Online Model-Free Influence Maximization with Persistence

Paul Lagrée, Olivier Cappé, Bogdan Cautis, Silviu Maniu

LRI, Univ. Paris-Sud, CNRS, LIMSI & Univ. Paris Saclay

May 9, 2017

Background & Motivations

Classic Influence Maximization [Kempe et al., 2003]

Important problem in social networks, with applications in marketing, computational advertising.

- **Objective:** Given a promotion budget, maximize the influence spread in the social network (word-of-mouth effect).
- Select k seeds (influencers) in the social graph, given an graph $G = (V, E)$ and a propagation model
- Edges correspond to follow relations, friendships, etc. in the social network

Problem

IM Optimization Problem

Denoting $S(I)$ the influence cascade starting from a set of seeds I, the objective of the IM is to solve the following problem

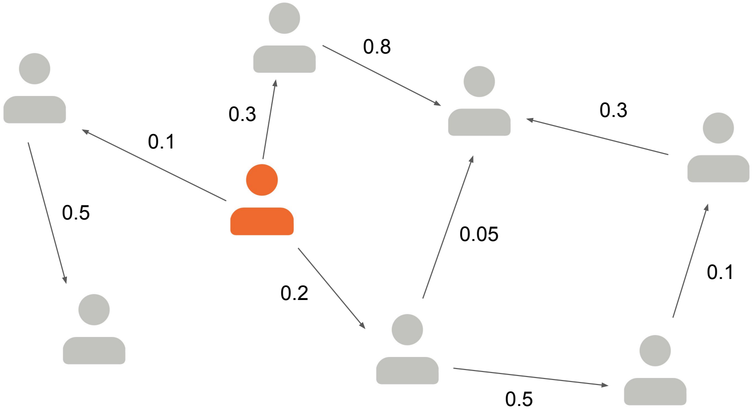
$$\underset{I \subseteq V, |I| = L}{\arg\max}\ \mathbb{E}[S(I)].$$

Independent Cascade Model

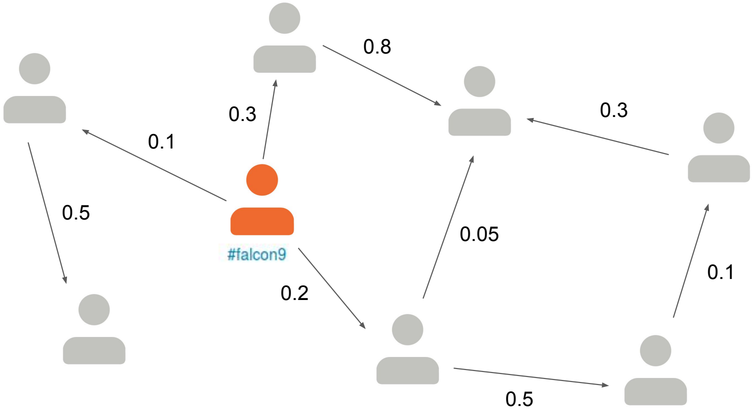
To each edge $(u, v) \in E$, a probability $p(u, v)$ is associated

1. at time 0 – activate seed s
2. node u activated at time t – influence is propagated at $t + 1$ to neighbours v independently with probability $p(u, v)$
3. once a node is activated, it cannot be deactivated or activated again.

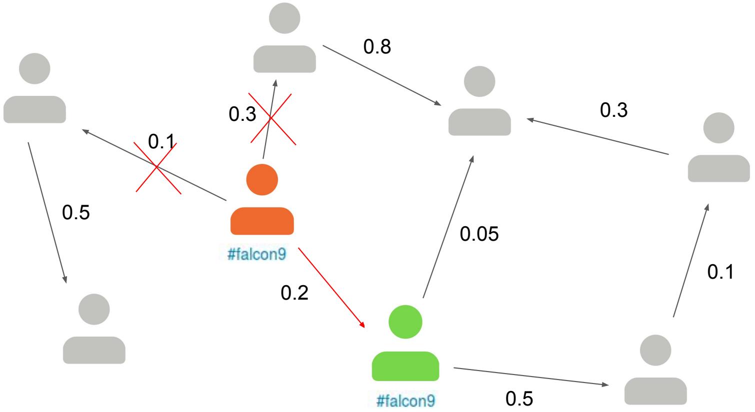
Independent Cascade Model – Example



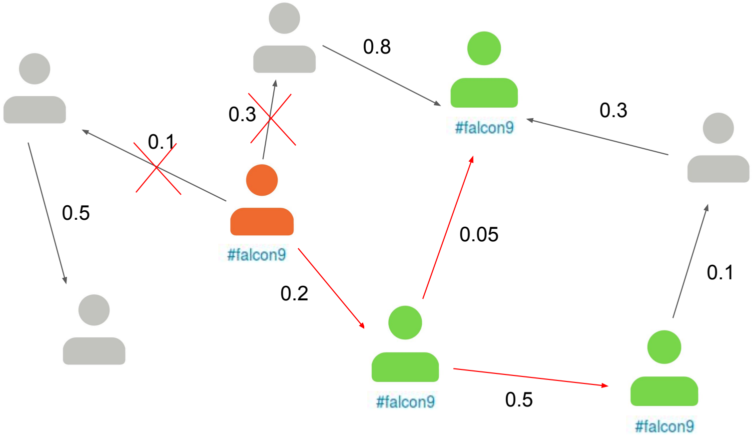
Independent Cascade Model – Example



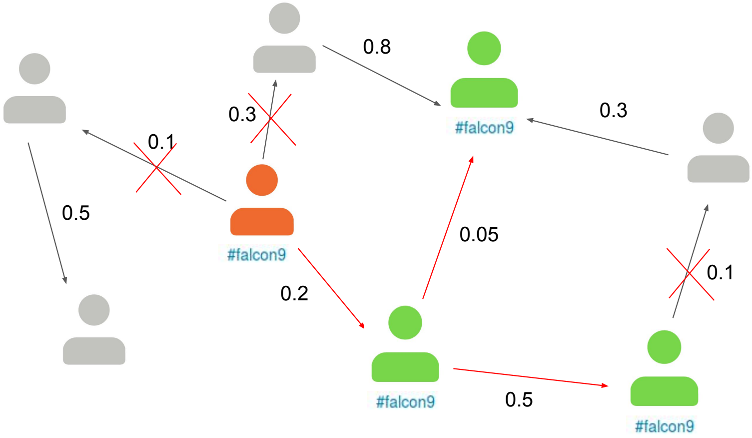
Independent Cascade Model – Example



Independent Cascade Model – Example



Independent Cascade Model – Example



Approximated IM algorithms

1. Computing expected spread: Monte Carlo simulations
2. for solving the IM: greedy approximation algorithm

Multiple algorithms and estimators: TIM / TIM+ [Tang et al., 2014], IMM [Tang et al., 2015], SSA [Nguyen et al., 2016], PMC [Ohsaka et al., 2014], . . .

Online Influence Maximization

Online Influence Maximization

We only know the social graph, but not edge probabilities. Problem introduced by [Lei et al., 2015] for the IC model.

1. at trial n — select a set of k seeds,
2. the diffusion happens, observe activated nodes and edge activation attempts
3. repeat to step 1 until the budget is consumed.

Online Influence Maximization with *Persistence*

OIMP Problem [Lei et al., 2015]

Given a budget N, the objective of the *online influence maximization with persistence* is to solve the following optimization problem

$$\underset{I_n \subseteq V, |I_n| = L, \forall 1 \leq n \leq N}{\arg\max} \mathbb{E} \left| \bigcup_{1 \leq n \leq N} S(I_n) \right| .$$

A node can be activated several times at different trials, but it will yield reward only once

Motivations

A campaign with several steps: different posts with a single semantics.

- people may transfer the information several times, but “adopting” the concept rewards only once (e.g. politics)
- brand fanatics, e.g., Star Wars, Apple, etc.
- social advertisement in users’ feed (e.g. Twitter / Facebook), people may transfer/ like the content several times across the campaign.

Online Model-Free Influence Maximization with Persistence

Setting

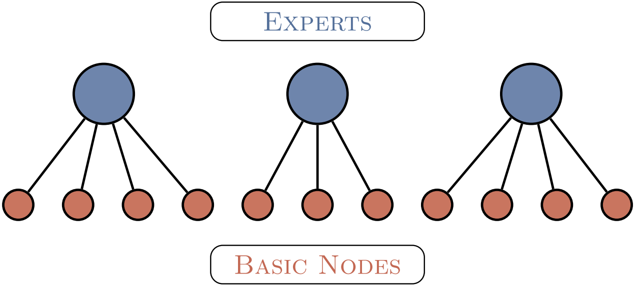
In the following, we work in

- the persistent setting
- no assumption regarding the diffusion model
- simple feedback: set of activated nodes

Simple, **realistic**, target **short horizons**

Model

To simplify the graph problem, we consider the corresponding graph (depth-1 trees):



- **Hypothesis:** empty intersection between experts
- **New problem:** estimating the missing mass of each expert, that is, the expected number of nodes that can still be reached from a given seed.

Missing mass

Following the work of [Bubeck et al., 2013]

- Missing mass $R_n := \sum_{u \in A} \mathbb{1}\left\{u \notin \bigcup_{i=1}^{n} S_i\right\} p(u)$
- Corresponds to the potential of the expert
- Missing mass estimator (known as the Good-Turing estimator)

$$\hat{R}_n := \sum_{u \in A} \frac{U_n(u)}{n},$$

where $U_n(u)$ is the indicator equal to 1 if x has been sampled exactly once.

The estimator is the fraction of hapaxes!

Confidence Bounds

Estimator Bias

$$\mathbb{E}[R_n] - \mathbb{E}[\hat{R}_n] \in \left[-\frac{\sum_{u \in A} p(u)}{n}, 0 \right]$$

Theorem

With probability at least $1 - \delta$, *denoting* $\lambda := \sum_{u \in A} p(u)$ *and* $\beta_n := (1 + \sqrt{2}) \sqrt{\frac{\lambda \log(4/\delta)}{n}} + \frac{1}{3n} \log \frac{4}{\delta}$, *the following holds:*

$$-\beta_n - \frac{\lambda}{n} \leq R_n - \hat{R}_n \leq \beta_n.$$

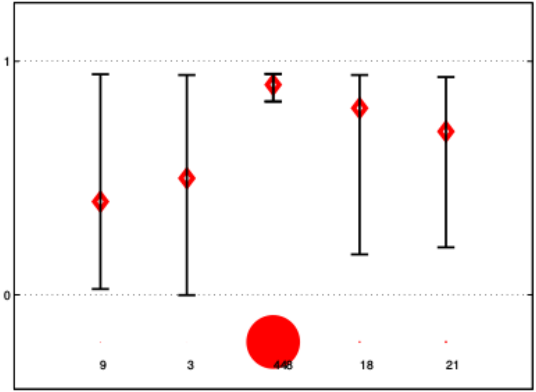
Algorithm

- UCB-like algorithm
- at round t, we play the expert k with largest index

$$b_k(t) := \hat{R}_k(t) + (1+\sqrt{2})\sqrt{\frac{\hat{\lambda}_k(t)\log(4t)}{N_k(t)}} + \frac{\log(4t)}{3N_k(t)},$$

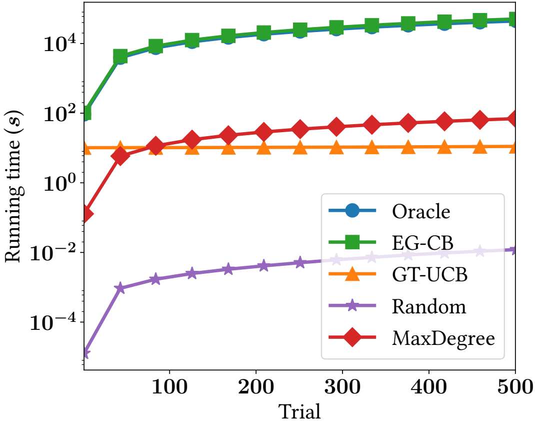
where $N_k(t)$ denotes the number of times expert k has been played up to round t

Optimism in Face of Uncertainty



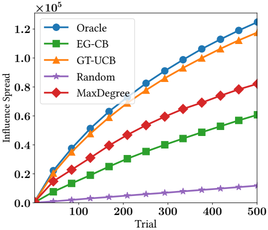
Experiments

Execution time (DBLP)

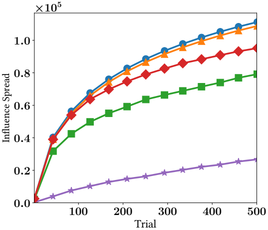


DBLP (WC – $L = 1$)

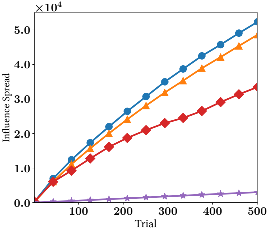
Growth of spreads (DBLP)



DBLP (WC – $L = 5$)



DBLP (TV – $L = 5$)



DBLP (LT – $L = 5$)

Thank you.

References

David Kempe, Jon Kleinberg and Éva Tardos (2003)
Maximizing the Spread of Influence Through a Social Network.
Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 137–146.

Siyu Lei, Silviu Maniu, Luyi Mo, Reynold Cheng and Pierre Senellart (2015)
Online Influence Maximization
SIGKDD.

Sébastien Bubeck, Damien Ernst and Aurélien Garivier (2013)
Optimal Discovery with Probabilistic Expert Advice: Finite Time Analysis and Macroscopic Optimality
Journal of Machine Learning Research, 601 – 623.

Wei Chen, Yajun Wang, Yang Yuan and Qinshi Wang (2016)
Combinatorial Multi-armed Bandit and Its Extension to Probabilistically Triggered Arms
Journal of Machine Learning Research, 1746 – 1778.

Sharan Vaswani, V.S. Lakshmanan and Mark Schmidt (2015)
Influence Maximization with Bandits
Workshop NIPS