Diffusion of Innovation

Leonid E. Zhukov

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

Social Network Analysis

MAGoLEGO course
1. Diffusion of innovation

2. Influence propagation models
   - Linear threshold model

3. Influence maximization problem
Diffusion process

Propagation process:
- Information based models:
  - ideas, knowledge
  - virus and infection
  - rumors, news
- Decision based models:
  - adoption of innovation
  - joining political protest
  - purchase decision

Local individual decision rules will lead to very different global results. "microscopic" changes $\rightarrow$ "macroscopic" results
Ryan-Gross study of hybrid seed corn delayed adoption (after first exposure)

Information effect vs adopting of innovation

Ryan and Gross, 1943
Ryan-Gross study

Hybrid corn adoption

Percentage of total acreage plated

Griliches, 1957
Diffusion of innovation

Everett Rogers, "Diffusion of innovation" book, 1962

Diagram showing the diffusion of innovation with different adopter groups:

- Innovators: 2.5%
- Early Adopters: 13.5%
- Early Majority: 34%
- Late Majority: 34%
- Laggards: 16%

Market share,%

0 25 50 75 100

- Growth model of how new products get adopted
- Two types of agents, two key parameters:
  - $p$ - innovation or spontaneous adoption rate (coefficient of innovation)
  - $q$ - rate of imitation (coefficient of imitation)
- Let $F(t)$ fraction of agents adopted by time $t$

\[
F(t + 1) = F(t) + p(1 - F(t))\delta t + q(1 - F(t))F(t)\delta t
\]

\[
\frac{dF(t)}{dt} = (p + qF(t))(1 - F(t))
\]
Bass Diffusion model

Solution of Bass model - S-curve. When $F(0) = 0$

$$F(t) = \frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p} e^{-(p+q)t}}$$

Empirical $p \sim 0.01 - 0.03$, $q \sim 0.3 - 0.5$, when $t$ in years
Diffusion of innovation

What influences potential adopters:

- relative advantage of the innovation
- compatibility with current ways of doing things
- complexity of the innovation
- triability - the ease of testing
- observability of results

Some questions remain:

- how a new technology can take over?
- who different technologies coexist?
- what stops new technology propagation?

Everett Rogers, 1962
Network structure

From the population level to local structure
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 $0$

$$
\begin{array}{c|cc}
  & A & B \\
  \hline
  A & a, a & 0, 0 \\
  B & 0, 0 & b, b \\
\end{array}
$$
Threshold model

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 $p \geq q$
Cascades

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

Let $a = 3$, $b = 2$, threshold $q = \frac{2}{2 + 3} = \frac{2}{5}$
Let $a = 3$, $b = 2$, threshold $q = \frac{2}{(2 + 3)} = \frac{2}{5}$

Start from nodes 7, 8: $\frac{1}{3} < \frac{2}{5} < \frac{1}{2} < \frac{2}{3}$

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 $\rho$ if every node in the set has at least fraction $\rho$ of its neighbors in the set.

Both clusters of density $\rho = 2/3$. For cascade to get into cluster $q \leq 1 - \rho$.

images from Easley & Kleinberg
Linear threshold model

- Influence comes only from NN $N(i)$ nodes, $w_{ij}$ 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 $A_o$, iterative process with discrete time steps
- Progressive process, only nonactive $\rightarrow$ active

![Graphs](image.png)
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
Influence maximization

Greedy maximization algorithm:
Given: Graph and set size $k$
Output: Maximum influence set $A$

1. Select a node $v_1$ that maximizes the influence $\sigma(v_1)$
2. Fix $v_1$ and find $v_2$ such that maximizes $\sigma(v_1, v_2)$
3. Repeat $k$ times
4. Output maximum influence set: $A = \{v_1, v_2...v_k\}$
Experimental results

Linear threshold model
network: collaboration graph 10,000 nodes, 53,000 edges

Greedy algorithm finds a set $S$ such that its influence set $\sigma(S)$ is $\sigma(S) \geq (1 - \frac{1}{e})\sigma(S^*)$ from the true optimal (maximal) set $\sigma(S^*)$

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

- 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
Course summary

1. Introduction to network science
2. Descriptive network analysis
3. Mathematical models of networks
4. Node centrality and ranking on networks
5. Network communities
6. Network structure and visualization
7. Epidemics and information spread
8. Diffusion of innovation