Limiting the Spread of Misinformation in Social Networks

By C. Budak, D. Agrawal and A. Al Abbadi
Department of Computer Science - UCSB
Article published in KDD10

Presented in CS512 by G. Pitois - 4/06/2011
Introduction

- Information spreads in Social Networks
  - TV, newspaper, web (Facebook, email)
- Information and Misinformation are competing
- How to limit the influence of Misinformation?
- Plan:
  - Modeling the spread of information in Social Networks
  - The Eventual Influence Limitation (EIL) problem
  - Experimentation on real data from Facebook
Spread of Information - 1

- Social Networks
  - Can be modeled as a (directed) graph $G = (N, V)$

- Independent Cascade
  - Time is modeled by a succession of steps
  - At each step $t$, a set of nodes are activated
  - An active node $v$ has a chance to activate its neighbours $w$ with a probability $p_{v,w}$ for step $t+1$
Spread of Information - 2

Cascade of information
Spread of Information - 3

- Two Diffusion Models are used:
  - Multi-Campaign Independent Cascade (MCICM)
    - Campaign C (bad information), origin from node $n_a$ (adversary node)
    - Limiting Campaign L starts with delay $r$
    - Campaign have independent probabilities of diffusion
  - Campaign-Oblivious Independent Cascade (COICM)
    - Probability of diffusion is independent of the campaign
    - Good model when both campaigns have similar quality of information
Spread of Information - 4

Multi-Cascade Independent Cascade Model (MCICM)

Probability of diffusion on each edge depends on the campaign

Campaign L
Starts with delay $r$
Spread of Information - 5

Probability of diffusion on each edge is independent of the campaign

Campaign-Oblivious Independent Cascade Model (COICM)
Eventual Influence Limitation (EIL) - 1

- Characteristics of the EIL problem
  - Minimize the number of nodes that would be infected by (bad) campaign C
  - Assume campaign C starts from a single node $n_a$
  - Determine an optimal set $A_L$ of $k$ nodes to start the limiting campaign L
  - NP-Hard problem 😞
Eventual Influence Limitation (EIL) - 2

MCICM with High efficiency

k=2
Eventual Influence Limitation (EIL) - 3

- Solutions to the EIL problem
  - Greedy algorithm
    - Too expensive for real social networks
  - Degree centrality
    - Target “influential people”
    - Do not consider if node may be infected
  - Early infectees
    - Choose nodes supposed to be infected at step r (delay time)
  - Largest infectees
    - Choose nodes that would infect the highest number if they were infected themselves
Experimentation

• Data sets
  • Regional networks from Facebook
    • SB2008: 13,000 nodes  184,000 edges
    • SB2009: 26,000 nodes  453,000 edges
    • Some others…

• Parameters
  • Degree of centrality of adversary node ($n_a$)
  • Delay $r$ of limiting campaign $L$
  • Weight distribution of probabilities for $C$ and $L$
SB 2008 - Multi-Campaign Independent Model with high efficiency
Results - 2

(a) $\text{delay} = 20\%$

(b) $\text{delay} = 70\%$

(c) $\text{delay} = 20\%, \text{adversary degree} \geq 40$

SB 2008 – Campaign-Oblivious Independent Model
Results - 3

- After running tests on the two models
  - MCICM (Multi-Campaign)
  - COICM (Campaign-Oblivious)

- Delay is very important

- Early infectees is unstable

- Largest infectees compare to greedy method

- To select the best method, consider:
  - Delay
  - Connectedness of adversary node ($n_a$)
Conclusion

- EIL problem is NP-Hard
- Heuristics can compare to greedy approach

Future work
- Parallel algorithm for computing saved nodes
- Minimize budget k (initial size of L campaign)
- Possibility for a node to change its mind

- Thank you for your attention!!