# Diffusion in Networks

Thodoris Lykouris

National Technical University of Athens

June 20, 2013

# Contents

# Influence in Networks

Usually the underlying social network of two forms:

# Influence in Networks

Usually the underlying social network of two forms:

1. Amorphous population of individuals and effects in aggregate
   - Poll

# Influence in Networks

Usually the underlying social network of two forms:

1. Amorphous population of individuals and effects in aggregate
   - Poll
2. Network as a graph and individuals influenced by network neighbors
   - Discussions

# Influence in Networks

Usually the underlying social network of two forms:

1. Amorphous population of individuals and effects in aggregate
   - Poll
2. Network as a graph and individuals influenced by network neighbors
   - Discussions

Local effects instead of global:

1. Direct-benefit effects
   - compatibility of technologies for operating systems
   - less cost for cell phones

# Influence in Networks

Usually the underlying social network of two forms:

1. Amorphous population of individuals and effects in aggregate
   - Poll
2. Network as a graph and individuals influenced by network neighbors
   - Discussions

Local effects instead of global:

1. Direct-benefit effects
   - compatibility of technologies for operating systems
   - less cost for cell phones
2. Indirect effects like informational effects
   - their choices often provide indirect information of things they know

# A networked Coordination game

- Social network, modeled as graph:
  - nodes are the people
  - edges denote friendship between them
- Two products $A$ and $B$ that are competitive

For each edge:

- If both adopt $A$, each gets payoff $a > 0$
- If both adopt $B$, each gets payoff $b > 0$
- If they adopt different products, each gets payoff 0

# Cascading Behavior and Viral Marketing

- Suppose $p$ of your neighbors have adopted $A$ and $1 - p$ B.
- Expected gain $pda$ if you adopt $A$ and $(1 - p)db$ if you adopt $B$
- If $p \geq \frac{b}{a+b}(= q)$, you prefer to adopt $A$.
- If a node has more than $q$ neighbors who have adopted $A$, he alters to $A$

# Cascading Behavior and Viral Marketing

- Suppose $p$ of your neighbors have adopted $A$ and $1-p$ B.
- Expected gain $pda$ if you adopt $A$ and $(1-p)db$ if you adopt $B$
- If $p \geq \frac{b}{a+b}(=q)$, you prefer to adopt $A$.
- If a node has more than $q$ neighbors who have adopted $A$, he alters to $A$
- Strategies for $A$ to ameliorate its situation:
  1. Increasing $a$ (making an existing innovation more attractive can generally increase its reach)
  2. Convincing a small number of key people to switch from $B$ to $A$
     - A small number of initial adopters essentially start a long fuse that eventually spreads globally
     - Great question: How to choose these key people?

# Cascades and Clusters

Suppose now everybody has adopted $B$ and you switch some nodes to $A$

- If a node has more than $q$ neighbors who have adopted $A$, he alters to $A$

# Cascades and Clusters

Suppose now everybody has adopted $B$ and you switch some nodes to $A$

- If a node has more than $q$ neighbors who have adopted $A$, he alters to $A$
- When does it stop?

# Cascades and Clusters

Suppose now everybody has adopted $B$ and you switch some nodes to $A$

- If a node has more than $q$ neighbors who have adopted $A$, he alters to $A$
- When does it stop?
  - When nobody has more than $q$ neighbors in $A$
- When will $A$ have a *complete cascade*?

# Cascades and Clusters

Suppose now everybody has adopted $B$ and you switch some nodes to $A$

- If a node has more than $q$ neighbors who have adopted $A$, he alters to $A$
- When does it stop?
  - When nobody has more than $q$ neighbors in $A$
- When will $A$ have a *complete cascade*?
  - When there exists a $B$-cluster with density $1 - q$
  - $1 - q$ of each cluster member's neighbors are in the cluster
  - Hence, dense clusters are the only obstacles to cascades

# Thresholds

Thresholds model the difference between learning for a product and adopting it

- Possibly heterogeneous
  - each node $v$ different $(a_v, b_v)$
  - $q_v = \frac{b_v}{a_v + b_v}$
  - blocking clusters: set of nodes where every node has $1 - q_v$ neighbors in the set
- Bridges between different connecting components not that useful
- More interested in the extent to which a node has access to easily infulenceable nodes

# Nature of networks

Generally weak ties vs strong ties?

- Access to more information via weak ties
- Common knowledge via strong ties (very important for a protest, for instance)

# From epidemic diseases to the diffusion of ideas

Contact networks model the ways in which a disease can spread:

- People are nodes
- Interactions are edges
- Danger of interaction is weight
  - when two people meet and one is infected, given probability for the disease to spread

Clear connection between diseases and diffusion of ideas in social networks

# The SIR Epidemic Model

Each node in three potential stages

- *Susceptible*: Before the node has caught the disease, it is susceptible to infection from its neighbors
- *Infectious*: Once the node has caught the disease, it has some probability to infect each of its susceptible neighbors
    - Probability $p$ to influence (extension: different for each edge)
    - Possibly $t_I$ steps
- *Removed*: After it has experienced full infectious period, it is removed from consideration

Variation: SIS Epidemic Model

# Contents

# The Problem

Given a social network, find an initial influential set of $k$ nodes that maximize the (expected) number of eventual adopters

- Many models to address this problem.
- Difference on
  - the way the influence is exerted
    - usually some function $f$ of the influence of its neighbors
  - the way a node decides to adopt a product
    - Cascade Models: subclass of epidemic models
    - Threshold Models

# Useful Properties

Supposing $\sigma(A)$ (expected) number of eventual adopters for influential set $A$

1. Monotonicity: if $A \subseteq B$ then $\sigma(A) \leq \sigma(B)$
   - Addition of a node in the influential set does not decrease (expected) number of eventual adopters
2. Submodularity: if $A \subseteq B$ then $\sigma(A \cup x) - \sigma(A) \geq \sigma(B \cup x) - \sigma(B)$
   - Marginal gain of a new node less for bigger influential sets

- Greedy algorithm (pick the node that maximizes the marginal gain of eventual adopters) has $(1 - \frac{1}{e})$-approx for monotone and submodular functions (Nemhauser)
- Generally difficult to find it exactly so emphasis on models with these nice properties

# Linear Threshold Model

- Each edge $e$ some weight $b_e$
- Each node $u$ some threshold $\theta_u$ uniformally at random from $[0, 1]$
- If $\sum_{w \in AN(u)} b_{w,u} \geq \theta_u$, $u$ adopts the product
- NP-hard. Reduction from Vertex Cover.
- Monotone and submodular. Idea of the proof:
  - One edge $e$ is considered *live* w.p. $b_e$
  - All the other blocked
  - All edges blocked w.p. $q - \sum w \in N(u) b_{w,u}$
  - The set of all reachable paths is the eventual adopting nodes
  - Stochastically equivalent to the Linear Threshold Model

# Independent Cascade Model

- When a node $u$ becomes active, it has a single chance to influence each neighbor $w$ w.p. $p_{u,w}$
- Both this and the previous model are *progressive* models (no colored node can become uncolored later)

- NP-hard.
  - Reduction from Set Cover.
  - 2 layers (sets and unions)
  - $p_{e,s} = 1$ iff $e \in S$
  - Can we have eventual influence of $n + k$?
- Monotone and submodular. Idea of the proof:
  - A biased coin $p_e$ for each edge to decide if it is *live* or not
  - A non-negative linear combination of submodular functions is itself submodular

# Decreasing Cascade Model

- Sometimes the first influencers tend to have more impact on us than the latter
- $p_v(u, S) \geq p_v(u, T)$ for $S \subseteq T$
- Order-independence necessary

- NP-hard. Reduction from Vertex Cover
- Monotone and submodular

# Deterministic Thresholds

- It is NP-hard to approximate anything with $\Omega(n^{1-\epsilon})$.
- That's why, randomized thresholds
- Same hardness even if the thresholds are majority thresholds
- Reduction from Set Cover, similar to Independent Cascade Model's

# Computation of Utility

- Exact computation of the utility is $\#P$-hard.
- Reduction from Counting simple paths in a graph.
- FPRAS for approximating it as well as we wish
- Hence the approximation algorithm gives $(1 - \frac{1}{e} - \epsilon)$-approximation for the best response

# Contents

# Competitive Cascade Model

- Competitive Influence Maximization Problem:
- when $m$ firms try to propagate their influence in a competitive way
- Simple extension of Independent Cascade Model with multiple firms
- Expected number of eventual adopters by last firm monotone and submodular
- Approximation algorithm for last firm's best response

# Weight-Proportional Competitive Linear Threshold Model

- Progressive process, Red and Blue firm
- Each edge $e$ has a weight $w_e$
- Each node $\theta_u$ chooses a threshold uniformally at random
- Node adopts if total influence exerted is greate or equal to its threshold
- It chooses the color, weight-proportionally to the influence it takes from each

# Weight-Proportional Competitive Linear Threshold Model

- Progressive process, Red and Blue firm
- Each edge $e$ has a weight $w_e$
- Each node $\theta_u$ chooses a threshold uniformally at random
- Node adopts if total influence exerted is greate or equal to its threshold
- It chooses the color, weight-proportionally to the influence it takes from each

- Generally not monotone nor submodular
- The reason is that the first influencers have double effect
  - When they exert influence at first, they take the node if they overcome the threshold
  - Unless this happens, they have second chances every time more influence is exerted to the node, due to the weight-proportional tie breaking rule

# Separated-Threshold Model for Competing Technologies

- Here firms influences aren't aggregated
- Each has its own influence and each node has two thresholds, one for each
- When $\sum_{u \in I_A t-1} w_{u,w}^A \geq \theta_w^A$ then $w$ adopts A's product.
- If both thresholds are overpassed, some tie-breaking rule, i.e. simple coinflip

- As expected, monotone
- Still not submodular: tie-breaking rule is to be blamed but unless extreme assumptions, no tie-breaking rule makes it submodular

# Switching-selection Model: A general framework

- General framework of models
- There is a switching function, which decides whether a node will adopt some product, according to the total influence exerted to it
- And a selection function, which decides which product it will adopt, given the influences exerted by each and given the fact that the node has decided to buy
- Includes many models as special cases
- Offers an intuition about how domineering the initial difference in budgets is (according to the nature of the functions)
  - Budget Multiplier
- Offers an intuition about how bad the equilibria can be
  - Price of Anarchy