Analysis of Social Information Networks

Thursday February 17th, Lecture 5: Influence (2)

Outline

* 2003: The algorithmic view, 'Exploiting Influence'
* 2010: effect of graph topology
  - Zoom on some recent results

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

CS@CU

Maximizing spread of influence

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

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

CS@CU

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)

CS@CU

Step 1

* 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

Step 2

* Key idea:
  – Propagation on edges (u,v) in E are event chosen independently.
  – It is equivalent to study a network where these events are decided in advance (i.e., conditioning).
* Assuming that edges in $E' \subseteq E$ propagates
  – f(S | E' propagates) = size of an union indexed by S
  – f is a sum of submodular function, proving the result

Step 2: Independent Cascade

* Key idea:
  – It is equivalent to study a network where these events are decided in advance (i.e., conditioning).
* Example 1: (Indep. Cascades: $p_v(u,X) = p_v(u)$)
* Assuming that edges in $E' \subseteq E$ propagates
  – f(S | E' propagates) = size of an union indexed by S
  – f is a sum of submodular function, proving the result

Step 2: Linear Threshold

* Key idea:
  - It is equivalent to study a network where these events are decided in advance (i.e., conditioning).
* Example 2: (Linear threshold: $g_v(X)=\sum_{u\in X} p_{uv}$)
* A random graph: each node v at most one in-edge
  - With prob. p_{uv} this edges is (u,v)
  - With prob. $1-\sum_{u\in V} p_{uv}$ v has no incoming edge
* Key observation: BFS from S on this graph and influence dynamics are statistically equivalent

CS@CU

Step 3 (for your information)

* We need to show that finding the best node v to add to S can be done with polynomial steps
  - Brute force method: (1) simulate the infection m times, then (2) use empirical average to choose v
  - Concentration result: with probability $1-\delta$, this yields a $(1-\varepsilon)$ approximation of v using $m=(n/\varepsilon)^2 \ln(2/\delta)$
  - Adapts approximation to show that this gives a $(1-1/e-\varepsilon)$ approximation

CS@CU

Summary

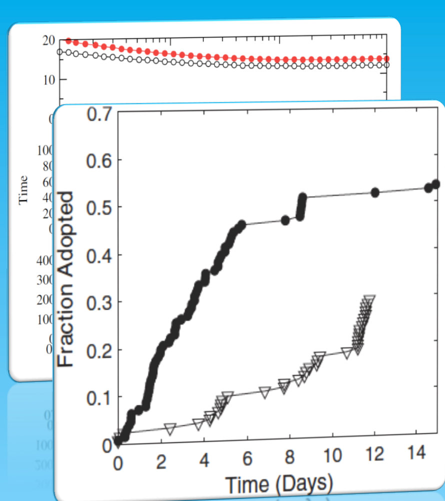
* Influence is prevalent:
  - Usually impacted by topology (cluster density) and local dynamics (critical mass vs. diminishing return)

* Generalization
  - Same result with renewed decision at each step.
  - Same result for the "linear threshold"

    $$g_v(X) = \sum_{u \in X} w_{uv}$$

  - Same result for any submodular function p and g_v

Impact of topology



- Test using small world rewired networks
- Non-monotonic behavior w.r.t. rewiring proba. P
- Empirical validation with real human subject

The Spread of Behavior in an Online Social Network Experiment, D Centola, Science (2010)

Cascade dynamics of complex propagation,
D Centola, V Eguíluz, M Macy, Phys. A (2007)

Impact of graph topology

* Intricate problems, results are recent!
* Random graph with arbitrary degree topology
  - Minimal size of infectious set has complex behavior
* Speed of propagation in
  - locally connected spreads faster (weak ties are weak)
  - High degree and expansion slow down propagation!

Diffusion and cascading behavior in random networks,
M. Lelarge, Preprint (2010)

The spread of innovations in social networks,
A Montanari, A Saberi, PNAS (2010)