



NATIONAL RESEARCH
UNIVERSITY

Cascades in Networks. Influence maximization

Network Science

Lecture 11

Leonid Zhukov

lzhukov@hse.ru

www.leonidzhukov.net/hse/2022/networks

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

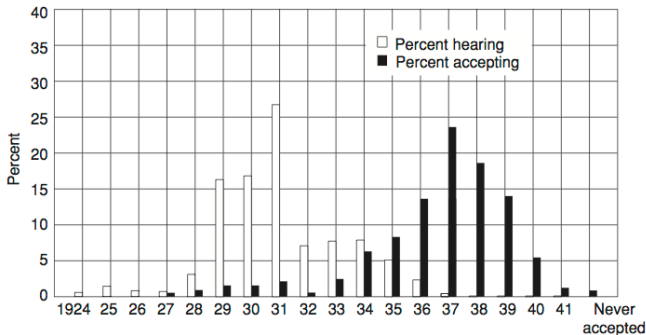
Propagation process:

- Viral propagation:
 - virus and infection
 - rumors, news
 - information
- Threshold (agent decision) models:
 - adoption of innovation
 - joining political protest
 - purchase decision
 - cascading failures

Local individual decision rules will lead to very different global results.

"microscopic" changes → "macroscopic" results

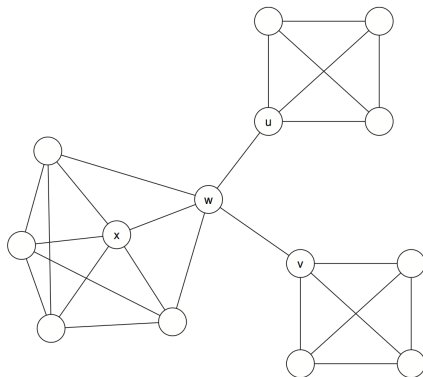
Ryan-Gross study of hybrid seed corn delayed adoption - diffusion of innovation



Information effect vs adopting of innnovation

Ryan and Gross, 1943

Information (awareness) vs adoption (decision) spreading



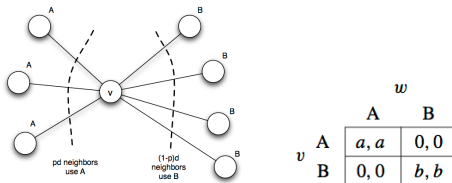
Local interaction game: Let u and v are players, and A and B are possible strategies

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

Network coordination game, direct-benefit effect



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

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

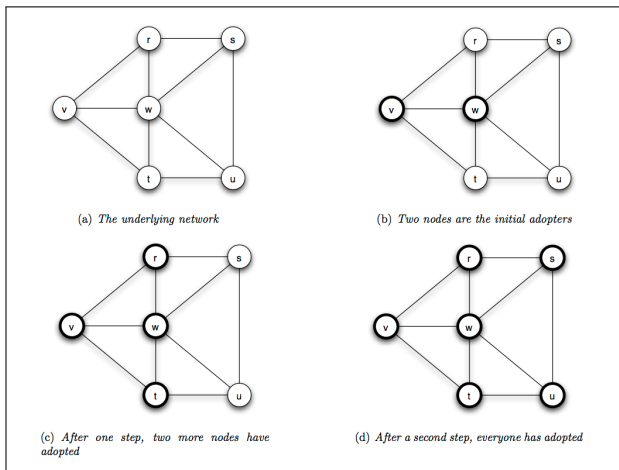
$$p \geq b/(a + b)$$

Threshold:

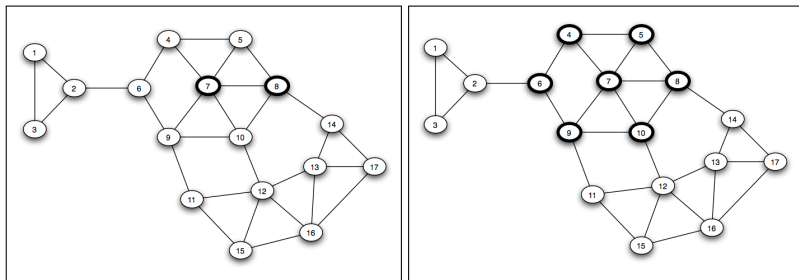
$$q = \frac{b}{a + b}$$

Accept new behavior A when $p \geq q$

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



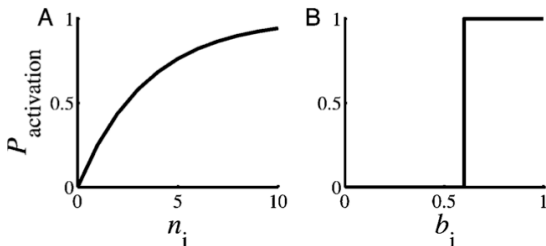
Let $a = 3$, $b = 2$, threshold $q = 2/(2 + 3) = 2/5$



- 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

Two models:

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



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

Influence response: diminishing returns and threshold

- 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

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

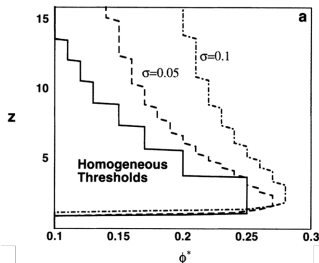
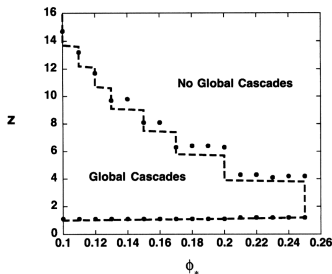
- Influence comes only from NN $N(i)$ nodes, w_{ji} influence $i \rightarrow j$
- Require $\sum_{j \in N(i)} w_{ji} \leq 1$
- Each node has a random acceptance threshold from $\theta_i \in [0, 1]$
- Activation: fraction of active nodes exceeds threshold

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

- Initial set of active nodes A_0 , iterative process with discrete time steps
- Progressive process, only nonactive $\rightarrow$ *active*

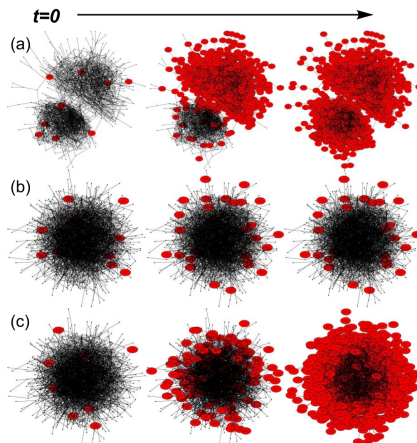
D. Kempe, J. Kleinberg, E. Tardos, 2003

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

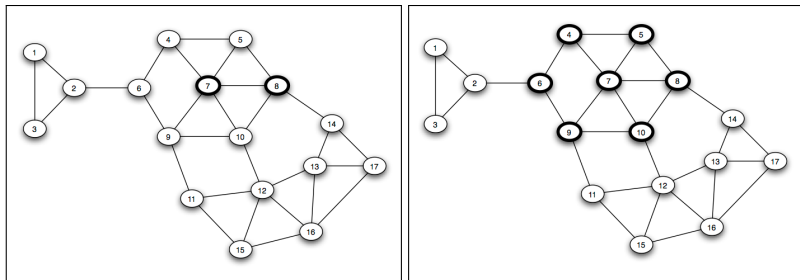


Cascade window: a) homogenous threshold b) normal threshold distribution

multiple seed nodes



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

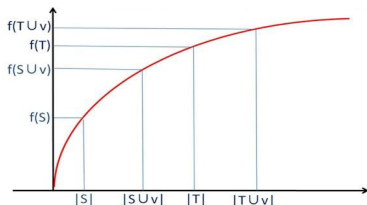


- Initial set of active nodes A_0
- Cascade size $\sigma(A_0)$ - expected number of active nodes when propagation stops
- Find k -set of nodes A_0 that produces maximal cascade $\sigma(A_0)$
- k -set of "maximum influence" nodes
- NP-hard

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



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 .*

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

Nemhauser, Wolsey, and Fisher, 1978

- 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 $\sigma(S^*)$, $\sigma(S) \geq 0.629\sigma(S^*)$

Approximation algorithm

Algorithm: Greedy optimization

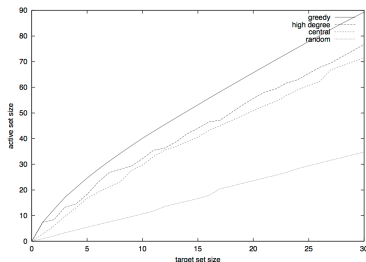
Input: Graph $G(V, E)$, k

Output: Maximum influence set S

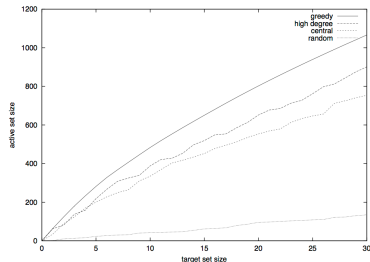
Set $S \leftarrow \emptyset$

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

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



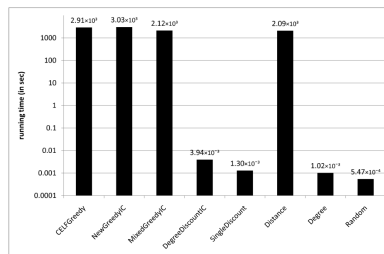
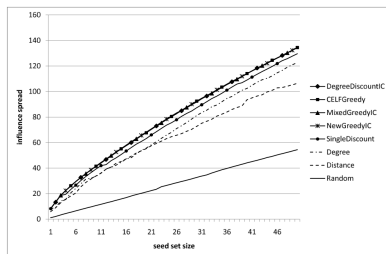
Independent cascade model



Linear threshold model

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

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



Independent cascade model: influence spread and running time

W. Chen et.al, 2009

- 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.