복잡계 망에서의 정보 흐름 모델과 분석

Kyomin Jung
Applied Algorithm Lab
KAIST
Information Diffusion

- Various networks play fundamental roles as a medium for diffusion of information, ideas, and influence among its members.
  - World Wide Web
  - Infection networks
  - Co-authorship networks
  - Social Networks

- Understanding how information flows on networks, how often and when it results in large spreadings are important problems.
Threshold phenomenon

- Some kinds of influences **spread greatly** compared to others.
  - Public protests in Tunisia, Egypt, and Libya in 2001
  - The tipping point of Harry Potter in 2000

- Threshold phenomenon (appearance of **large spreading**)
  - When an information spreads **rapidly and dramatically** at a certain moment.
  - In sociology, this moment is called **tipping point**.
Applications

- **Maximization** of spreading of influences
  - Advertisement
  - Opinion spreading

- **Minimization** of spreading of bad information
  - Prevention of epidemics (vaccination)
  - Public abnormality control
Studies on Information Diffusion

- Traditionally the diffusion of innovation studied in **Sociology**
  - Adoption of hybrid corn (Ryan and Gross, 1943)
  - Diffusion of innovations among physicians (Coleman et al., 1957)
  - Innovation decision process theory (Rogers, 1962)

- Lots of models have been investigated
  - Linear threshold model
  - SIR model
Information Diffusion Models

- A network is represented as a graph. Each user is considered as a node.
- Each node can be either active or inactive.
- By the “word-of-mouth” effects, each node’s tendency to become active increases monotonically as more of its neighbors become active.
  - A node can switch to **active** from **inactive**, but does not switch in the other direction.
Individuals make their decisions based on their neighbor’s decisions.

Each individual has a threshold value $\phi_v \in [0,1]$ drawn from a distribution $f \in C^1$ in an i.i.d. manner.

If the number of neighbor nodes that accepted the innovation exceed $\phi_v$, then $v$ adopts it.

I’ll buy a smart phone, if 60% of my friends use it.
Linear Threshold Diffusion Process

Initial adopter: 1
Final cascade size: 4

Stop!
Previous Work on Linear Threshold Model

- Information spreading and the occurrence of a tipping point have been analyzed for special cases
  - Complete graph with any $f$
    (Granovetter, *The American Journal of Sociology*, 1978)
  - Infinite and locally tree-like graph with any $f$
    (Watts, *PNAS*, 2002)
  - Erdős-Rényi random network with constant $f$
Main Question

- Let \( t(k) \) be the cascade size with \( k \) proportion of initial adopters.

- Select \( k \) proportion of initial adopters uniformly at random and independently

Can we predict \( t(k) \) with high probability for a more general class of network structures and threshold distribution \( f \)?

Based on this analysis, can we predict when a tipping point will appear?

- We provide positive answers
  - Work with Seulki Lee and Hyuna Kim
Experiments

- Dataset
  - Facebook network
    - New Orleans regional network
    - $|V| = 60,290$, $|E| = 1,545,686$, average degree = 23
  - MySpace network
    - $|V| = 100,000$, $|E| = 6,854,231$, average degree = 137
  - Erdős-Rényi random network
    - $|V| = 100,000$, average degree = 100
  - Complete graph
    - $|V| = 100,000$

- Setup
  - $f \sim N(\mu, \sigma)$ with various $\mu$ and $\sigma$ values

Diffusion of innovations curve (Rogers, 1962)

KAIST Applied Algorithm Lab
Experiment Results

- For many values of $\mu$ and $\sigma$, we observe that tipping point occurs for both real world social networks and synthetic networks.
Experiment Results

\[ \mu = 0.4 \quad \sigma = 0.1 \]

\[ \mu = 0.4 \quad \sigma = 0.2 \]

Initial adopter size (%)

Cascade size (%)
**SIR Model**

- The SIR model is originally used to model diffusion of epidemics.
- An individual in a network is **susceptible** for the first time, having a possibility to be infected. After **infected**, it remains infected for a while, infecting contactees. Finally, it is **cured** (removed).

This process explains a **simple** way of information diffusions or social interactions.

- Facebook, Twitter retweet, information spreading in the blog space, etc
SIR Model

The SIR spreading procedure examples

$p_{i,j} = 0.5$

$p_{i,j} = 0.75$

kaist applied algorithm lab
Our Interests

- We are interested in ...
  - **Probabilities** of large spreading
  - **Sizes** of large spreading
  - **Conditions** under which large spreading occurs
Outline of Our Results

- Work with Sungsu Lim and Namju Kwak

- Previous work considers only the case when the diffusion probability is a constant for each edge.

- We consider when the diffusion probability depends on the local information of the two end nodes, which appears often in social networks and complex networks.

- We obtain formula to exactly compute probabilities and sizes of large spreading of a network under the SIR model using the degree distribution of the network.

- The results of our mathematical calculations are very similar to the empirical results.
Simulations

- Use the SIR spreading model.
  \[ p_{i,j} = f(d_i, d_j) = \frac{c}{d_i} \text{ and } p_{i,j} = f(d_i, d_j) = \frac{c}{d_j} \]

- Simulations are performed on ...
  - Preferential attachment graph
  - General random graph
  - Facebook and Myspace friendship graph

- A single initial infectious (I) node is randomly picked. All the other nodes are susceptible (S).

- Observe probabilities and sizes of large spreading at the end of the procedure.
Simulations

pref_att_dj_c1.0

0.01% of experiment cases have resulted in large spreadings with an average size 100.00%.

pref_att_dj_c1.5

37.87% of experiment cases have resulted in large spreadings with an average size 57.91%.

pref_att_dj_c2.0

61.61% of experiment cases have resulted in large spreadings with an average size 80.45%.

pref_att_dj_c2.5

75.80% of experiment cases have resulted in large spreadings with an average size 90.21%.
Simulations

- Facebook_dj_c1.0
  - 0.01% of experiment cases have resulted in large spreadings with an average size 100.00%.

- Facebook_dj_c1.5
  - 22.82% of experiment cases have resulted in large spreadings with an average size 54.24%.

- Facebook_dj_c2.0
  - 40.46% of experiment cases have resulted in large spreadings with an average size 80.93%.

- Facebook_dj_c2.5
  - 52.87% of experiment cases have resulted in large spreadings with an average size 91.38%.
Simulations

- For 'facebook di c2.0', 80.60% of experiment cases have resulted in large spreadings with an average size 40.86%.

- For 'general di c2.0', 77.66% of experiment cases have resulted in large spreadings with an average size 32.89%.

- For 'facebook di c2.0', 40.46% of experiment cases have resulted in large spreadings with an average size 80.93%.

- For 'general di c2.0', 32.54% of experiment cases have resulted in large spreadings with an average size 77.85%.