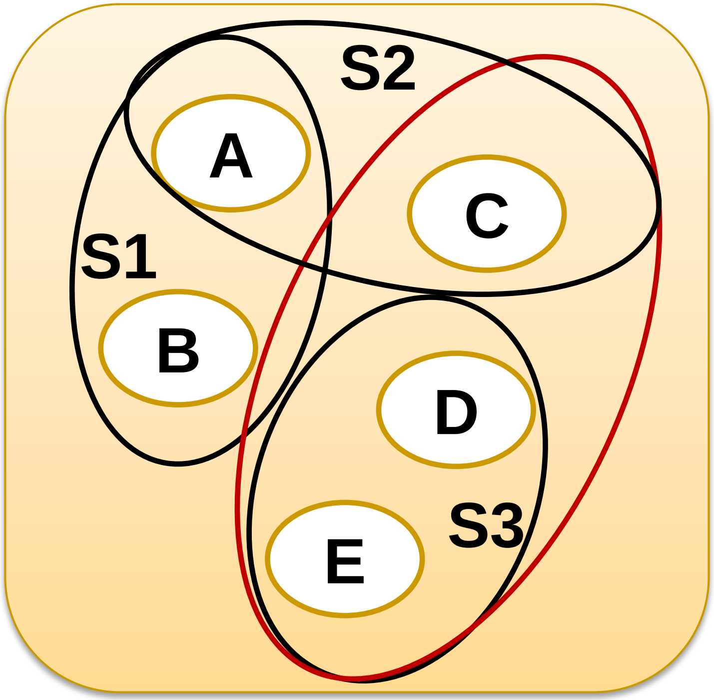
- K=1: No
- K=2: No
- K=3: Yes
- K=4: …

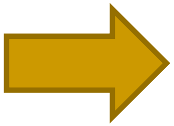


NP-Hard (Cascade Model)

- Solve Set Cover

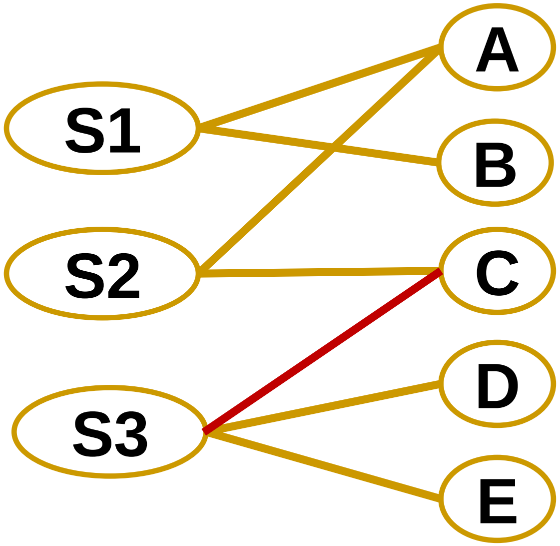
Q: 2 subsets cover all ?





- Influence maximization

Q: $|S| = 2, \sigma(S) \geq 2 + 5$?



NP-Hard (Cascade Model)

Influence Maximization Problem

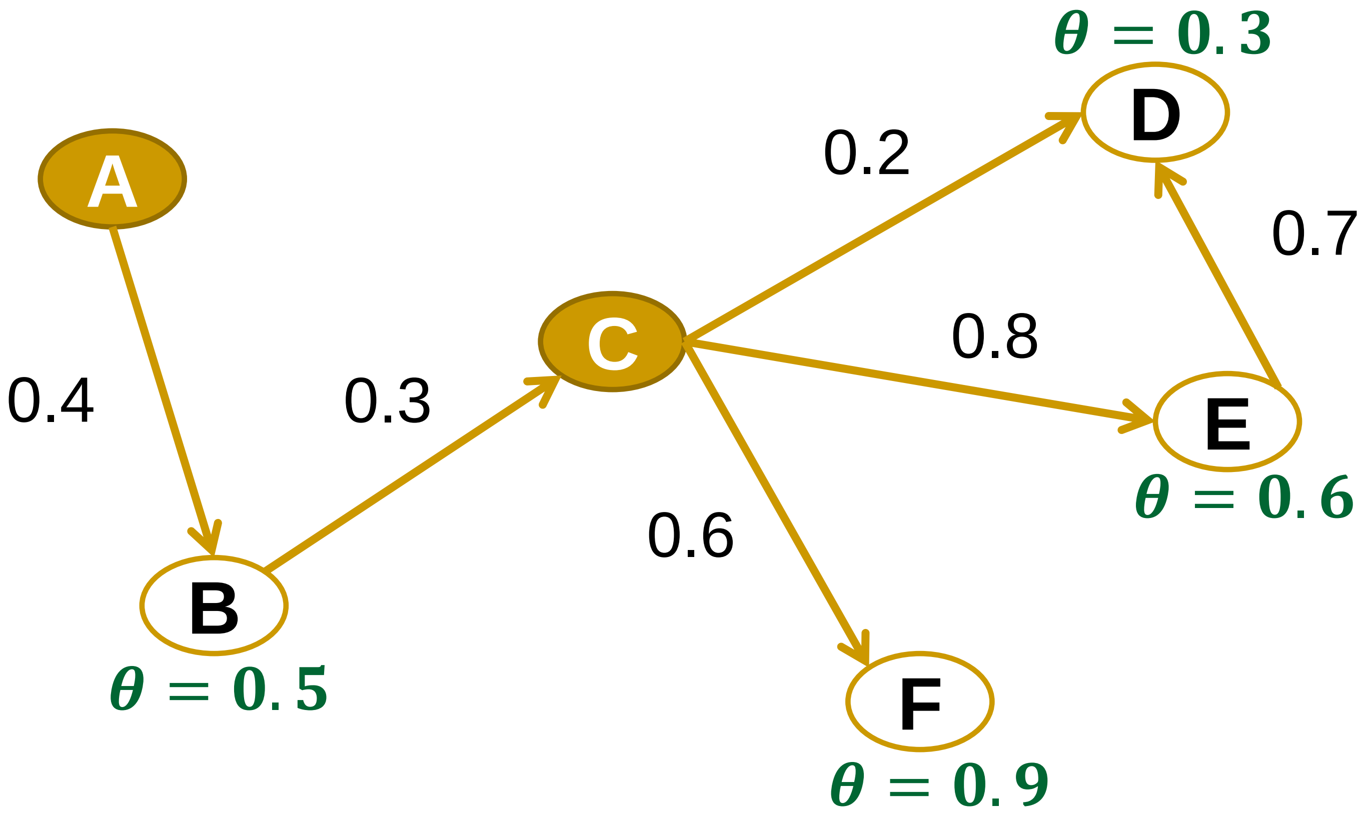
is at least as difficult as

Set Cover Problem

Prove: Submodularity

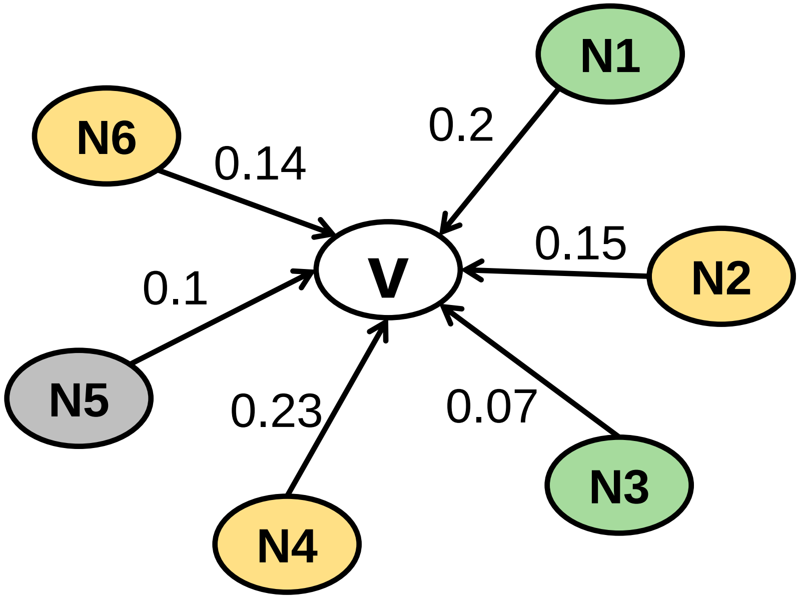
Linear Threshold Model

Recall: Threshold Model

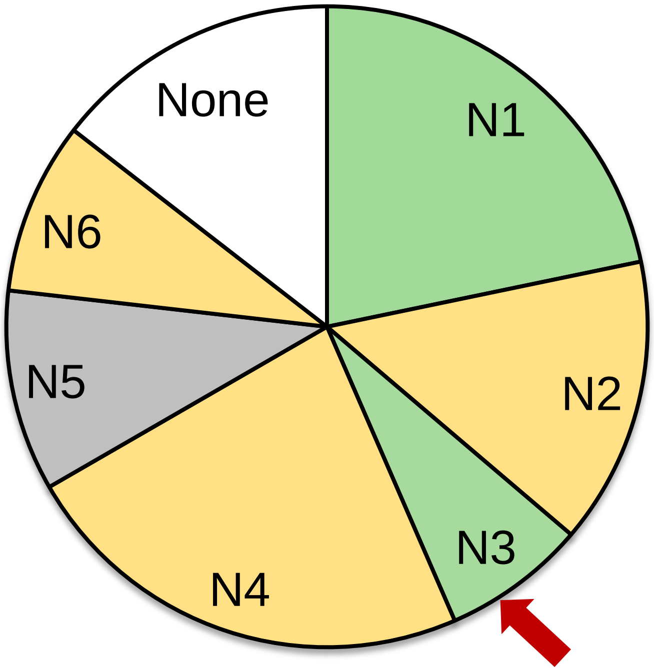


Gamble: Roulette

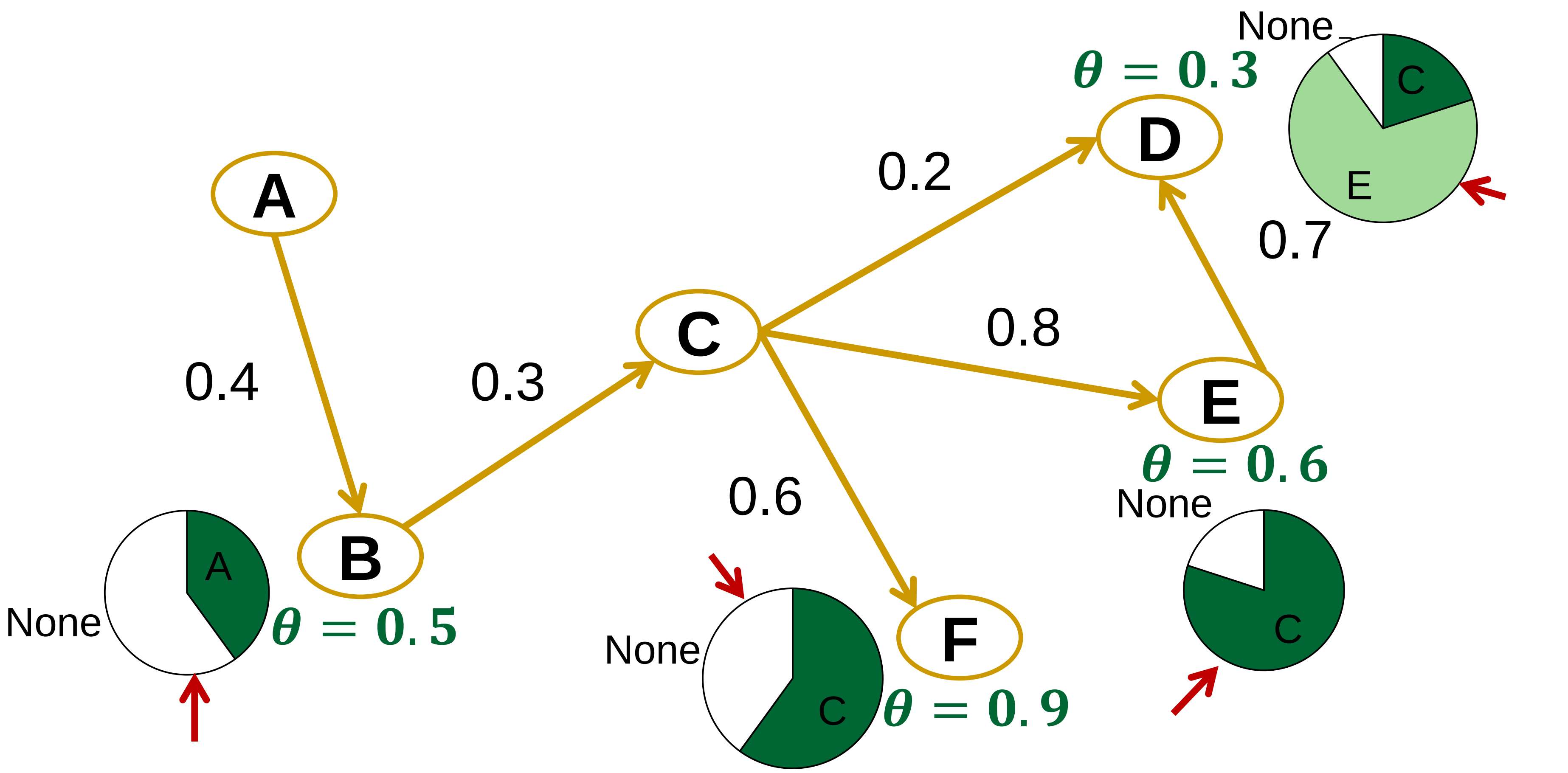
Gamble: Roulette



$\boldsymbol{\theta = 0.4}$

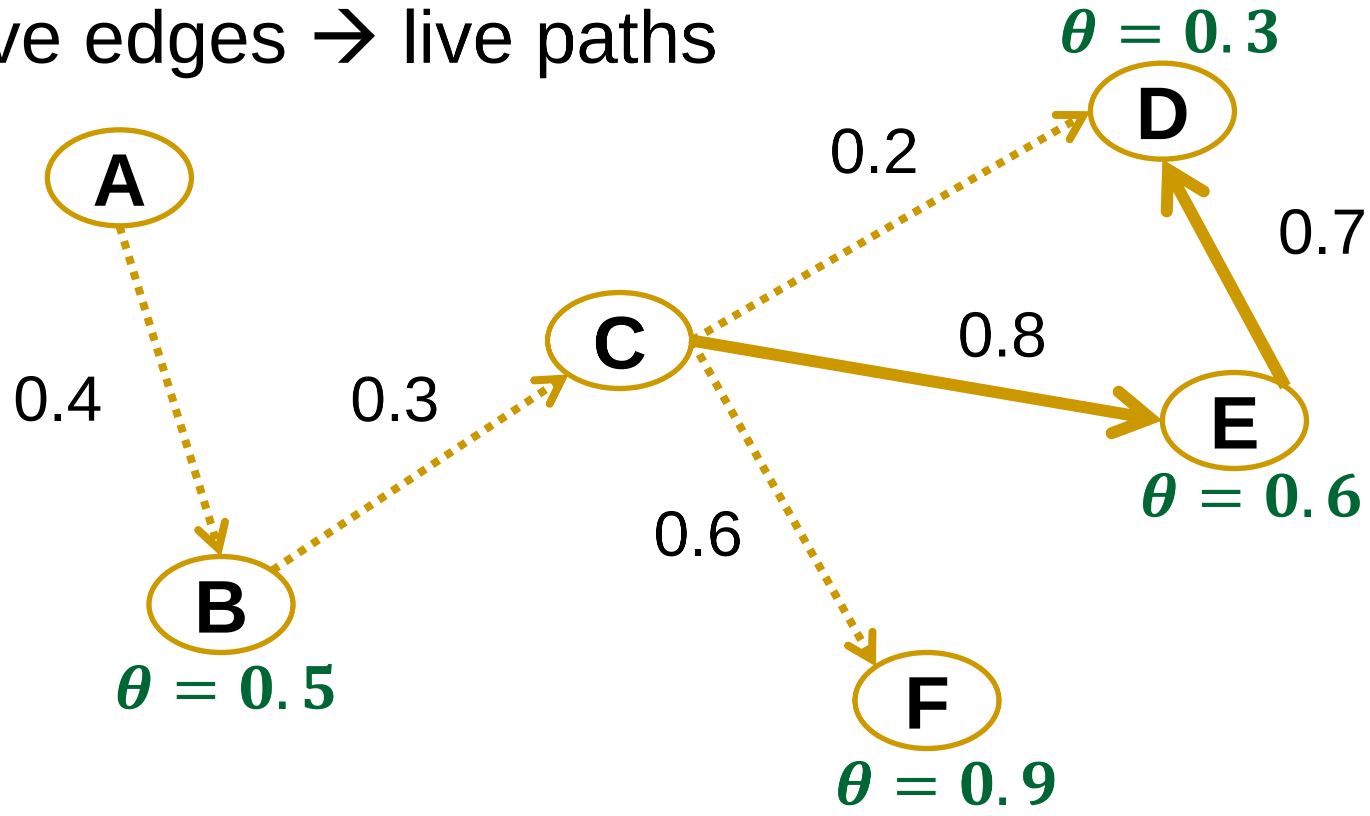


Submodularity (Threshold Model)



Submodularity (Threshold Model)

- Live edges → live paths



Correctness of Simplification

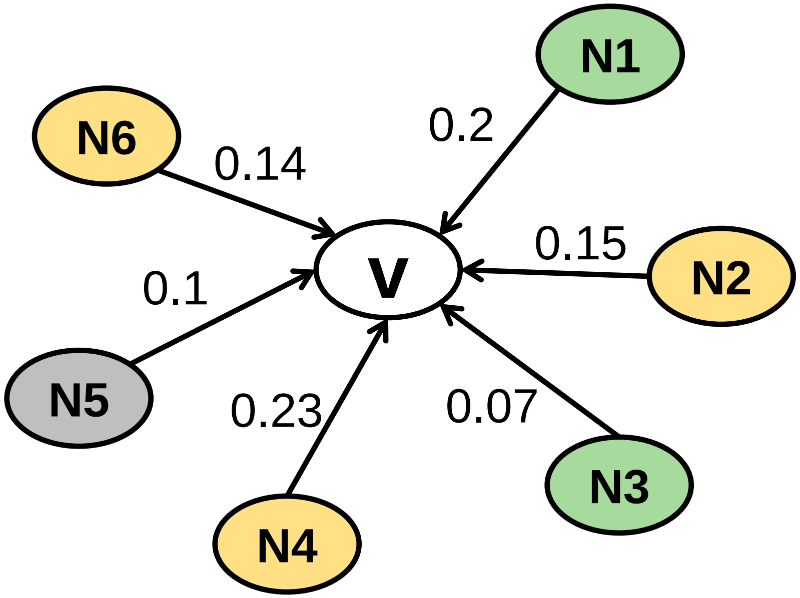
$$For\ node\ v:$$

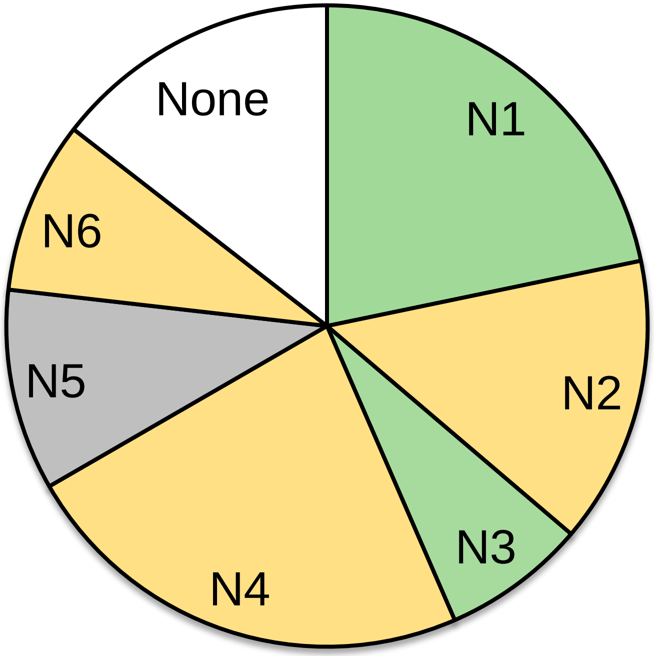
$$P(\ active\ in\ Iteration\ t+1\ |inactive\ in\ Iterations \leq t)$$

$$= \frac{P(active\ in\ Iteration\ t+1)}{P(inactive\ in\ Iterations \leq t)}$$

Simplified Model

Active before iteration 5
becomes active in iteration 5



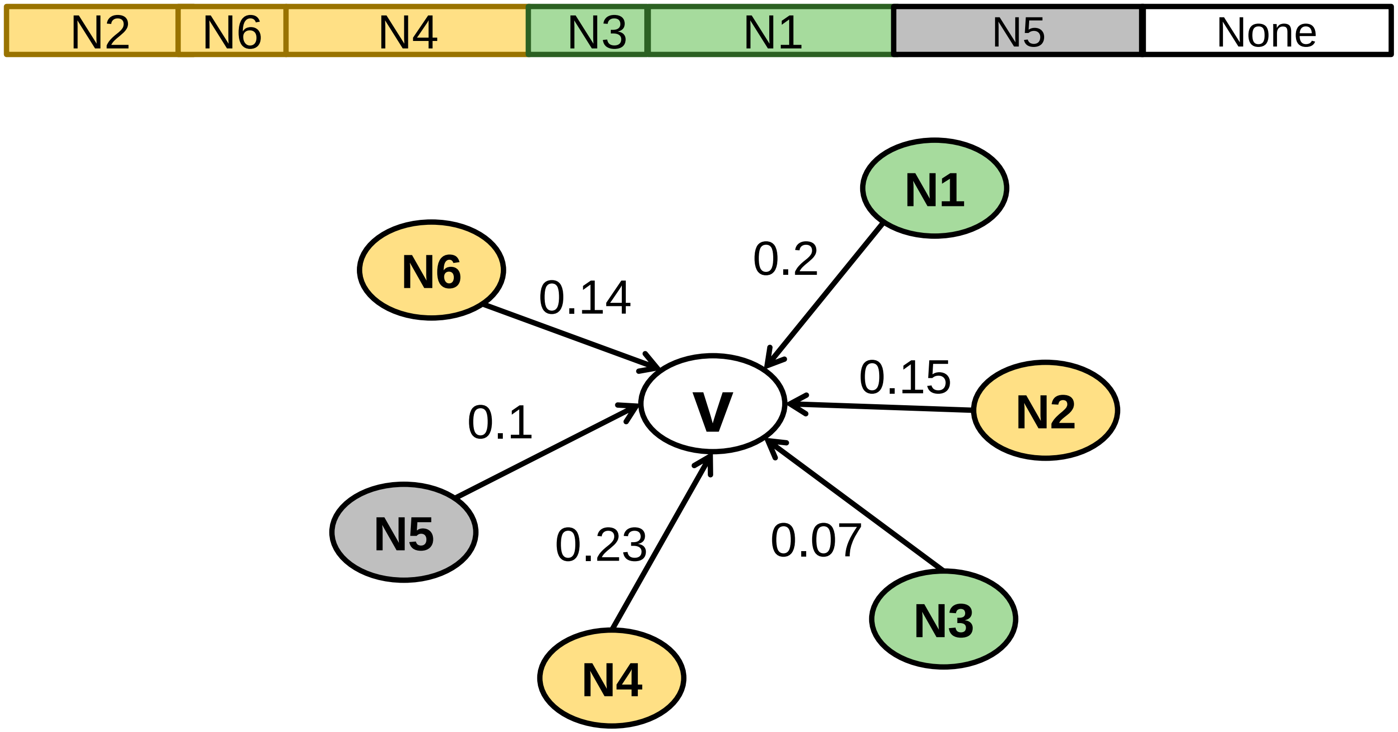


Simplified Model

A_t: Nodes becoming active in iteration t

$$\frac{\sum_{u \in A_t} b_{u,v}}{1 - \sum_{u \in A_1 \cup A_2 \cup \cdots \cup A_{t-1}} b_{u,v}}$$

Original Model



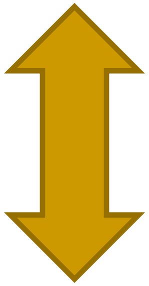
Original Model

A_t: Nodes becoming active in iteration t

$$\frac{\sum_{u \in A_t} b_{u,v}}{1 - \sum_{u \in A_1 \cup A_2 \cup \cdots \cup A_{t-1}} b_{u,v}}$$

Simplify Threshold Model

Node v ends up active



A live path: some seed → v

Similarly, we have…

Active = Has a live path

$\sigma_X(S)$ is submodular

$\sigma(S)$ is submodular

Prove:
NP-hard

Linear Threshold Model

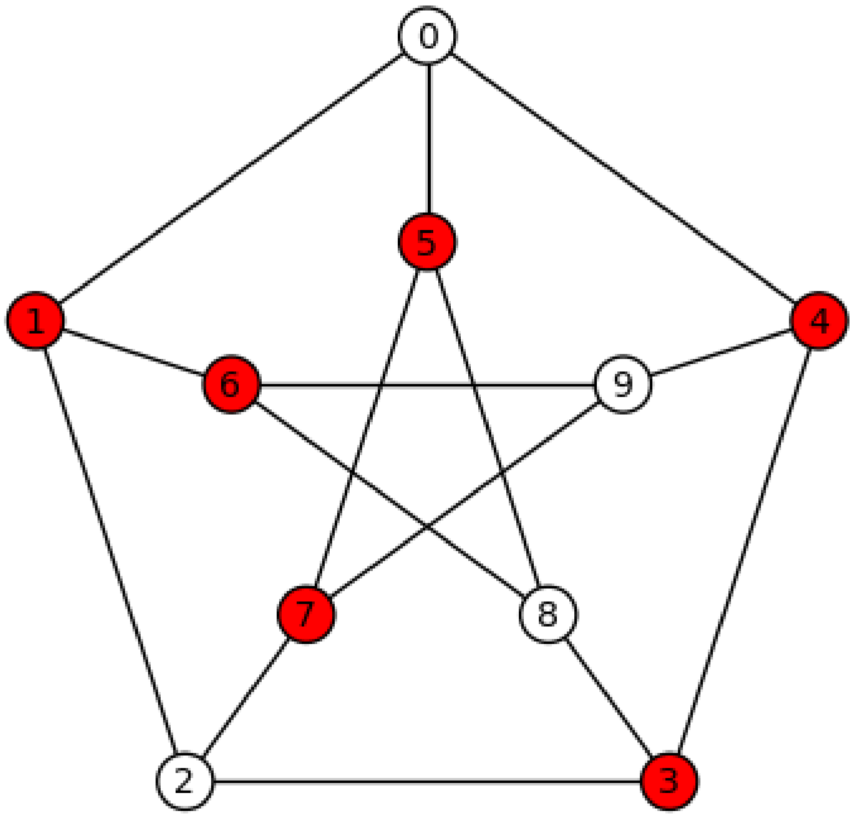
NP-Hard (Threshold Model)

- Vertex Cover Problen
  - k vertexes (S)

  each edge

  is incident to

  at least one vertex in S



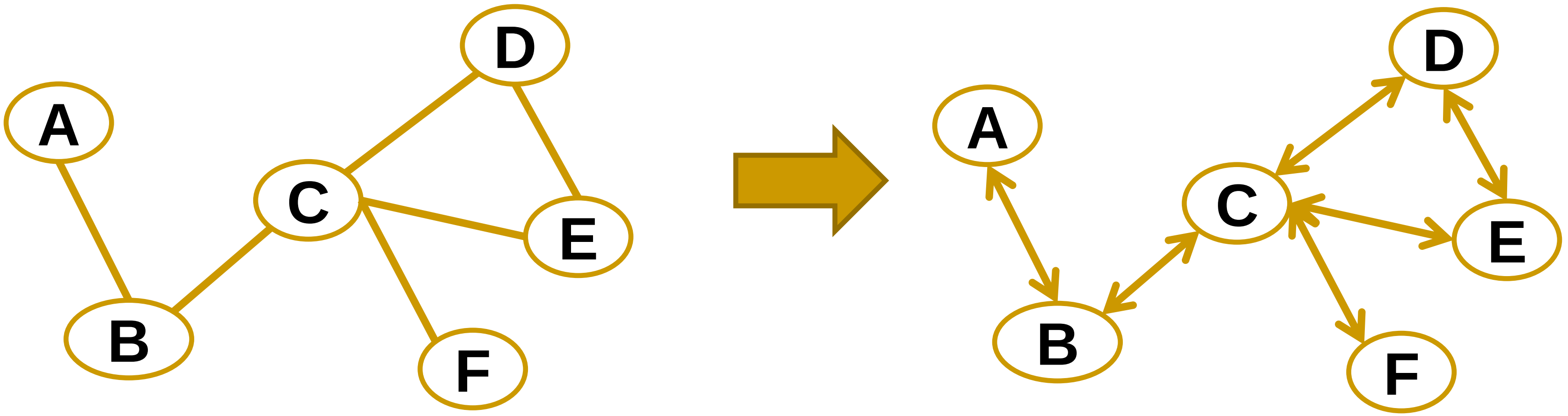
NP-Hard (Threshold Model)

- Vertex Set Cover

Q: 3 vertexes cover all ?

- Influence maximization

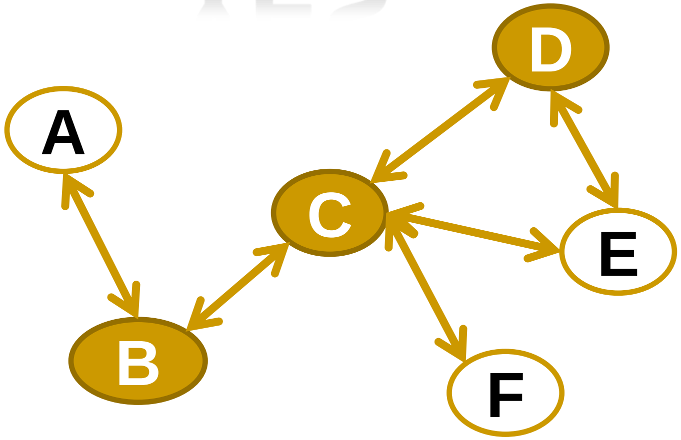
Q: $|S| = 3, \sigma(S) = 6$?



Influence Maximization

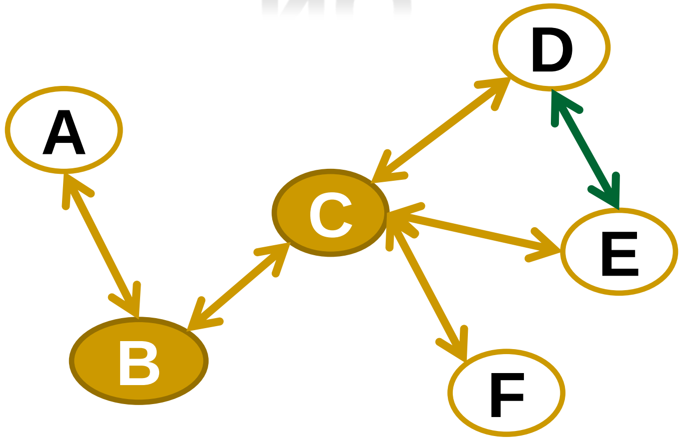
Q: $|S| = 3, \sigma(S) = 6?$

YES



Q: $|S| = 2, \sigma(S) = 6?$

NO



NP-Hard (Threshold Model)

Influence Maximization Problem

is at least as difficult as

Vertex Cover Problem

End of Proofs

- Influence Maximization Problem

Model	$\boldsymbol{\sigma}(S)$ is Submodular	NP-hard
Independent Cascade	✓	✓
Linear Threshold	✓	✓

Initial Problem

Given a spread model

find S, s.t.

$$\sigma(S) \geq \left(1 - \frac{1}{e} - \epsilon\right) \cdot \sigma(S^*)$$

Prove:

$\boldsymbol{\sigma(S)}$ is Submodular

f(S):

Non-negative monotone **submodular**

Greedy Hill Climbing

$\boldsymbol{MAX_v \; f(S \cup \{v\}) - f(S)}$

(Maximize Marginal Gain)

find S, s.t.

$$f(S) \geq \left(1 - \frac{1}{e}\right) \cdot f(S^*)$$

Summary

- Problem Description
- Two Models
  - Independent Cascade Model
  - Linear Threshold Model
- Submodular Functions
- Proof of Approximation Guarantee
- Proof of NP-Hardness

Q&A