Project1

Information Exposure Maximization

Heuristic Search

Outline

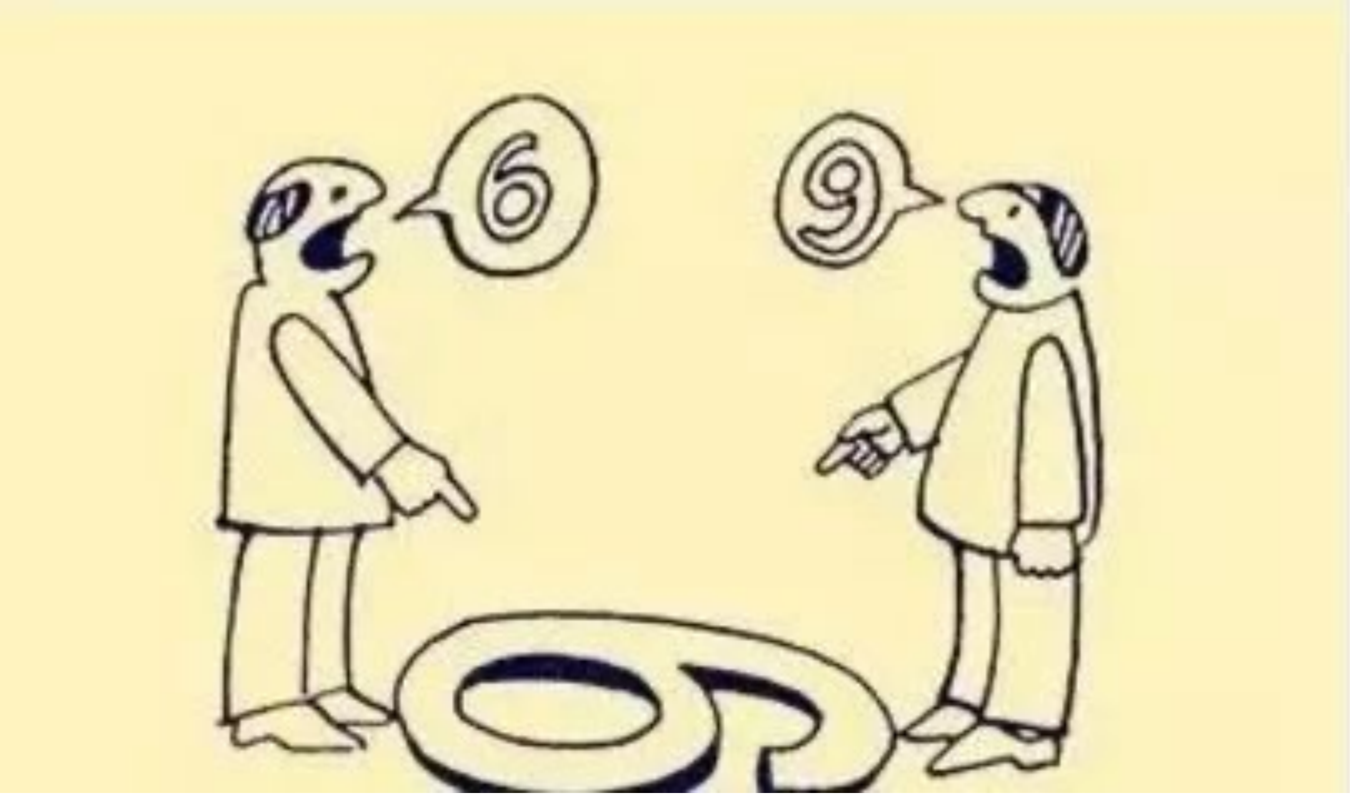
- A brief review of information exposure maximization
- An estimation method for balanced information exposure
- A heuristic algorithm for information exposure maximization
- Summary

Outline

- **A brief review of information exposure maximization**
- An estimation method for balanced information exposure
- A heuristic algorithm for information exposure maximization
- Summary

Brief review of IEM

- The Information Exposure Maximization (IEM) problem is proposed to solve the echo chamber effect on social media.



Brief review of IEM

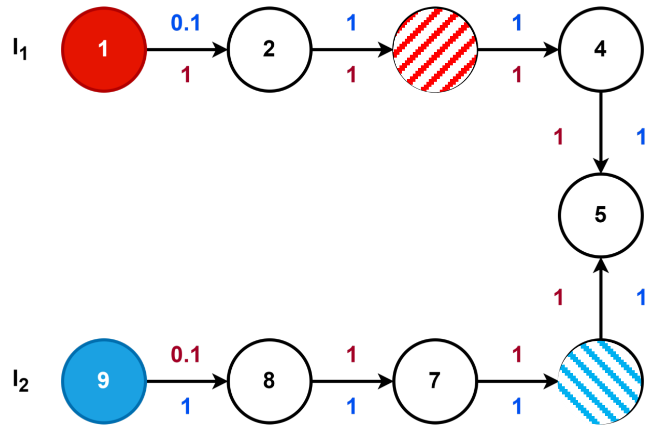
Given a social network $G = (V, E)$, two initial seed sets I_1 and I_2, and a budget k.

The IEM is **to find two balanced seed sets S_1 and S_2**, where $|S_1| + |S_2| \leq k$, and **maximize the balanced information exposure**, i.e.,

$$\max \Phi(S_1, S_2) = \max \mathbb{E}\left[\left|V \setminus \left(r_1(I_1 \cup S_1) \,\triangle\, r_2(I_2 \cup S_2)\right)\right|\right]$$

$$\text{s.t.} \quad |S_1| + |S_2| \leq k$$

$$S_1, S_2 \subseteq V$$



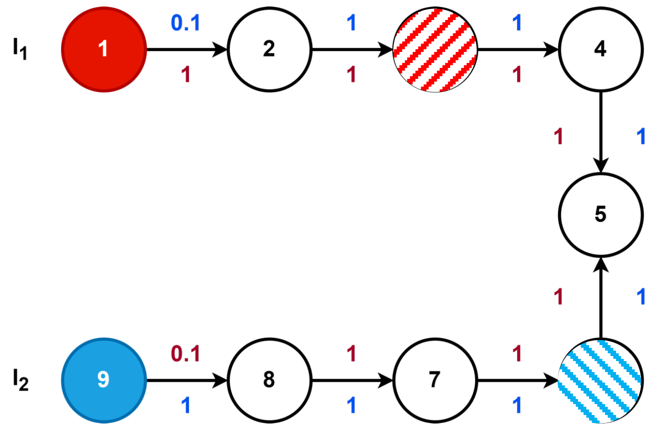
Brief review of IEM

Given a social network $G = (V, E)$, two initial seed sets I_1 and I_2, and a budget k.

The IEM is **to find two balanced seed sets S_1 and S_2**, where $|S_1| + |S_2| \leq k$, and **maximize the balanced information exposure**, i.e.,

$$\max \Phi(S_1, S_2) = \max \mathbb{E}\left[\left|V \setminus \left(r_1(I_1 \cup S_1) \triangle r_2(I_2 \cup S_2)\right)\right|\right]$$

Both are **random variables** determined by the stochastic process of the diffusion model and their diffusion probabilities



Brief review of IEM

SUSTech Southern University of Science and Technology NICAL

- Finding an optimal solution of IEM is NP-hard.
- Computing the balanced information exposure for a given solution is NP-hard.

Outline

- A brief review of information exposure maximization
- **An estimation method for balanced information exposure**
- A heuristic algorithm for information exposure maximization
- Summary

Balanced information exposure estimation

Monte Carlo simulation

- A computational algorithm that uses **repeated random sampling** to obtain the likelihood of a range of results of occurring

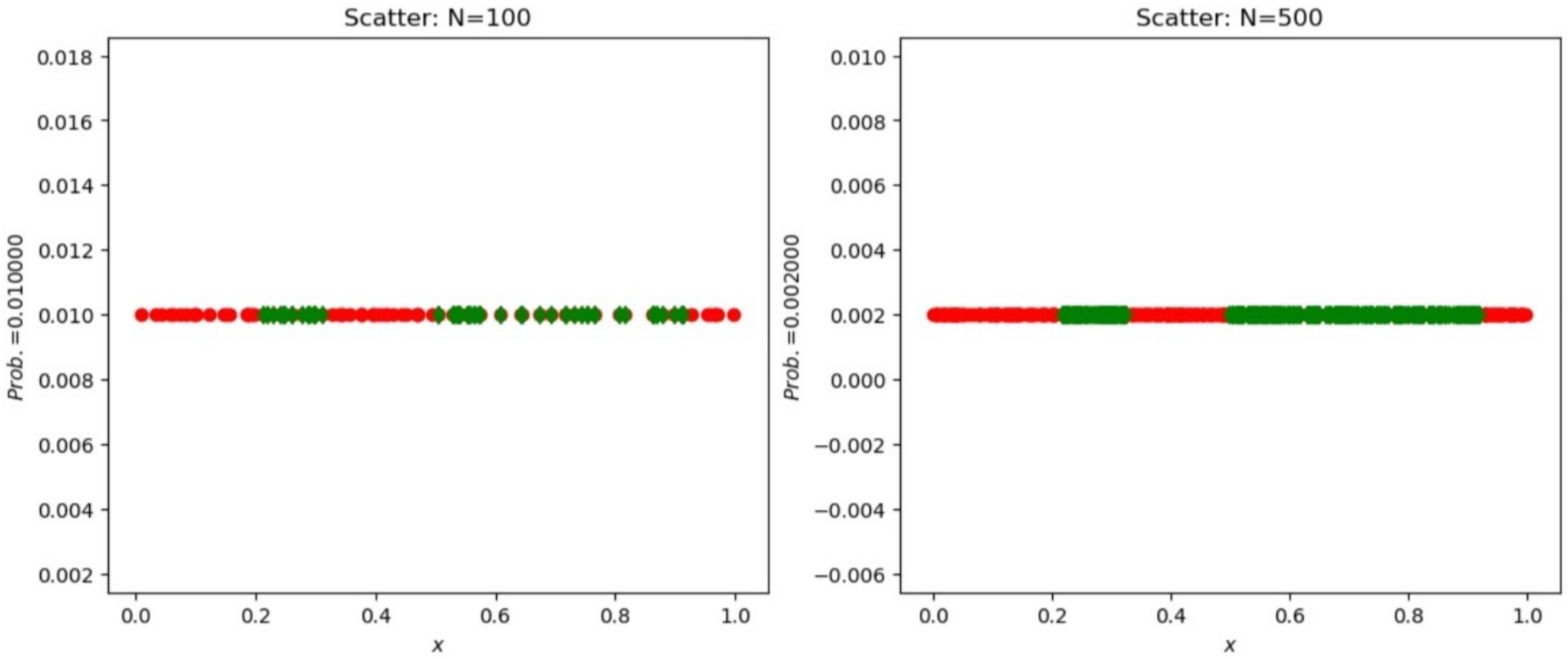
Balanced information exposure estimation

Monte Carlo simulation

- A computational algorithm that uses **repeated random sampling** to obtain the likelihood of a range of results of occurring

Example1:

Estimate the length of the green segments on a line



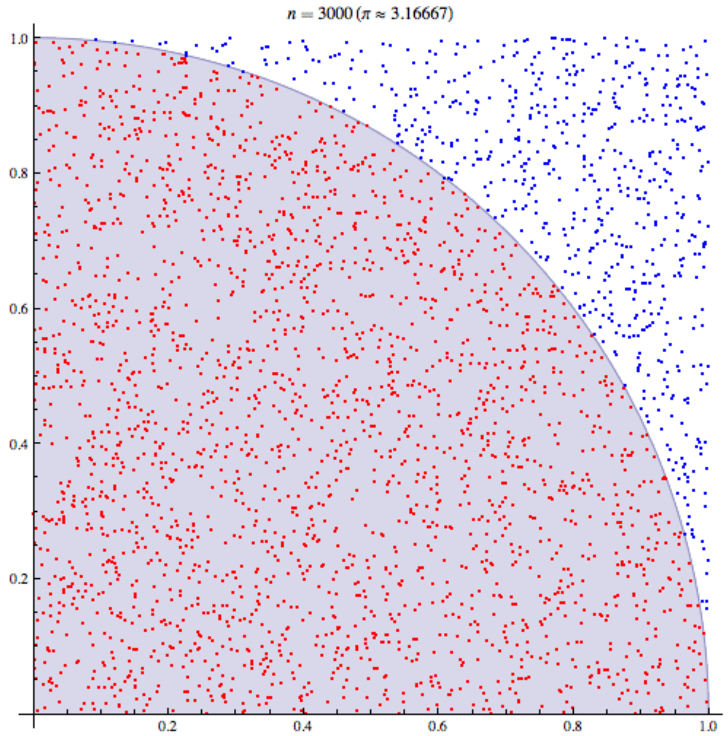
Balanced information exposure estimation

Monte Carlo simulation

- A computational algorithm that uses **repeated random sampling** to obtain the likelihood of a range of results of occurring

Example2:

Estimate π



Balanced information exposure estimation

Monte Carlo simulation

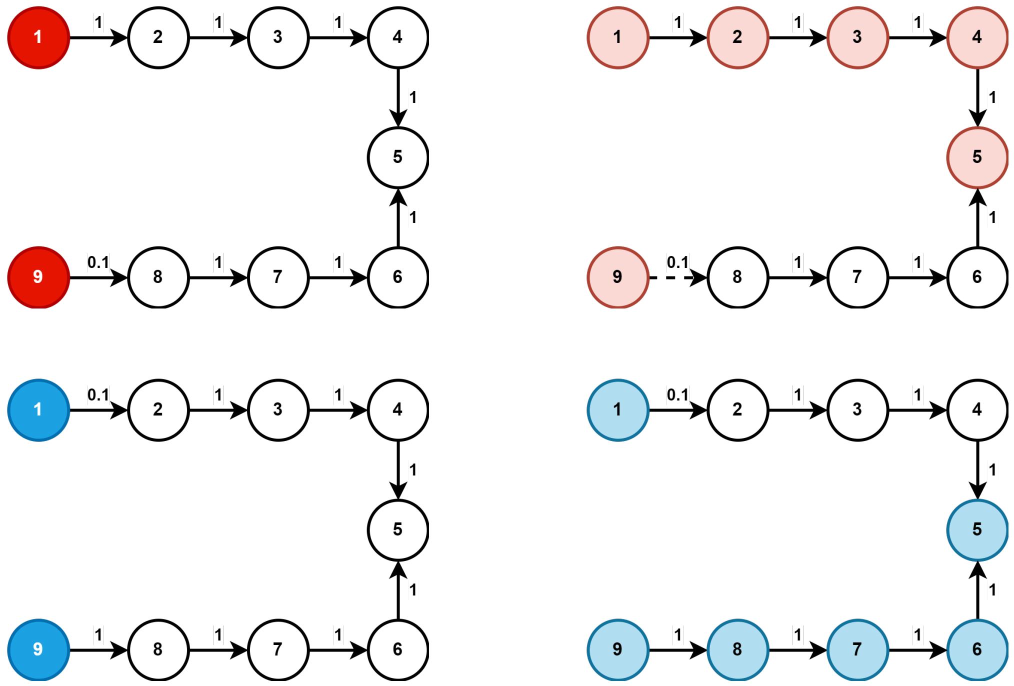
- A computational algorithm that uses **repeated random sampling** to obtain the likelihood of a range of results of occurring

Estimate balanced

information exposure:

$$\Phi_{g\sim G}(S_1, S_2) = \left|V \setminus \left(r_1(I_1 \cup S_1) \triangle r_2(I_2 \cup S_2)\right)\right|_g$$

$$= |\{1, 2, 5, 8, 9\}| = 5$$



Balanced information exposure estimation

Monte Carlo simulation

- A computational algorithm that uses **repeated random sampling** to obtain the likelihood of a range of results of occurring

Estimate balanced information exposure:

$$\max \Phi(S_1, S_2) = \max \mathbb{E}\left[\left|V \setminus \left(r_1(I_1 \cup S_1) \vartriangle r_2(I_2 \cup S_2)\right)\right|\right]$$

$$\boxed{\hat{\Phi}(S_1, S_2) = \frac{\sum_{i=1}^{N} \Phi_{g_i}(S_1, S_2)}{N}}$$

Outline

- A brief review of information exposure maximization
- An estimation method for balanced information exposure
- **A heuristic algorithm for information exposure maximization**
- Summary

Heuristic algorithm for IEM

Greedy best-first search

- Main idea: expand the node with the largest $h(v)$ value

$h(v) =$ increment to the balanced information exposure

Algorithm: Greedy best-first search

$S_1 \leftarrow S_2 \leftarrow \emptyset$;

while $|S_1| + |S_2| \leq k$ do

$v_1^* \leftarrow \arg\max_v \big(\Phi(S_1 \cup \{v\}, S_2) - \Phi(S_1, S_2)\big)$;

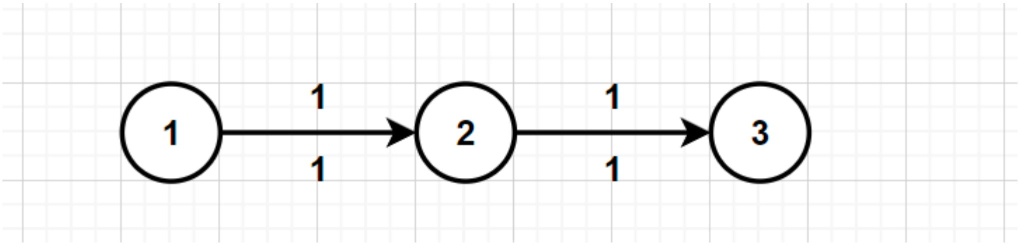
$v_2^* \leftarrow \arg\max_v \big(\Phi(S_1, S_2 \cup \{v\}) - \Phi(S_1, S_2)\big)$;

add the better option between $< v_1^*, \emptyset >$ and $< \emptyset, v_2^* >$ to $< S_1, S_2 >$ while respecting the budget.

Heuristic algorithm for IEM

Greedy best-first search

- Not optimal!



$I_1 = I_2 = \{2\}, S_1 = S_2 = \{\}$

+1, U1={1,2}, U2={2}, E1={1,2,3},E2={2,3}, Φ =2

+2, U1={2}, U2={2}, E1={2,3}, E2={2,3}, Φ =3

+3, U1={2,3}, U2={2}, E1={2,3}, E2={2,3}, Φ =2

Worse

$I_1 = I_2 = \{2\}, S_1 = \{2\}, S_2 = \{\}$

+1, U1={1,2}, U2={2}, E1={1,2,3},E2={2,3}, Φ =2

+3, U1={2,3}, U2={2}, E1={2,3}, E2={2,3}, Φ =2

Outline

- A brief review of information exposure maximization
- An estimation method for balanced information exposure
- A heuristic algorithm for information exposure maximization
- **Summary**

Summary

SUSTech Southern University of Science and Technology NICAL

- Information exposure maximization is computationally complex
- **Monte Carlo simulations** for balanced information exposure estimation
- **Greedy best-first search** to find balanced seed sets
- **Improvements in solution quality or computing efficiency are encouraged**