Lecture 18: Influence (4/4)

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

COMS 4995-1: Introduction to Social Networks
Part II: Cascade
Thursday, November 17th
* 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 threshold $t_v$ 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 $X = \{\text{nodes in } N(v) \text{ previously active}\}$
  - 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
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: $S \leftarrow S \cup \arg\max_v \{ f(S \cup \{v\}) - f(S) \}$

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.
     \[ S \subseteq T \text{ Implies } f(S \cup \{v\}) - f(S) \geq f(T \cup \{v\}) - f(T), \]
  2. Show $f$ is submodular under this condition on $p_v$ (we will prove it for $p_v$ constant)
  3. Finally, prove that each step is polynomial more involved (will be admitted here)
What is an example of a submodular function?

- \( f(S) = |S| \)
- \( f(S) = |S|^2, f(S) = \ln(|S|) \)
- \( f(S) \) indicator of \( x \): \( f(S) = 1 \) if \( x \) is in \( S \), 0 otherwise

- What about \( f(S) = 0 \) if \( x \) is in \( S \), 0 otherwise
A union function: Let $C_1, C_2, \ldots, C_n$ subsets of $V$

- For any $S$ subset of $\{1,2,\ldots,n\}$, let $f(S) = \left| \bigcup_{i \in S} C_i \right|$

$$f(S) \cup \delta(S) = f(T) \cup \delta(T) \quad \text{for } S \subseteq T$$

$$\left| \bigcup_{i \in S} C_i \right| - \left| \bigcup_{j \in S} C_j \right| \geq \left| C_i \setminus \bigcup_{j \in T} C_j \right|$$

* What about a positive linear combinations?

$$f + g = \sum t_i f_i$$
Key idea:
- Propagation on edges \((u,v)\) in \(E\) are event chosen independently.
- Conditioning: It is equivalent to study a network where these events are decided in advance.

\[ \sum_{E' \subseteq E} P(E') f(S \mid E') \]

Assuming that edges in \(E' \subseteq E\) propagates
- \(f(S \mid E' \text{ propagates}) = \text{size of an union indexed by } S\)
- \(f\) is a sum of submodular function, proving the result
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
- $f(S \mid E' \text{ propagates}) = \text{size of an union indexed by } S$
- $f$ is a sum of submodular function, proving the result

$$f(S) = \sum_{E \subseteq E} \left| \bigcup_{s \in S} E_s \right|$$
Example

What are the nodes that I cannot if I start from $S_0$?

\[ \{ \text{nodes that I cannot} \mid E \text{ is successful edge} \} = \bigcup_{s \in S_0} \mathcal{E}_s \]

\[ f(s_0) = \left| \bigcup_{s \in S_0} \mathcal{E}_s \right| \]
Key idea:
- It is equivalent to study a network where these events are decided in advance (i.e., conditioning).

Example 2: \( g_v(X) = \sum_{u \in X} p_{uv} \)

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
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,
  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:**
  This gives a $(1-1/e-\varepsilon)$ approximation
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 \( g_v \)
Critical Mass vs. Dimin. Return

- Live journal
- Friendship + topics
- Conference
- Co-authorship

Diminishing return dominates, except for the first 2 nodes