Social contagion

Leonid E. Zhukov

School of Applied Mathematics and Information Science
National Research University Higher School of Economics

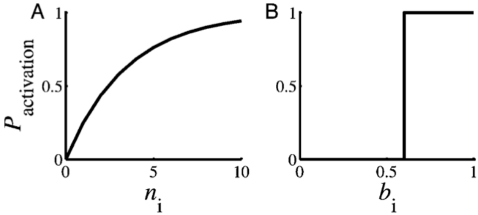
28.04.2014

Social contagion

- Diffusion of innovation - Threshold model:
  - Adoption of innovations
  - Joing a protest group
  - Making a purchase (?)
- Social diffusion - Independent cascade model:
  - News propagation
  - Retweets, mentions, email forwarding
  - Recommendations

Influence response

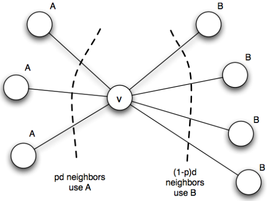
Influence response: diminishing returns and critical mass (threshold)



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

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:

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

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

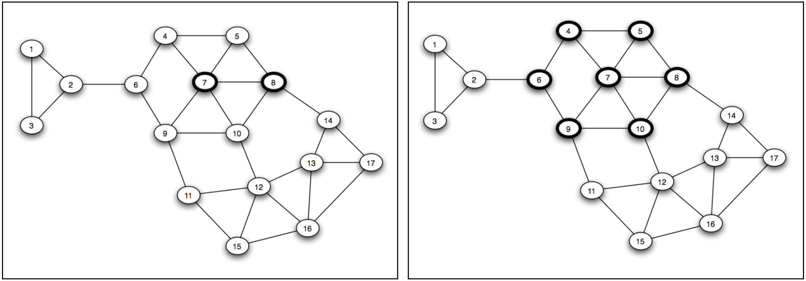
Thresholod:

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

Accept new behavior A when $p \geq q$

Cascade propagation

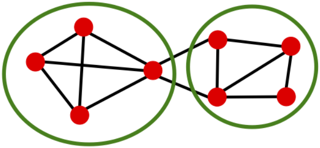
Cascade - sequence of changes of behavior, ''chain reaction''



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

Cascades and clusters

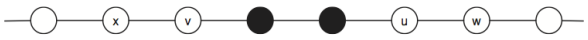
Group of nodes form a cluster of density ρ if every node in the set has at least fraction ρ of its neighbors in the set



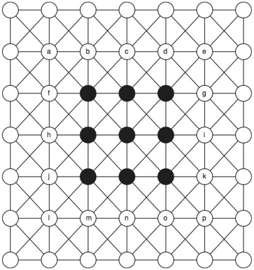
- If the network contains a claster with density $\rho > 1 - q$, then the cluster will block the propagation of cascade anf there will be no complete cascade
- For cascade to get into cluster $q \leq 1 - \rho$.
- Adoption threshold to get into the left cluster $q = 2/5$, the right cluster $q = 1/3$

Cascades capacity

- Cascade capacity - the largest threshold $q_\infty = \max b/(a+b)$ at which small set of initial adopters can cause a complete cascade?
- Finite set of initial nodes on an infinte network



$q_\infty = 1/2$

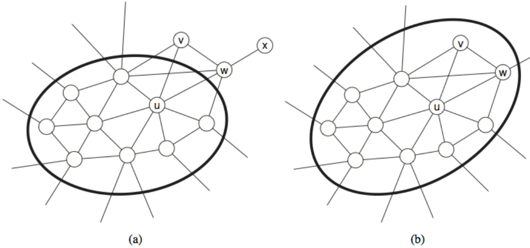


$q_\infty = 3/8$

- With smaller $q < q_\infty$ cascade spreads faster
- For $q < 1/2$, $a > b$; for $q < 3/8$, $a > 5/3b$

Cascade capacity

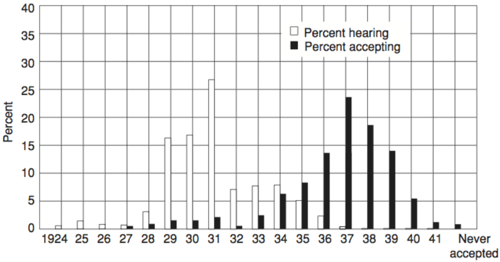
- The bigger the cascade capacity q_∞, the easier cascade spreads on network
- If $q_\infty > 1/2$ exists, then $a < b$ can cause a complete cascade, i.e. inferior technology could displace superior with a small set of initial adoptors
- Claim: There is no nework with cascade capacity $q_\infty > 1/2$



Shrinking interface: since $p > q_\infty > 1/2$, there are more connections inside interface then outside on every step, moving in decreases interface size

Ryan-Gross study

Ryan-Gross study of hybrid seed corn delayed adoption (after first exposure)



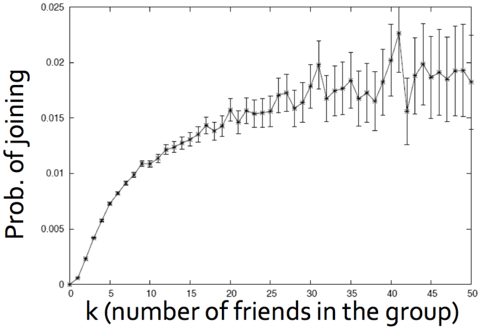
Ryan and Gross, 1943

Independent cascade model

- Initial set of active nodes S_0
- Discrete time steps
- On every step an active node v can activate connected neighbour 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

Live journal

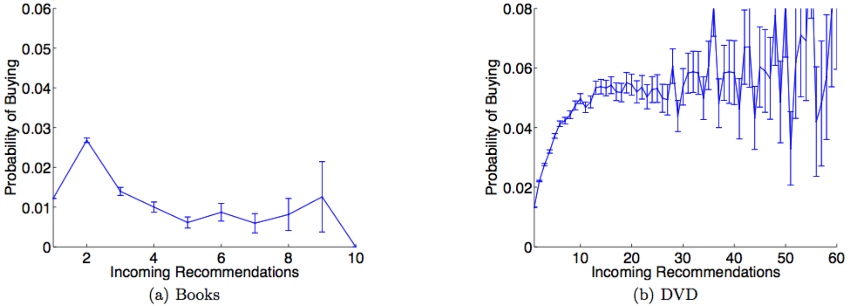
10 mln members, joining groups



Backstrom, 2006

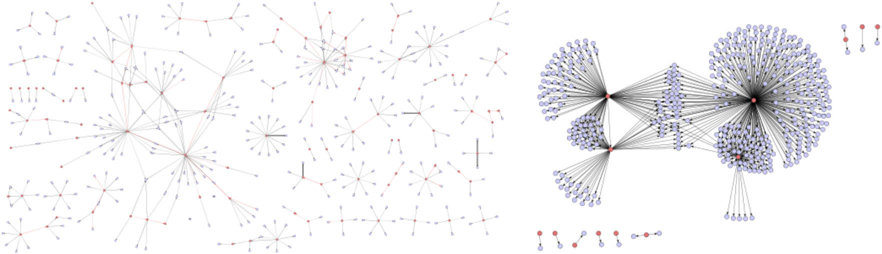
Online recommendation network

- Large on-line retailer. After purchase one could send recommendation to a friend, both receive discount/referral credit
- 15 mln recommendations, 3 mln users



Leskovec et.al, 2006

Online recommendation network

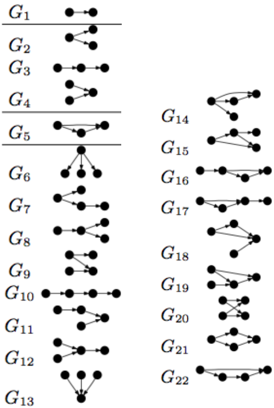


(a) Medical book (b) Japanese graphic novel

Leskovec et.al, 2006

Online recommendation network

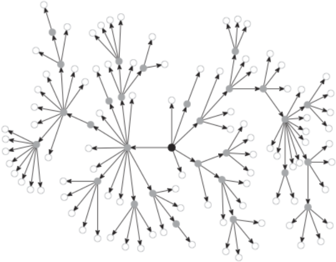
Most frequent subgraphs in cascade (network motifs)



Leskovec et.al, 2006

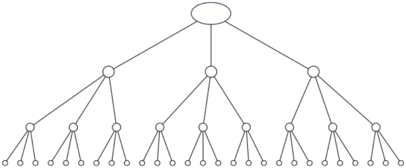
Online newsletter experiment

- Online email newsletter + reward for recommending it
- 7,000 seed, 30,000 total recipients, around 7,000 cascades, size form 2 nodes to 146 nodes, max depth 8 steps
- on average $k = 2.96$, infection probability $p = 0.00879$
- Suprisingly: no loops, triangles, closed paths

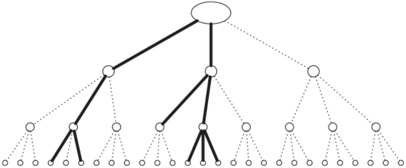


Iribarren et.al, 2009

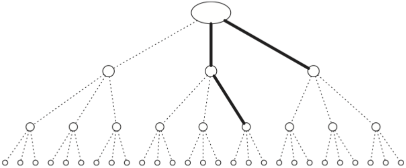
Cascade - branching process



(a)



(b)



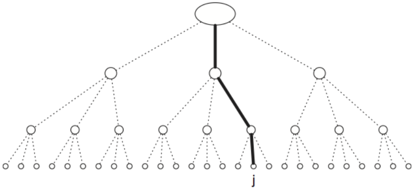
(c)

Branching process

Branching process:

- each node has k neighbours, transmits "infection" with probability p
- let q_n - probability that infection persits n steps (levels of the tree)
- pq_{n-1} - probability that spreads through first contact and then survives $n-1$ levels
- $(1-pq_{n-1})^k$ - probability that will not spread through any of the nodes

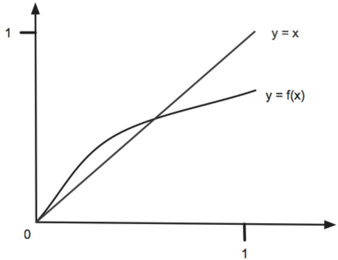
$$1 - q_n = (1 - pq_{n-1})^k$$



Branching process

- limiting probability $q^* = \lim_{n\to\infty} q_n$

$$q^* = 1 - (1 - pq^*)^k$$



- Slope: $pk(1 - pq)^{k-1}$, at $q = 0$, $R_0 = pk$ - reproductive number
- When $R_0 > 1$, there is a non zero probability of infection persits
- For newsletter email campaign: $R_0 = 0.00879 * 2.96 \approx 0.26 < 1$

Two main models

- Independent Cascade Model - diminishinng returns (epidemic model):
  - spread quickly in higly connected networks
  - long range links benifit spreading
  - high degree nodes (hubs) help spreading
- Linear Threshold model - critical mass (game-theoretic model):
  - spread quickly in locally connected networks (inside clusters)
  - outside clusters slows the spreading
  - high degree nodes slow spreading

References

- Contagion, S. Morris, 2000.
- Group Formation in Large Social Networks: Membership, Growth and Evolution, L.Backstrom et.al, 2006
- Maximizing the Spread of Influence through a Social Network, D. Kempe, J. Kleinberg, E. Tardos, 2003
- Impact of Human Activity Patterns on the Dynamics of Information Diffusion, J.L. Iribarren, E. Moro, 2009
- The Dynamics of Viral Marketing, J. Leskovec, L. Adamic, B. Huberman, 2006