Lecture 17: Influence (3/4)

How do influence spread?
(i.e., how innovations get adopted?)

COMS 4995-1: Introduction to Social Networks
Part II: Cascade
Tuesday, November 15th

Outline

* Life under the influence

* 1978: The global view, ‘Network effect’
* 2000: The local view, ‘Neighbors effect’
* 2003: The algorithmic view, ‘Exploiting Influence’

Outline

* Life under the influence

* 1978: The global view, ‘Network effect’
* 2000: The local view, ‘Neighbors effect’
* 2003: The algorithmic view, ‘Exploiting Influence’

A general algorithmic problem

* How to find the best initial seeding set S_0?
  - Maximizing the total spread, with a fixed size
* A more general model of neighbor influence
  - Assumes threhold t_v uniform in [0;1] and v becomes active as soon as $t_v \leq g_v(X)$
  - as u becomes active, activates neighbor v with prob. $p_v(u,X)$, where X = {nodes in N(v) previously active}
  - Special cases: Granovetter, Morris, Independent
  - If p order independent, the two models equivalent

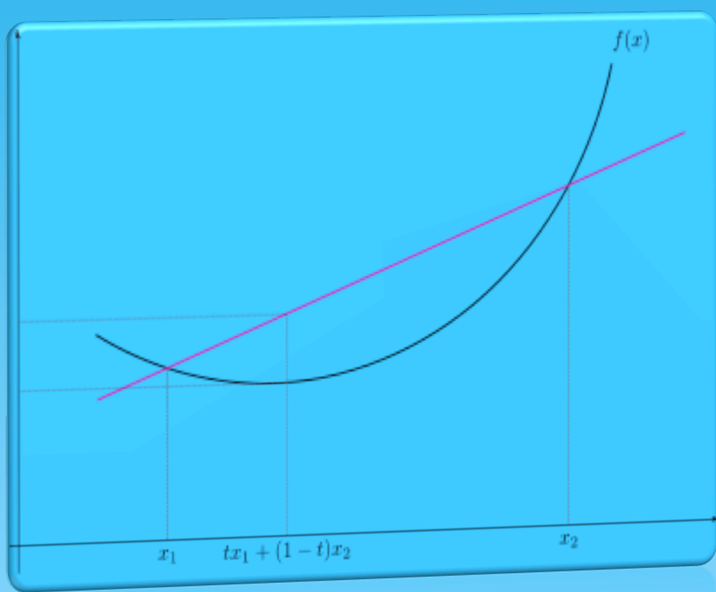
Critical Mass vs. Dim. Return

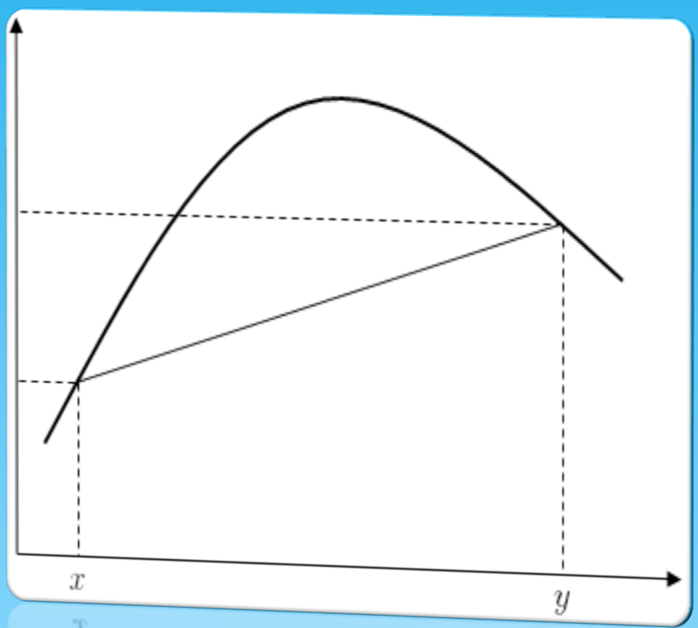
$p_v(u,X)$ "increases" with X

– Infection gets easier

$p_v(u,X)$ "decreases" with X

– infection gets harder





Maximizing spread of influence

- Let us denote the objective function by f(S)
  - f(S) = E[# of active nodes] at the end of the process
  - Assuming process starts with S
- Thm: Computing max{f(S) | |S|=k} is NP hard
  - Even computing a $n^{1-\varepsilon}$ approximation is hard (ε>0)
  - Proof: If g_v shows "critical mass" then relates to a set covering problem
  - Can we avoid "critical mass" and compute it?

CS@CU

Maximizing spread of influence

* Thm: Whenever p_v show diminishing return
  - There exists a simple polynomial algorithm computing S such that $f(S) \geq (1-1/e)f(S^*)$ where S^* is the optimal subset of size k
  - Algorithm follows greedy "one node at a time" rule
    Do k times: $S \leftarrow S \cup \text{argmax}_v \{ f(S \cup \{v\}) - f(S) \}$

Maximizing the spread of influence through a social network,
D. Kempe, J. Kleinberg, E. Tardos, ACM KDD (2003)

Proof

* Three steps:
  1. Show that the result holds if f is submodular, i.e. $S \subseteq T$ Implies $f(S \cup \{v\}) - f(S) \geq f(T \cup \{v\}) - f(T),$
  2. Show f is submodular under this condition on p_v
  3. Finally, prove that each step is polynomial more involved (will be admitted here)

Step 1

f submodular, ≥0

* Thm: If f non-negative, non-decreasing, submodular
  – Then greedy algorithm provides (1-1/e) approximation of maximizing f(S) subject to |S|=k.

  – Proof: First,

$$
\begin{aligned}
f(S_{i+1}) &\geq f(S_i) + \frac{1}{k} \cdot (f(T) - f(S_i)) \\
&= (1 - \frac{1}{k}) f(S_i) + \frac{1}{k} \cdot f(T)
\end{aligned}
$$

which implies by recurrence, $f(S_i) \geq (1 - (1 - \frac{1}{k})^i) \cdot f(T)$

An analysis of approximations for maximizing submodular set functions, G Nemhauser, L Wolsey, M Fisher, Math. Prog. (1978)

CS@CU

Proof

T = optimal $\{v_1, \dots, v_k\}$ $f(T)$

$S = \{v_1, v_2, \dots, v_k\}$ $S_i = \{v_1, \dots, v_i\}$ $1 \le i \le k$

sufficient to show with

$$f(S_{i+1}) - f(S_i) \geq \frac{1}{k}\left(f(T) - f(S_i)\right)$$

$\geq \frac{1}{k}\left(f(T \cup S_i) - f(S_i)\right)$

$T \cup S_i$ $\{v_1, \dots, v_i, u_1, \dots, u_k\}$ $f(T) - f(S_i)$

S_{i+1} $\{v_1, \dots, v_i, v_{i+1}\}$ $k\left(f(S_{i+1}) - f(S_i)\right) \geq f(T \cup S_i) - f(S_i)$

$$f(T) - f(S_{i+1}) = f(T) - f(S_i) - \left(f(S_{i+1}) - f(S_i)\right)$$

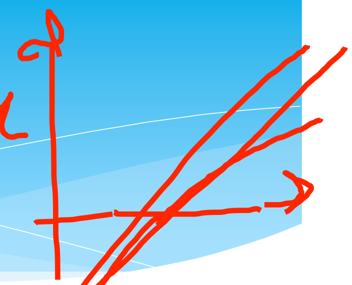
$$\leq \left(1 - \frac{1}{k}\right)\left(f(T) - f(S_i)\right)$$

CS@CU

UPCOMING

$f(\emptyset) = 0$

$\ln(1+x) \le x$



* ~~Follow up of the proof in the influence context~~

$$f(T) - f(S_{i+1}) \le \left(1 - \frac{1}{k}\right)\left(f(T) - f(S_i)\right)$$

$S_0 = \emptyset$

$\left(1 - \frac{1}{2}\right)^2 = \frac{1}{4}$

$$f(T) - f(S_k) \le \left(1 - \frac{1}{k}\right)^k \left(f(T) - f(\emptyset)\right)$$

$$\le \frac{1}{e}\left(f(T) - f(\emptyset)\right)$$

$$\left(1 - \frac{1}{k}\right)^k = \exp\left(k \ln\left(1 - \frac{1}{k}\right)\right) \le \exp\left(k\left(-\frac{1}{k}\right)\right) \le \frac{1}{e}$$

$$f(S_k) \ge \left(1 - \frac{1}{e}\right) f(T)$$

0.63%