An Efficient Randomized Algorithm for Rumor Blocking in Online Social Networks

Shao-Heng Ko

Theory Study Group, August 2, 2017

A survey on randomized algorithms in social network diffusion

~~An Efficient Randomized Algorithm for Rumor Blocking in Online Social Networks~~

Shao-Heng Ko

Theory Study Group, August 2, 2017

Info Diffusion in Social Networks

- **SIR (Susceptible-Infected-Recovered) model**
  - biology, physics
- **Tipping model (deterministic LT)** 1978
  - economics, sociology
- **IC (Independent Cascade) &**
  LT (Linear Threshold) models Kempe, 2003
  - computer science, data mining

The IC and LT models Kempe, 2003

- Directed Social Graph $G = (V, E)$
- Nodes are either **inactive** or **active** (model the spread of an innovation/idea)
- **Progressive** : nodes do not switch from active to inactive
- **Seeds** : some nodes are active at the beginning
- **Edge Weights** (model the influence)
- **Activation Rules**

IC model

- **Edge Weights** : the probability of information diffusion
- **Activation Rules** : each node, once activated,
  has exactly one chance to activate others (by all its out-edge)
- **History-ignorant**

LT model

- **Edge Weights** : represents the amount of influence
- **Activation Rules** : each node is activated once enough portion of its neighbors are active
- **Deterministic / Random Threshold**

Influence Maximization Problem

- Denote A_0 to be the set of seed nodes
- Denote $\sigma(A_0) = \mathrm{E}[\#\text{ of active nodes at the end, given } A_0]$
- **1. Given $\boldsymbol{A_0}$, find $\boldsymbol{\sigma(A_0)}$: #P-hard for IC** Chen, 2010
- **2. Find $\boldsymbol{A_0}$ to maximize $\boldsymbol{\sigma(A_0)}$** (Influence Maximization Problem) :
  - **NP-hard for both IC and LT** (and all extensions)
  - **$(1-\frac{1}{e})$-approx. greedy algorithm (by submodularity)** (for most extensions)

Caveat

- The $(1-\frac{1}{e})$-approx. greedy algorithm is polynomial in **queries of prob.1**!

◢▆▅▄▃ 崩╰(〒皿〒)╯潰 ▃▄▅▆◣

- Exact computation is infeasible → sampling
- **Q: How to sample?**

Randomized Greedy Algorithms (1)

- Naïve Monte Carlo simulations: Chen, 2009; Chen, 2010
  $\Omega(k \cdot m \cdot n \cdot poly(\epsilon^{-1}))$ to achieve $(1 - \frac{1}{e} - \epsilon)$-approx. with const. prob.

- **Reverse Sampling** Borgs, 2014
  $O(k(m+n)\log(n)\epsilon^{-3})$ to achieve $(1 - \frac{1}{e} - \epsilon)$-approx. with const. prob. $\frac{3}{5}$

Randomized Greedy Algorithms (2)

- Reverse Sampling Borgs, 2014
  $O(k(m+n)\log(n)\epsilon^{-3})$ to achieve $(1-\frac{1}{e}-\epsilon)$-approx. with const. prob. $\frac{3}{5}$
  With amplifying
  $O(k\boldsymbol{l^2}(m+n)\log^{\boldsymbol{2}}(n)\epsilon^{-3})$ to achieve $(1-\frac{1}{e}-\epsilon)$-approx. w. p. $1-\frac{1}{n^l}$
- TIM (Two-Phase Influence Maximization) Tang, 2014; Tang, 2015
  $O((\boldsymbol{k}+\boldsymbol{l})(m+n)\boldsymbol{\log(n)}\epsilon^{\boldsymbol{-2}})$ to achieve $(1-\frac{1}{e}-\epsilon)$-approx. w. p. $1-\frac{1}{n^l}$

Extensions of IC/LT model

- Weighted and Competitive Scenario (To name a few)
  - Many influence parties on IC Carnes, 2007, on LT Borodin, 2010
  - Sequential Competitive Bharathi, 2007, PoA = 2
- Rumor Blocking
  - Care about **# of nodes not activated by rumor sources**
  - Formulated on IC model Budak, 2011, on LT model He, 2012

Rumor Blocking

- The particular paper Tong, 2017 is utilizing **Reverse Sampling** method on **Rumor Blocking**
- **Main Result:**
  A $\boldsymbol{O}\left(\frac{\boldsymbol{km}\ \mathbf{ln}(\boldsymbol{n})}{\boldsymbol{\epsilon}^{\mathbf{2}}}\right)$ -randomized algorithm to achieve $(1-\frac{1}{e}-\epsilon)$-approx. with high probability.
- Monte-Carlo-based: $\boldsymbol{O}\left(\frac{\boldsymbol{k}^{\mathbf{3}}\boldsymbol{mn}\ \mathbf{ln}(\boldsymbol{n})}{\boldsymbol{\epsilon}^{\mathbf{2}}}\right)$

Reverse Sampling Borgs, 2014

- **Build a sparse, undirected hypergraph H by repeated simulation**
  - Start from a random node u
  - Stochastic-DFS backward to find all vertices that “influences” u
  - All these vertices become one edge in H
  - Repeat for R steps (not rounds!), return H
- **Greedily choose nodes with highest degree in H**

Intuitive Idea

- Influence of a node set $S = n \cdot \Pr[\text{DFS from random } u \text{ reaches } S]$
- In H, $\deg(u)$ $\approx$ influence of u if H big enough to estimate the influence
- Greedy on H is a good approximate to the original greedy
- achieve $(1 - \frac{1}{e} - \epsilon)$-approx. with prob. $\frac{3}{5}$
- Simulate many times, take the largest generated hypergraph: success prob. amplified

Runtime-Quality Tradeoff

- We can obtain $(1 - \frac{1}{e} - \epsilon)$ – approx. in time $O((m+n)\log(n)\epsilon^{-3})$
- Want to obtain $O(\beta)$ - approx. in time $O(\beta(m+n)\log(n))$
- Simply modify the hypergraph size: sometimes fail
  - Too less data to guess which node has maximum influence
  - Large edges in the hypergraph
  - Randomly select nodes with prob. proportional to its hypergraph degree

Dynamic Runtime

- β need not to be given
- Run the greedy algorithm when the hypergraph construction reaches 2^i steps
- When the algorithm is terminated without warning, return the most recent solution

Q&A

Kempe, 2003

- Kempe, D., Kleinberg, J., & Tardos, É. (2003, August). Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 137-146). ACM.

Carnes, 2007

- Carnes, T., Nagarajan, C., Wild, S. M., & Van Zuylen, A. (2007, August). Maximizing influence in a competitive social network: a follower's perspective. In *Proceedings of the ninth international conference on Electronic commerce*(pp. 351-360). ACM.

Bharathi, 2007

- Bharathi, S., Kempe, D., & Salek, M. (2007). Competitive influence maximization in social networks. *Internet and Network Economics*, 306-311.

Chen, 2009; Chen, 2010

- Chen, W., Wang, Y., & Yang, S. (2009, June). Efficient influence maximization in social networks. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 199-208). ACM.
- Chen, Wei, Chi Wang, and Yajun Wang. "Scalable influence maximization for prevalent viral marketing in large-scale social networks." *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2010.

Borodin, 2010

- Borodin, A., Filmus, Y., & Oren, J. (2010, October). Threshold Models for Competitive Influence in Social Networks. In *WINE* (Vol. 6484, pp. 539-550).

Budak, 2011

- Budak, C., Agrawal, D., & El Abbadi, A. (2011, March). Limiting the spread of misinformation in social networks. In *Proceedings of the 20th international conference on World wide web* (pp. 665-674). ACM.

He, 2012

- He, X., Song, G., Chen, W., & Jiang, Q. (2012, April). Influence blocking maximization in social networks under the competitive linear threshold model. In *Proceedings of the 2012 SIAM International Conference on Data Mining* (pp. 463-474). Society for Industrial and Applied Mathematics.

Borgs, 2014

- Borgs, C., Brautbar, M., Chayes, J., & Lucier, B. (2014, January). Maximizing social influence in nearly optimal time. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms* (pp. 946-957). Society for Industrial and Applied Mathematics.

Tang, 2014

- Tang, Y., Xiao, X., & Shi, Y. (2014, June). Influence maximization: Near-optimal time complexity meets practical efficiency. In *Proceedings of the 2014 ACM SIGMOD international conference on Management of data* (pp. 75-86). ACM.

Shakarian, 2015 (A survey book)

- Shakarian, P., Bhatnagar, A., Aleali, A., Shaabani, E., & Guo, R. (2015). *Diffusion in social networks*. Springer International Publishing.

Tong, 2017

- Tong, G., Wu, W., Guo, L., Li, D., Liu, C., Liu, B., & Du, D. Z. (2017). An Efficient Randomized Algorithm for Rumor Blocking in Online Social Networks. *arXiv preprint arXiv:1701.02368*.