Diffusion of Innovation and Influence Maximization

Leonid E. Zhukov

School of Data Analysis and Artificial Intelligence
Department of Computer Science
National Research University Higher School of Economics

Network Science

Lecture outline

1. Diffusion of innovation
2. Influence propagation models
   - Independent cascade model
   - Linear threshold model
3. Influence maximization problem
   - Submodular function optimization

Diffusion process

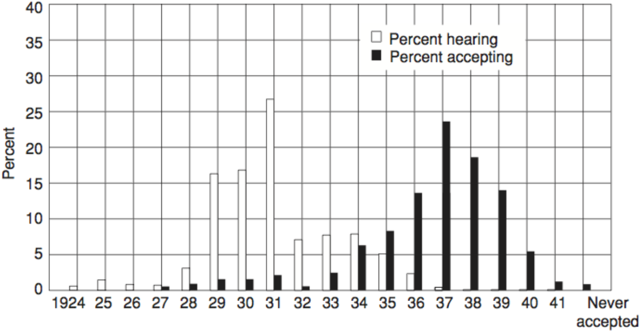
Propagation process:

- Viral propagation:
  - virus and infection
  - rumors, news
  - information
- Decision based models:
  - adoption of innovation
  - joining politcal protest
  - purchase decision

Local individual decision rules will lead to very different global results.
"microscopic" changes $\rightarrow$ "macroscopic" results

Diffusion of innovation

Ryan-Gross study of hybrid seed corn delayed adoption - diffusion of innovation

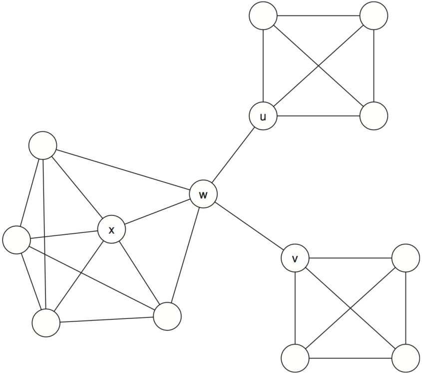


Information effect vs adopting of innnovation

Ryan and Gross, 1943

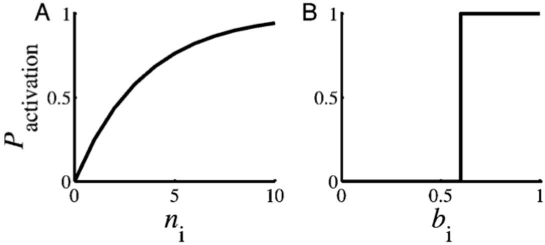
Diffusion of innovaton

Information (awareness) vs adoption (decision) spreading



Influence response

Influence response: diminishing returns and threshold (critical mass)



$$P(n) = 1 - (1 - p)^n \qquad P(b) = \delta(b > b_0)$$

Influence propagation models

Two basic models:

- Linear Threshold Model (critical mass)
- Independent Cascade Model (diminishing returns)

Network coordination game

Local interaction game: Let u and v are players, and A and b are possible strategies

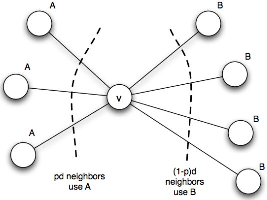
Payoffs

- if u and v both adopt behavior A, each get payoff $a > 0$
- if u and v both adopt behavior B, each get payoff $b > 0$
- if u and v adopt opposite behavior, each get payoff 0

		w	
		A	B
v	A	a, a	$0, 0$
	B	$0, 0$	b, b

Threshold model

Network coordination game, direct-benefit effect



		w: A	w: B
v	A	a, a	$0, 0$
	B	$0, 0$	b, b

Node v to make decision A or B, p - portion of type A neighbors to accept A:

$$a \cdot p \cdot d > b \cdot (1 - p) \cdot d$$

$$p \geq b/(a + b)$$

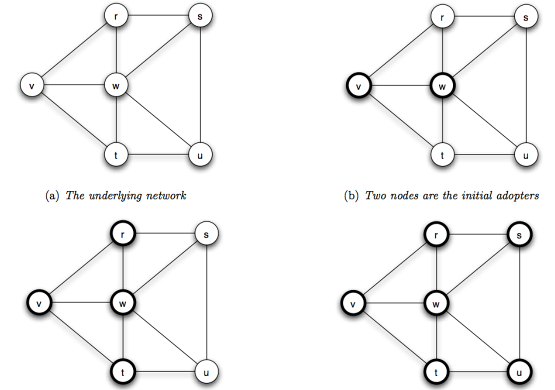
Threshold:

$$q = \frac{b}{a + b}$$

Accept new behavior A when $p \geq q$

Cascades

Cascade - sequence of changes of behavior, "chain reaction"



(a) *The underlying network*

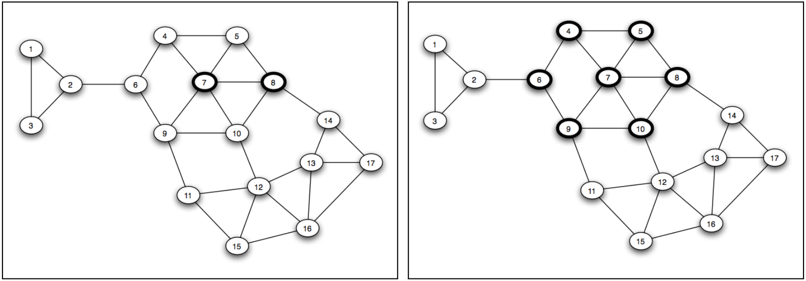
(b) *Two nodes are the initial adopters*

(c) *After one step, two more nodes have adopted*

(d) *After a second step, everyone has adopted*

Let $a = 3$, $b = 2$, threshold $q = 2/(2+3) = 2/5$

Cascade propagation



- Let $a = 3$, $b = 2$, threshold $q = 2/(2+3) = 2/5$
- Start from nodes 7,8: $1/3 < 2/5 < 1/2 < 2/3$
- Cascade size - number of nodes that changed the behavior
- Complete cascade when every node changes the behavior

Linear threshold model

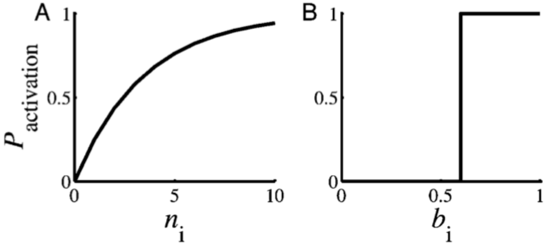
- Influence comes only from NN $N(i)$ nodes, w_{ij} influence $i \to j$
- Require $\sum_{j \in N(i)} w_{ji} \leq 1$
- Each node has a random acceptance threshold from $\theta_i \in [0, 1]$
- Activation: fraction of active nodes exceeds threshold

$$\sum_{active\ j \in N(i)} w_{ji} > \theta_i$$

- Initial set of active nodes A_o, iterative process with discrete time steps
- Progressive process, only nonactive $\to$ *active*

Influence response

Influence response: diminishing returns and threshold (critical mass)



$$P(n) = 1 - (1 - p)^n \qquad P(b) = \delta(b > b_0)$$

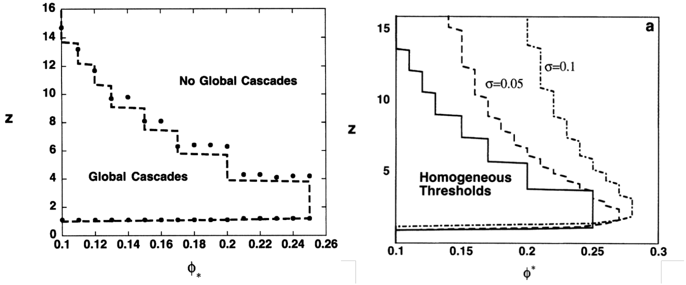
Independent cascade model

- Initial set of active nodes S_0
- Discrete time steps
- On every step an active node v can activate connected neighbor w with a probability $p_{v,w}$ (single chance)
- If v succeeds, w becomes active on the next time step
- Process runs until no more activations possible

If $p_{v,w} = p$ it is a particular type of SIR model, a node stays infected for only one step

Cascades in random networks

- Global cascade (sufficiently large)
- Triggered by single node (or small set)
- Random graphs ER p_k
- Threshold distribution ϕ

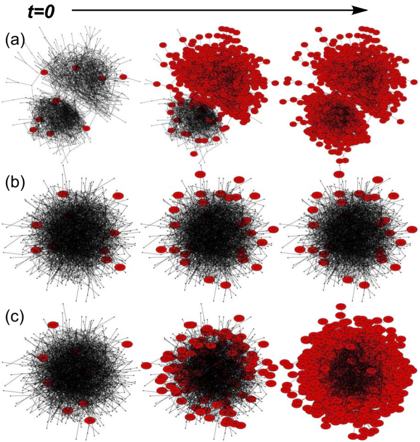


Cascade window: a) homogenous threshold b) normal threshold distribution

D. Watts, 2002

Cascades in random networks

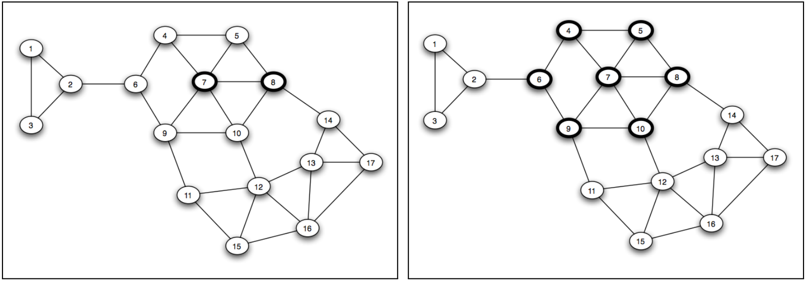
multiple seed nodes



(a) Empirical network; (b), (c) - randomized network

P. Singh, 2013

Influence maximization problem



- Initial set of active nodes A_o
- Cascade size $\sigma(A_o)$ - expected number of active nodes when propagation stops
- Find k-set of nodes A_o that produces maximal cascade $\sigma(A_o)$
- k-set of "maximum influence" nodes
- NP-hard

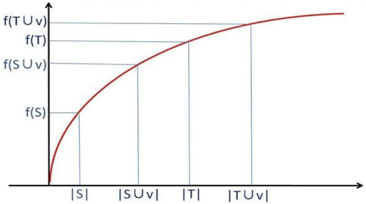
D. Kempe, J. Kleinberg, E. Tardos, 2003, 2005

Submodular functions

- Set function f is submodular, if for sets S, T and $S \subseteq T$, $\forall v \notin T$

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

- Function of diminishing returns ("concave property")
- Function f is monotone, $f(S \cup \{v\}) \geq f(S)$



Submodular functions

Theorem

Let F be a monotone submodular function and
let S^* be the k-element set achieving maximal f.
Let S be a k-element set obtained by repeatedly, for k-iterations, including an element producing the largest marginal increase in f.

$$f(S) \geq (1 - \frac{1}{e}) f(S^*)$$

Nemhauser, Wolsey, and Fisher, 1978

Influence maximization

- Cascade size $\sigma(S)$ is submodular function (D. Kempe, J. Kleinberg, E. Tardos, 1993)

$$\sigma(S) \geq (1 - \frac{1}{e})\sigma(S^*)$$

- Greedy algorithm for maximum influence set finds a set S such that its influence set $\sigma(S)$ is within $1/e = 0.367$ from the optimal (maximal) set $\sigma(S^*)$, $\sigma(S) \geq 0.629\sigma(S^*)$

Influence maximization

Approximation algorithm

Algorithm: Greedy optimization

Input: Graph $G(V, E)$, k

Output: Maxumum influence set S

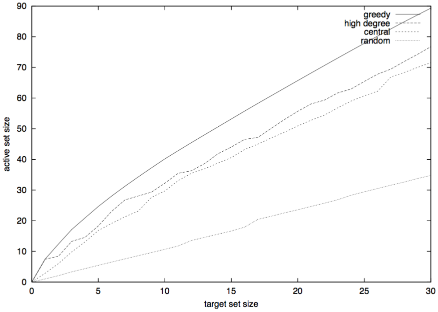
Set $S \leftarrow 0$

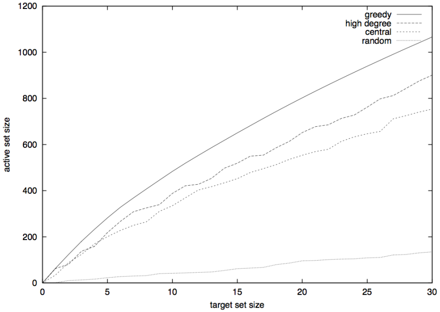
for $i = 1 : k$ **do**

  select $v = \arg\max_{u \in V \setminus S}(\sigma(S \cup \{u\}) - \sigma(S))$

  $S \leftarrow S \cup \{v\}$

Experimental results





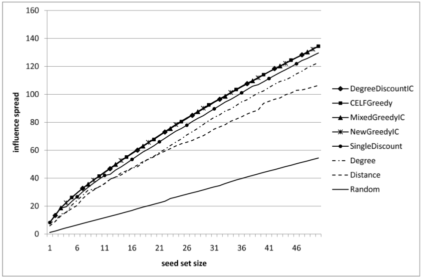
Independent cascade model

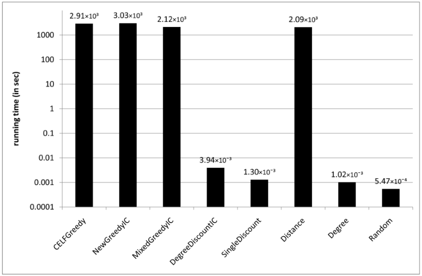
Linear threshold model

network: collaboration graph 10,000 nodes, 53,000 edges

D. Kempe, J. Kleinberg, E. Tardos, 2003

Computational considerations





Independent cascade model: influence spread and running time

W. Chen et.al, 2009

References

- Contagion, S. Morris, Review of Economic Studies, 67, p 57-78, 2000
- Maximizing the Spread of Influence through a Social Network, D. Kempe, J. Kleinberg, E. Tardos, 2003
- Influential Nodes in a Diffusion Model for Social Networks, D. Kempe, J. Kleinberg, E. Tardos, 2005
- Efficient Influence Maximization in Social Networks, W. Chen, Y. Wang, S, Yang, KDD 2009.
- A Simple Model of Global Cascades on Random Networks. D. Watts, 2002.