Influence Maximization in the Cascade Model

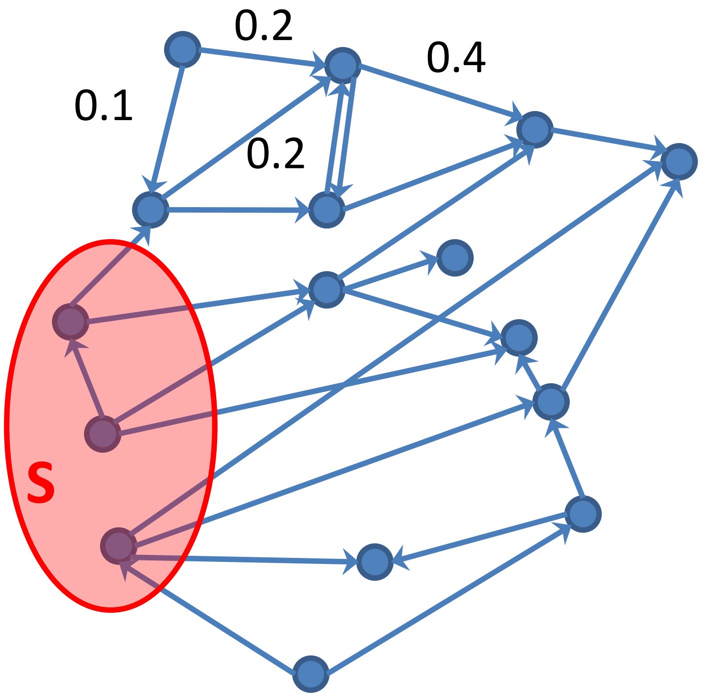
Finding Most Influential Nodes

- We want to find the set of nodes that can cause the highest effect to the network

- Applications:
  - Viral marketing: Find a set of users to give coupons
  - Network mining: Find out most important/infectious blogs

Influence Maximization

- We are given a graph, and probabilities on the edges.
- f(S): Expected # active nodes at the end with the cascade model if we start with a set S of active nodes
- Problem: Find set S: |S| ≤ k that maximizes f: $\max_{S \subset V : |S| \le k} f(S)$



The problem is NP-hard (reduction from set cover)

Can we show that f is nondecreasing and submodular?

Submodular Functions

- Let V be a set of elements
- Let f be a set function:

  $f: V \rightarrow R$

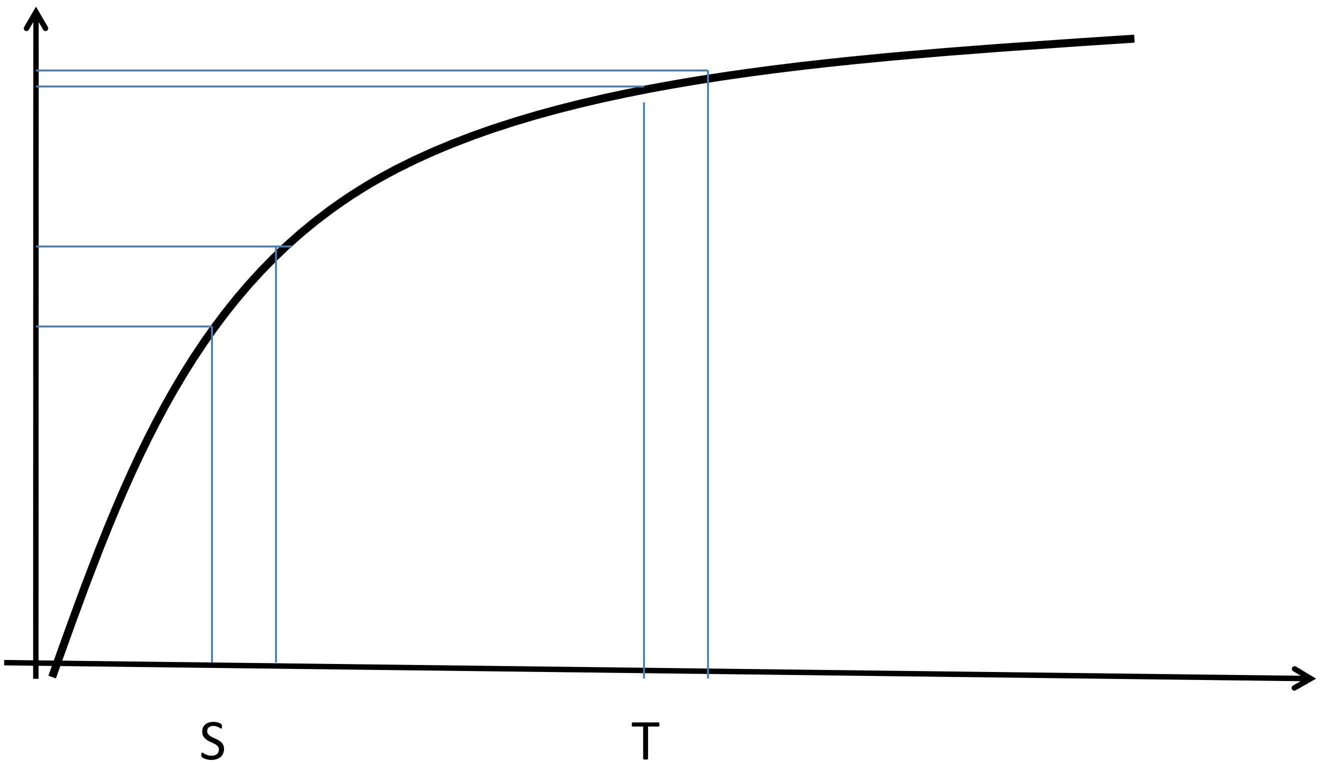
- f is **nondecreasing** if $f(S \cup \{v\}) - f(S) \geq 0$
- f is **submodular** if

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

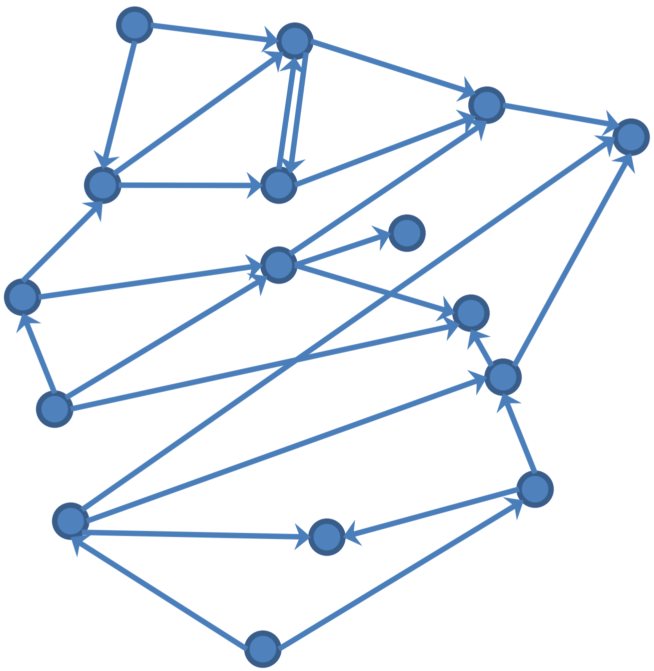
for $S \subset T$.

Submodular Functions II

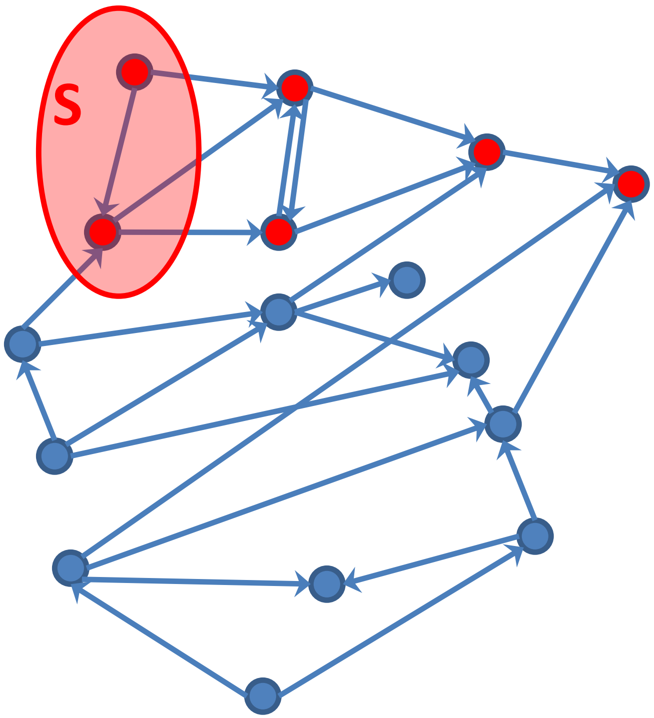
- Submodularity is similar to concavity (but for sets)
- **Diminishing returns**



Submodular Function Example



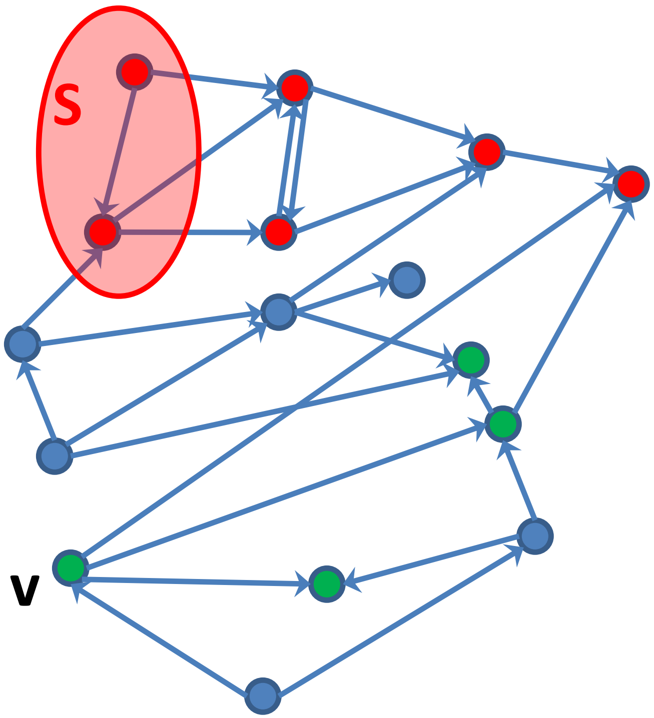
Submodular Function Example



S: set of nodes
R(S): Set of nodes reachable from S
f(S) = |R(S)| = # nodes reachable from S

Here:
f(S) = 6

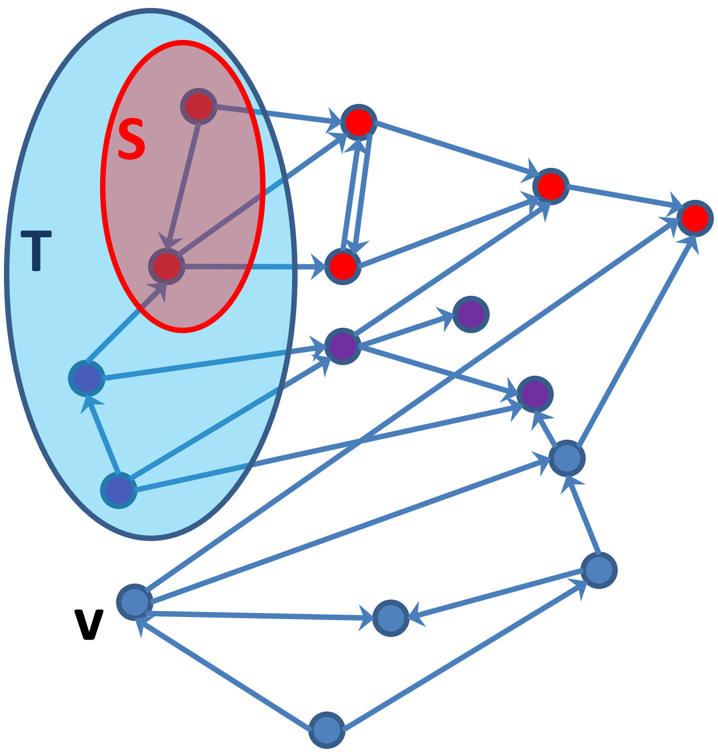
Submodular Function Example



S: set of nodes
R(S): Set of nodes reachable from S
f(S) = |R(S)| = # nodes reachable from S

Here:
f(S) = 6
f(S∪{v}) = 10
f(S∪{v}) – f(S) = 4

Submodular Function Example

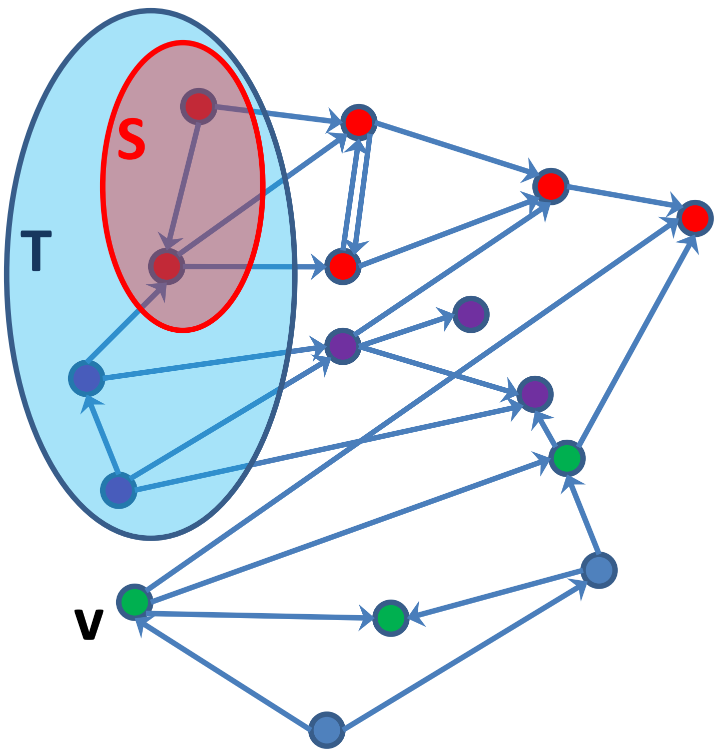


S: set of nodes
R(S): Set of nodes reachable from S
f(S) = |R(S)| = # nodes reachable from S

Here:
f(S) = 6
f(S∪{v}) = 10
f(S∪{v}) – f(S) = 4

f(T) = 11

Submodular Function Example



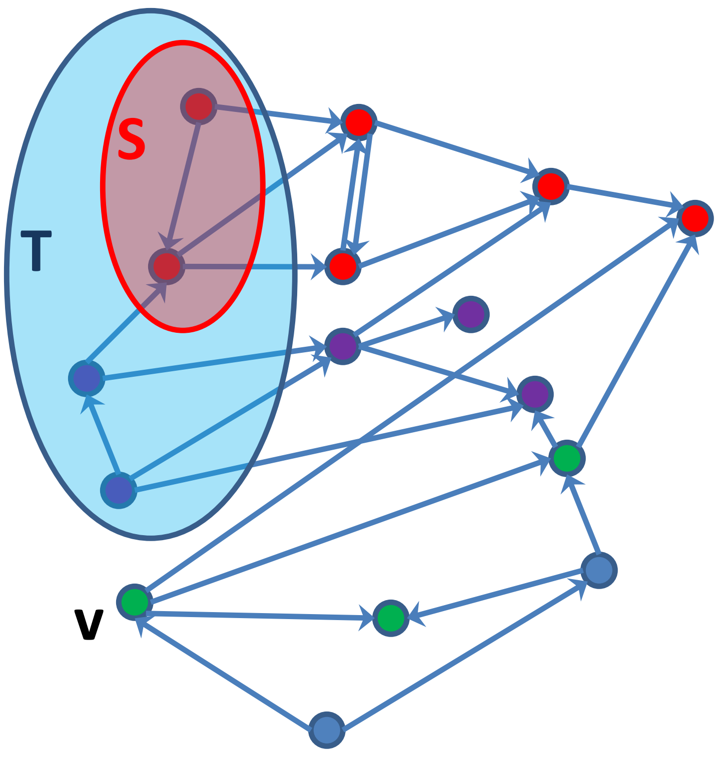
S: set of nodes
R(S): Set of nodes reachable from S
f(S) = |R(S)| = # nodes reachable from S

Here:
f(S) = 6
f(S∪{v}) = 10
f(S∪{v}) – f(S) = 4

f(T) = 11
f(T∪{v}) = 14
f(T∪{v}) – f(T) = 3

f(S∪{v}) – f(S) ≥ f(T∪{v}) – f(T)

Submodular Function Example



S: set of nodes
R(S): Set of nodes reachable from S
f(S) = |R(S)| = # nodes reachable from S

Here:
f(S) = 6
f(S∪{v}) = 10
f(S∪{v}) – f(S) = 4

f(T) = 11
f(T∪{v}) = 14
f(T∪{v}) – f(T) = 3

f(S∪{v}) – f(S) ≥ f(T∪{v}) – f(T)
f is a submodular function

Whatever I gain by adding v to T
I also gain by adding v to S

Submodular Function Maximization

- Consider a set function f: V → R that is **nondecreasing** and **submodular**
- We want to find a subset S of k elements from V that maximizes f:

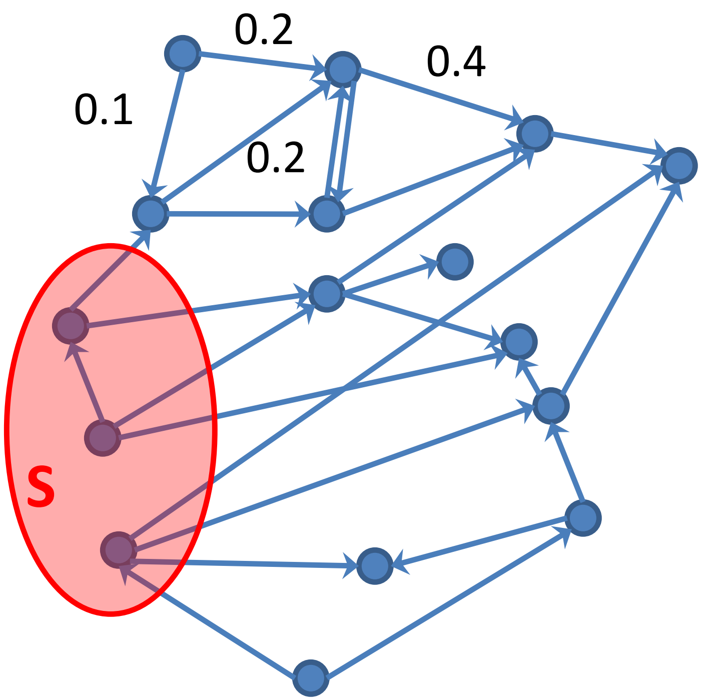
$$\max_{S \subset V : |S| \leq k} f(S)$$

- An easy strategy is the **greedy**:
  - S = ∅
  - While (|S| < k)
    - Find an element v that maximizes f(SU{v})
    - S = SU{v})
  - Return S

- **Theorem.** The **greedy** algorithm gives a (1-1/e) ≈ 0.63 approximation.

Back to Influence Maximization

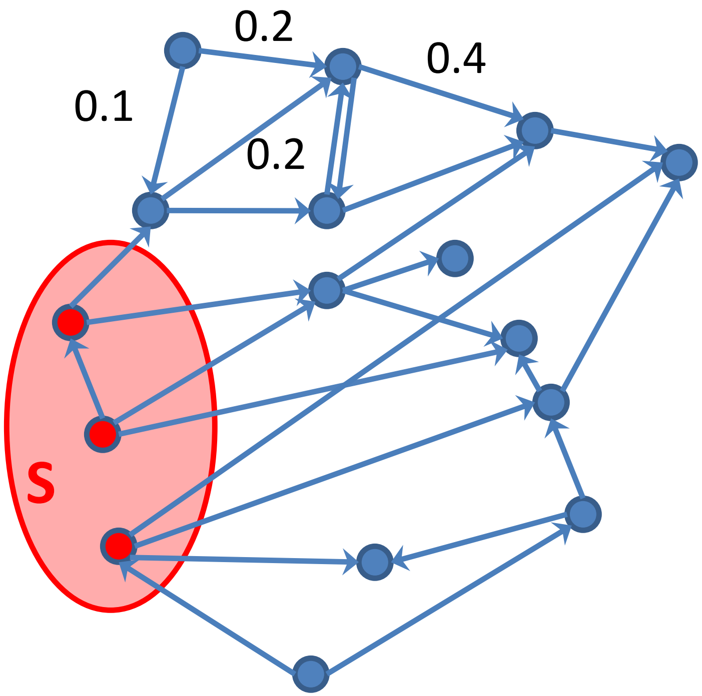
- We are given a graph, and probabilities on the edges.
- f(S): Expected # active nodes at the end with the cascade model if we start with a set S of active nodes
- Problem: Find set S: |S| ≤ k that maximizes f: $\max_{S \subset V : |S| \leq k} f(S)$



Can we show that f is nondecreasing and submodular?

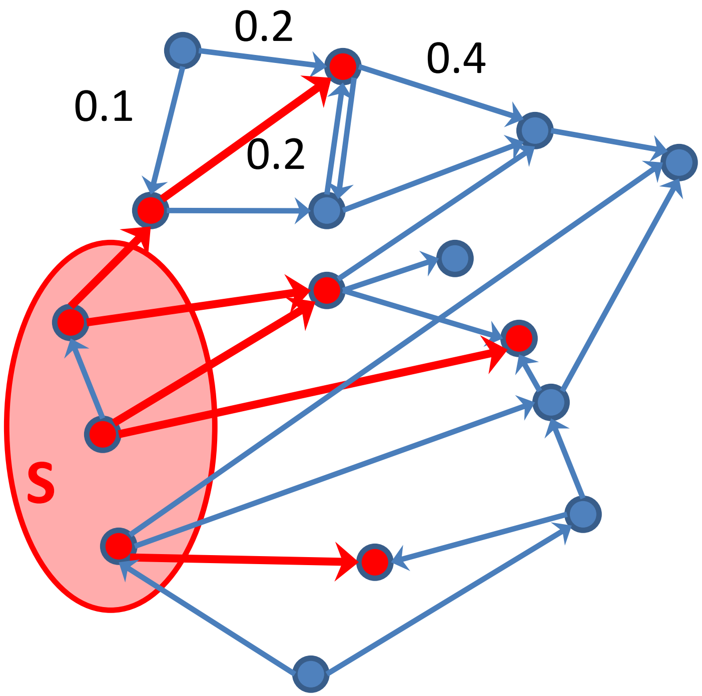
If we show it then we can get a (1-1/e) approximation.

Show that f(S) is submodular



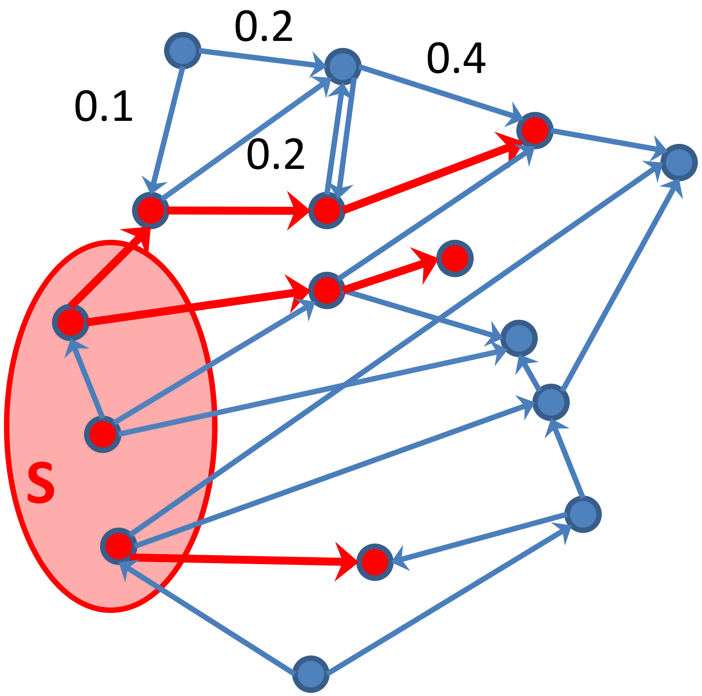
- Fix a set S and consider a particular scenario ω of the cascade model .
- f(S, ω): # active nodes at the end
- Then $f(S) = \sum_{\omega} \Pr(\omega) \cdot f(S, \omega)$

Show that f(S) is submodular



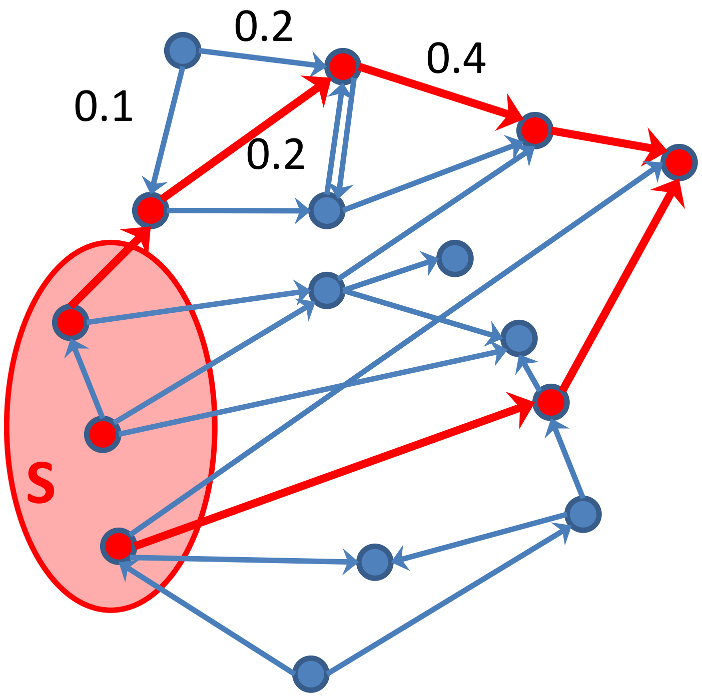
- Fix a set S and consider a particular scenario ω of the cascade model .
- f(S, ω): # active nodes at the end
- Then $f(S) = \sum_{\omega} \Pr(\omega) \cdot f(S, \omega)$

Show that f(S) is submodular



- Fix a set S and consider a particular scenario ω of the cascade model .
- f(S, ω): # active nodes at the end
- Then $f(S) = \sum_{\omega} \Pr(\omega) \cdot f(S, \omega)$

Show that f(S) is submodular

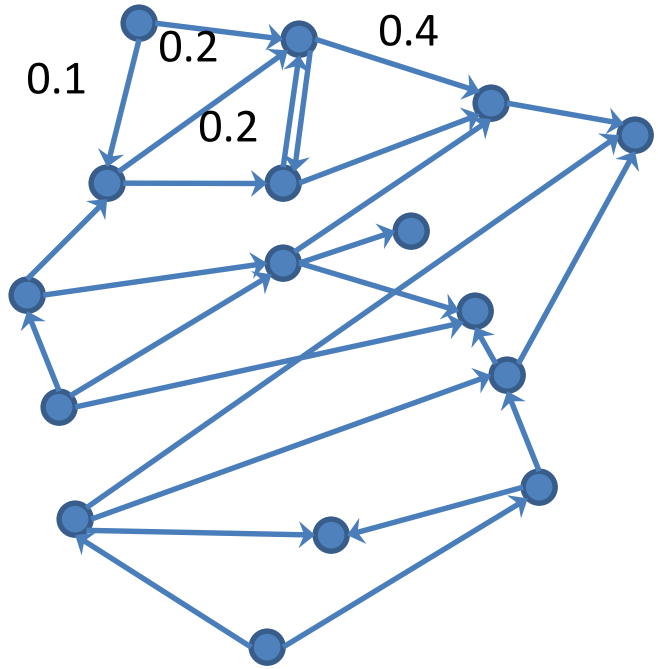


- Fix a set S and consider a particular scenario ω of the cascade model .
- f(S, ω): # active nodes at the end
- Then $f(S) = \sum_{\omega} \Pr(\omega) \cdot f(S,\omega)$

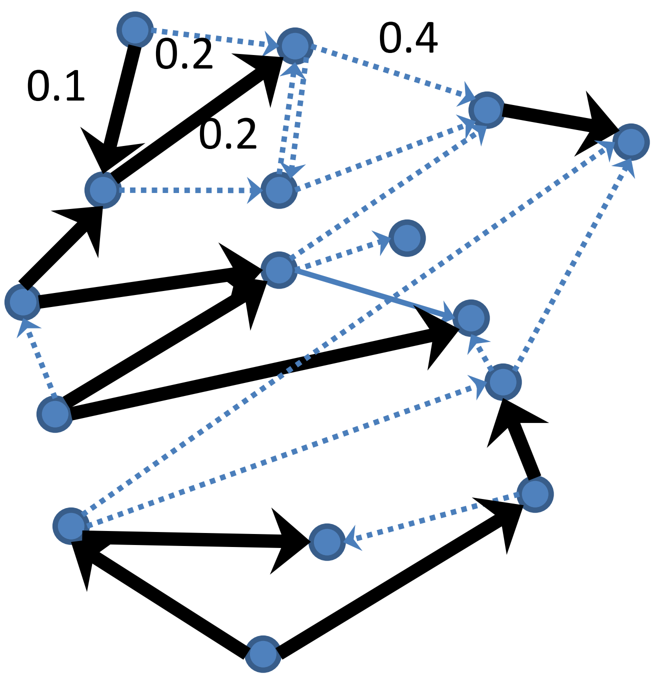
Show that g(S) = f(S,ω) is submodular

- We first show that for a fixed scenario ω, g(S) = f(S,ω) is submodular.
- To show that we will view the cascading model in a different way

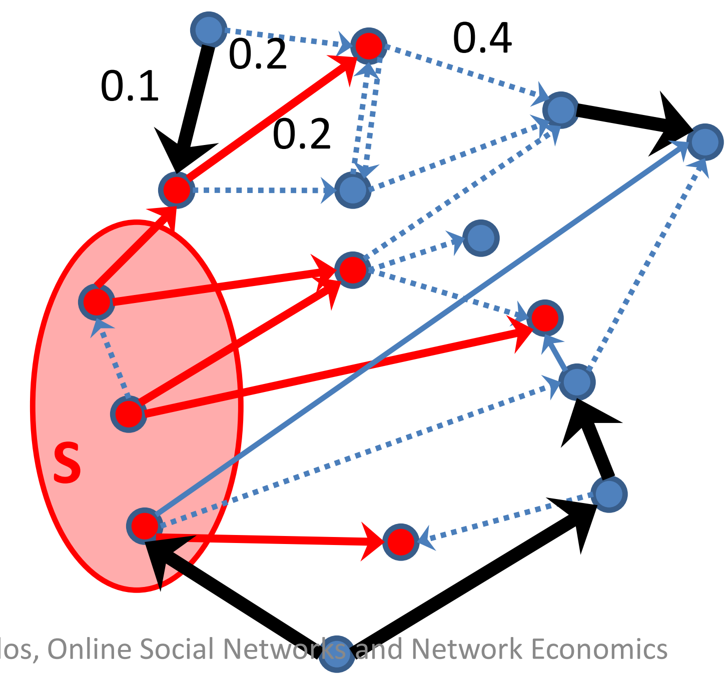
A different view of the process



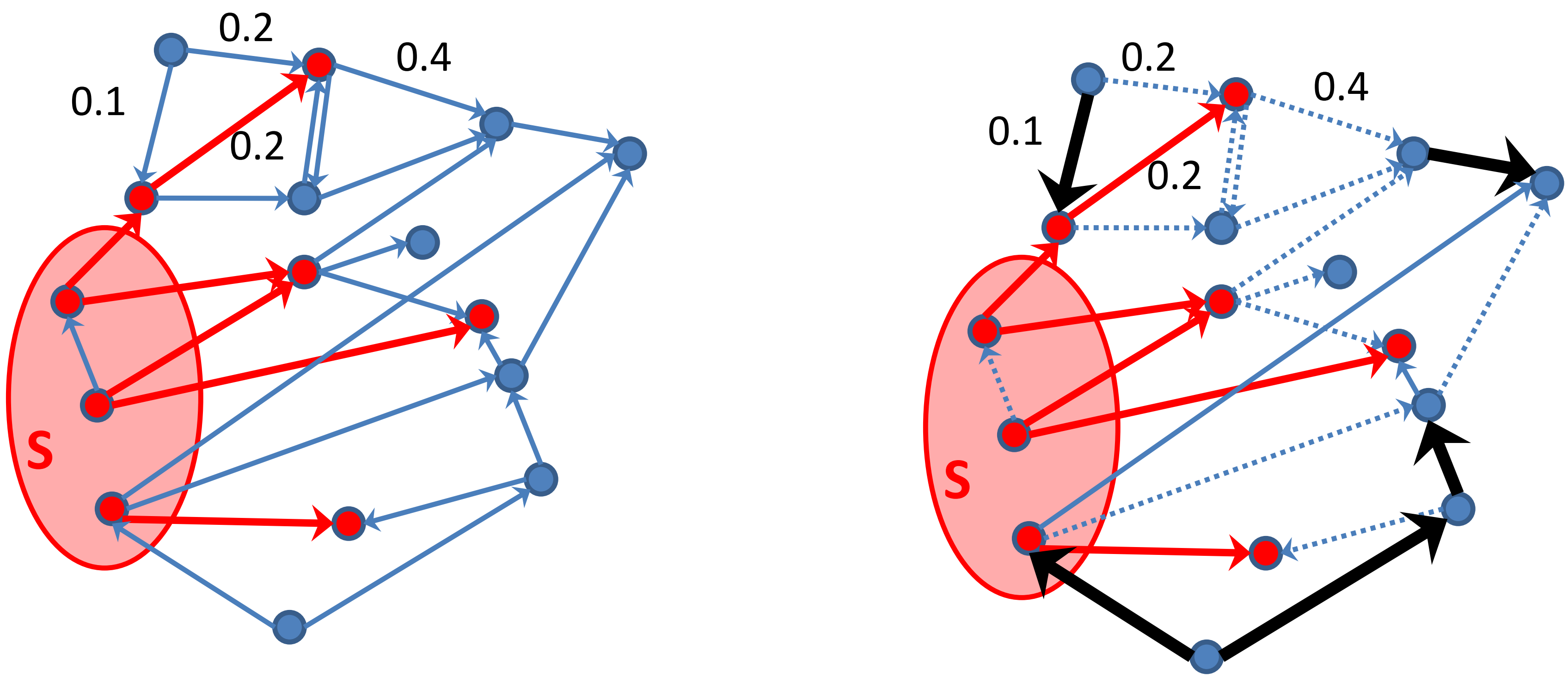
Assume that we flip the coins for the edges in the beginning



Given an initial set S look at the points reachable from S in the modified graph



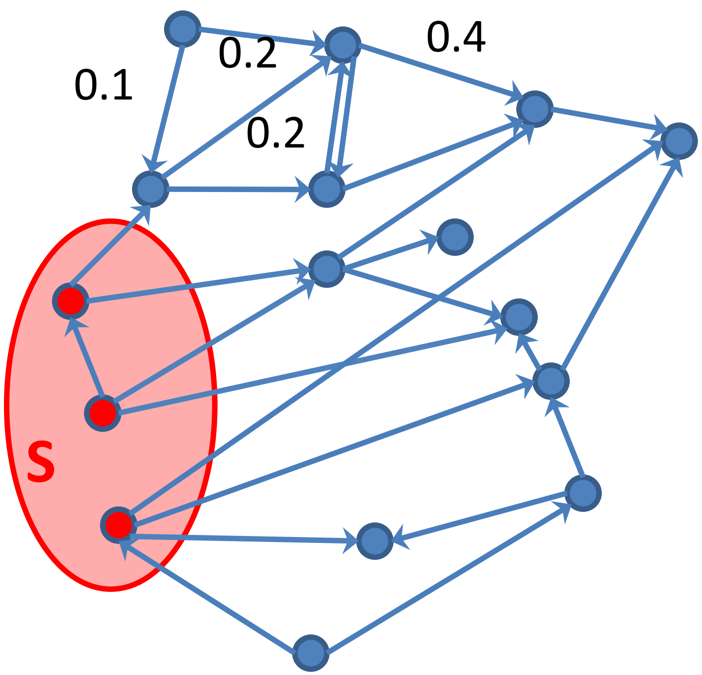
Another view of the cascading model



The cascading model and the new model give the same set of points in the end

But we already shown that g(S) is submodular

Back to f(S)



- For a fixed ω we showed that the function g(S) = f(S, ω) is submodular
- But we want to show that f(S) is submodular
- We have: $$f(S) = \sum_{\omega} \Pr(\omega) \cdot f(S, \omega)$$

- **Theorem.** A nonnegative linear combination of submodular functions is submodular

- We are DONE