Optimizing Information Propagation in Social Network

OIE 559 Advanced Prescriptive Analytics

Aukkawut Ammartayakun

Worcester Polytechnic Institute

October 11, 2022

Why?

- A lot of information
- Let say that we know the truth that we are living in the simulation; who should spread the news so that most people are *aware*?

Why not just look into the cascade or influence maximization problem?

- Traditionally, we look into the spread by looking at the behavior of vertices.
- Suppose we have a friend who has a different political view. They don't need to share the news from their side!
  - They can also share the news that opposes them and criticize that too!
  - "We are living in the simulation" I agree.
  - Who is dumb enough to trust this news "Scientist confirmed! We are not living in the simulation".
- In this case, looking at the side the vertices are on is not enough.

If that is the case, why don't we say it is "independent"?

- Although the flow of information now seems to be independent, the decision of the spreading is depending on the side they are in!
  - That is, there will be more tendency to spread the information they are sided on.

Assumptions

- Each person in the social network has their own probability of spreading the information and adopting the information on their own.
  - That is, even if you don't join the force to break the simulation (or even if you join the enemy team), you can still spread the information about fighting against it.
  - However, the probability of spreading information will depend on which side they are in given the information.
- Unlike game theory problem, enemy will not just choose the best candidate for this problem. Instead, they will choose the node randomly.

Problem Statement

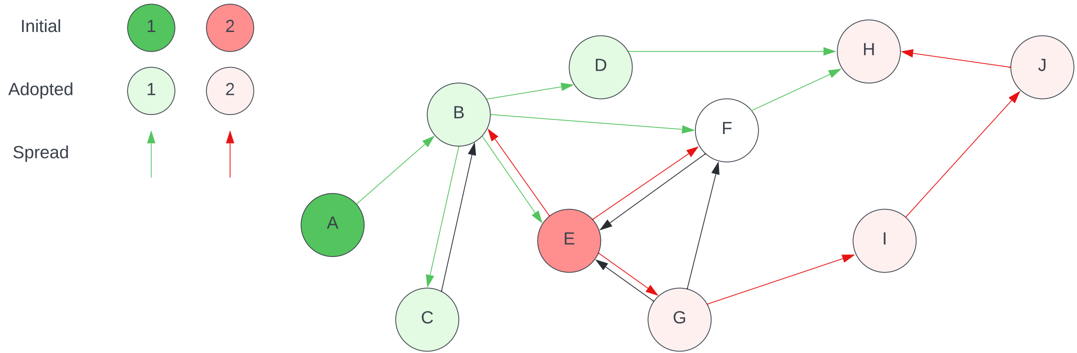


Figure: Problems

Exploration

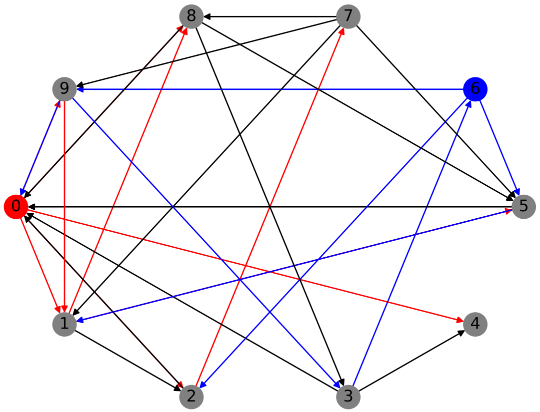


Figure: Example run with $ER(10, 0.3)$ and $p_P = 0.5$. Node 0 is picked as our initial propagator. In this case $\mathbb{E}_G\left[|E_A|\,|v_A = 0, v_B = 6\right] = 8$ and $\mathbb{E}_G\left[|E_B|\,|v_B = 6, v_A = 0\right] = 7$

Exploration

- This game is not fair!
  - That means, for some graph, uniformly random strategy will result in a lose.
  - In 500 random trials (of 100 graphs, i.e., 50000 runs) of size 10 with $p_P = 0.5$ will favors on one player (88%) compared to the tie game (12%).

Analysis of the Game

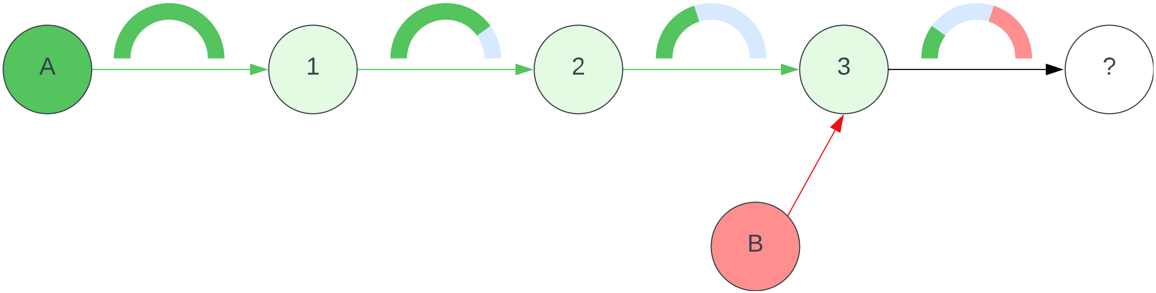


Figure: Toy example

- The information will diffused. i.e., the longer the chain is, the less chance of propagating the information I_A

Naive Approach

- We want to maximize the propagation probability of the furthermost node that is connected to the source.

Naive Approach

- We want to maximize the propagation probability of the furthermost node that is connected to the source.
- **Decision variable**: an initial node $v^{\text{init}} \in V_A$

Stochastic Programming

Traditional two-state stochastic programming formulation would be

$$\max_{x \in \Omega_x} c^\top x + \mathbb{E}_\xi \left[Q(x,\xi) \right] \tag{1}$$

where

$$Q(x,\xi) = \max_{y \in \Omega_Q(x)} q(\xi)^\top y \tag{2}$$

in which $Q(x,\xi)$ defines the second stage (recourse) decision which is probabilistic.

- In this problem, $c^\top x = 0$ because there is no cost on picking the point.
- However, there is a decision constraint where we can pick the available point.

Reinforcement Learning?

Recall the stochastic programming formulation in (1)

$$\max_{x \in \Omega_x} c^\top x + \mathbb{E}_\xi \left[Q(x, \xi) \right]$$

This process is Markovian. Thus, from the value function:

$$v_\pi(s) = \mathbb{E}_\pi \left[\sum_{k=t}^{\infty} \gamma^k r_k \middle| S_t = s \right]$$

is essentially telling us that we can explore the environment, even if we have only one action after the point is selected (i.e., the path after we picked the point is always optimal)!

- Suppose we don't discount the reward and let $r = 1$ at that state if we convert one edge.
- This is just a max flow problem.

Naive Formulation

This method is the simplified version of the stochastic programming proposed in (Wu *et al.*, 2018)

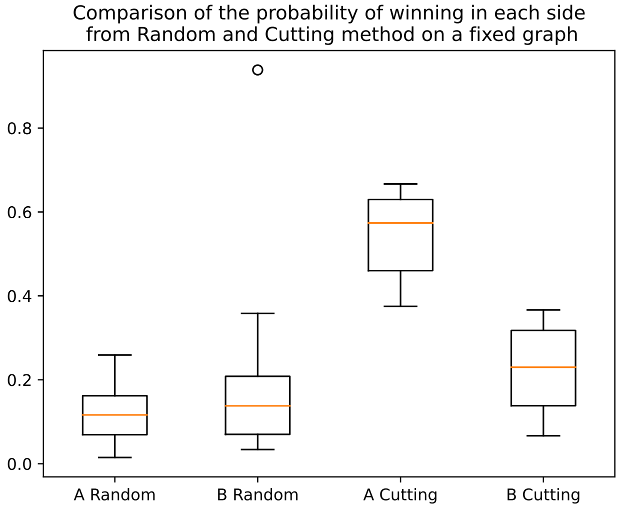
- Consider the graph G with initial node s. We initialize the flow $f = 1$ to the network from x. Our goal is to maximize the flow. This is the max-flow problem.

$$\begin{aligned} \max \quad & v \\ \text{s.t.} \quad & \sum_{j:(s,j)\in E} x_{sj} \leq v \\ & \sum_{i:(i,t)\in E} x_{it} \leq v \\ & \sum_{(i,j)\in E} x_{ij} = \sum_{(j,i)\in E} x_{ji} \\ & x_{ij} > 0 \end{aligned}$$

Naive Competitive Max Flow Formulation

- We will use the Markov decision process (MDP) to find the policy (i.e., our initial points) such that $v_{\pi=x}(s)$ is maximize (i.e., satisfying equation (1))
  - Let say the agent will get $+1$ reward if it wins the competition (that purely based on chance) on the edge and 0 otherwise.
- Without reward discounting factor, it is just saying that estimate the distribution of payoff on each edges.
- Then, let say we want to win, why don't we forfeit the edge that we will lose anyways?
  - **Pruning** the edge that are not favorable to us. Then, find the node that gives maximum flow.

Simulations



Future Works

- Prove the optimality (or suboptimality).
  - This heuristic model is sub-optimal. i.e., Not guaranteed to provide an optimal solution
- Optimize the model computational complexity
- What if discount factor is not 1?

References I

Erdős, P., and Rényi, A.
On the evolution of random graphs.
In *Publication of the Mathematical Insitute of the Hungarian Academy of Science* (1960), pp. 17–61.

Goldberg, A., and Stein, S. K.
Beyond social contagion: Associative diffusion and the emergence of cultural variation.
American Sociological Review 83, 5 (2018), 897–932.

Güney, E.
An efficient linear programming based method for the influence maximization problem in social networks.
Information Sciences 503 (2019), 589–605.

References II

Li, C., and Grossmann, I. E.
A review of stochastic programming methods for optimization of process systems under uncertainty.
Frontiers in Chemical Engineering 2 (2021).

Newman, M. E. J., Watts, D. J., and Strogatz, S. H.
Random graph models of social networks.
Proceedings of the National Academy of Sciences 99, suppl_1 (2002), 2566–2572.

Rabb, N., Cowen, L., de Ruiter, J. P., and Scheutz, M.
Cognitive cascades: How to model (and potentially counter) the spread of fake news.
PLOS ONE 17, 1 (01 2022), 1–32.

Sutton, R. S., and Barto, A. G.
Second ed.

References III

WU, H.-H., AND KÜÇÜKYAVUZ, S.
A two-stage stochastic programming approach for influence maximization in social networks.
Computational Optimization and Applications 69, 3 (Apr 2018), 563–595.