Project1

Information Exposure Maximization

Phase3-Evolutionary Optimization

Outline

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

Brief review of IEM

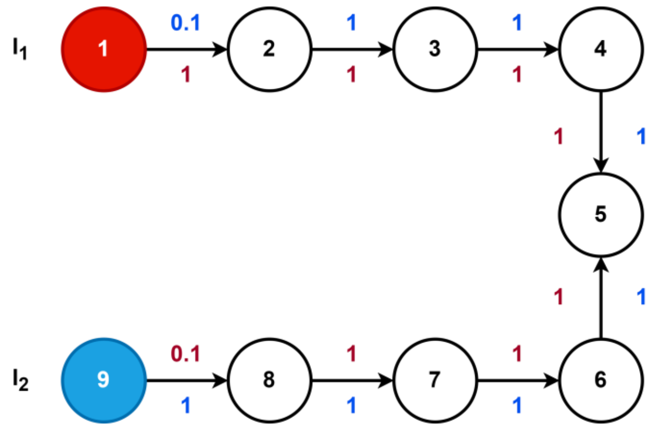
Given a social network $G = (V, E)$, two initial seed sets I_1 and I_2, and a budget k.

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.,

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

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

$$S_1, S_2 \subseteq V$$



Brief review of IEM

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

Brief review of objective function estimation

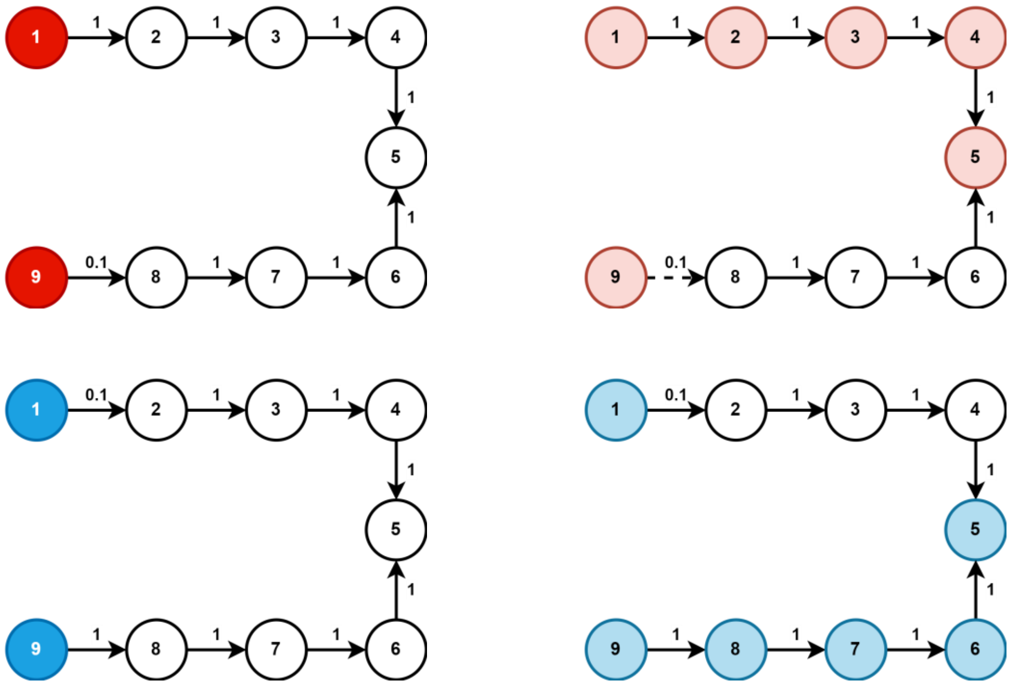
Monte Carlo simulation

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

Estimate balanced

information exposure:

$$
\begin{aligned}
&\Phi_{g\sim G}(S_1, S_2) \\
&= \left|V \setminus \left(r_1(I_1 \cup S_1) \triangle r_2(I_2 \cup S_2)\right)\right|_g \\
&\qquad = |\{1, 2, 5, 8, 9\}| = 5
\end{aligned}
$$



Brief review of objective function 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:

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

$$\Downarrow$$

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

Outline

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

Beyond Classical Search for IEM

Basic issues

- Solution representation
  - e.g. continuous, discrete (binary, integer, permutation, etc.)
- Fitness function
  - differ from the objective function
- Search method
  - e.g., simulated annealing, evolutionary algorithms, etc.

Evolutionary optimization for IEM

Solution Representation

- Binary representation

$$x = \{x_1, x_2, \cdots, x_{|V|}, x_{|V|+1}, x_{|V|+2}, \cdots, x_{|V|+|V|}\}$$

$$x_i \in \{False, True\}$$

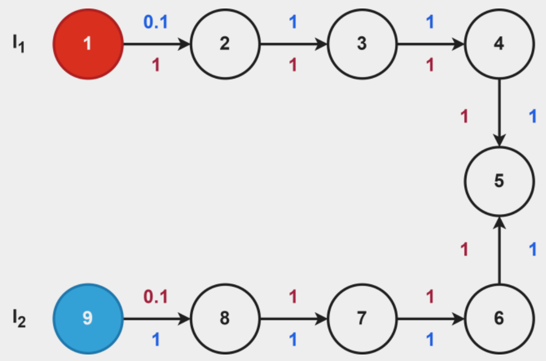
ith node is added into S_1, $i \in [1, |V|]$

ith node is added into S_2, $i \in [|V+1|, |V| + |V|]$

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

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

$$S_1, S_2 \subseteq V$$



Evolutionary optimization for IEM

Fitness Function

- Distinguish between feasible and infeasible solutions

$$fitness(S_1, S_2) = \begin{cases} \hat{\Phi}(S_1, S_2) & \text{if } |S_1| + |S_2| \le k, \\ -(|S_1| + |S_2|) & \text{otherwise.} \end{cases}$$

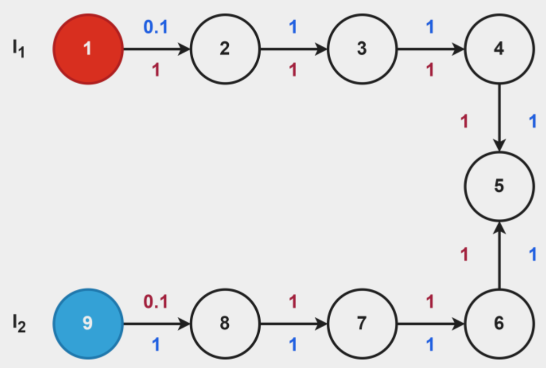
Punish infeasible solutions

according to the degree of violation

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

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

$$S_1, S_2 \subseteq V$$



Evolutionary optimization for IEM

Fitness Function (differs from objective function)

- Distinguish between feasible and infeasible solutions

$$
fitness(x) = \begin{cases} \hat{\Phi}(x) & \text{if } \sum x \leq k, \\ -\sum x & \text{otherwise.} \end{cases}
$$

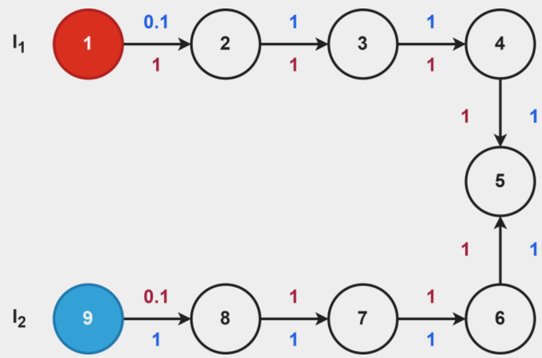
$$
x = \{x_1, x_2, \cdots, x_{|V|}, x_{|V|+1}, x_{|V|+2}, \cdots, x_{|V|+|V|}\}
$$

$$
x_i \in \{False, True\}
$$

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

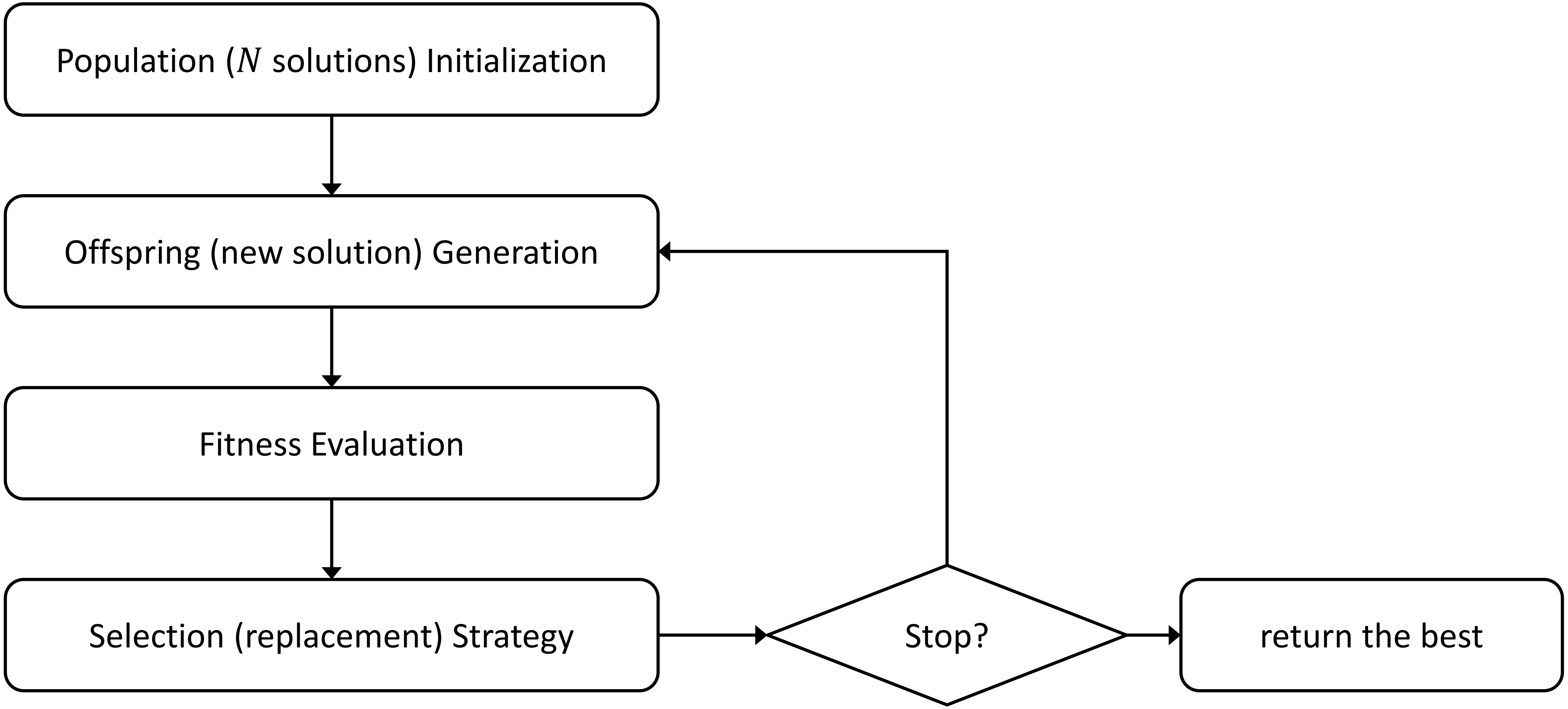
$$
\text{s.t.} \quad |S_1| + |S_2| \leq k
$$

$$
S_1, S_2 \subseteq V
$$



Evolutionary optimization for IEM

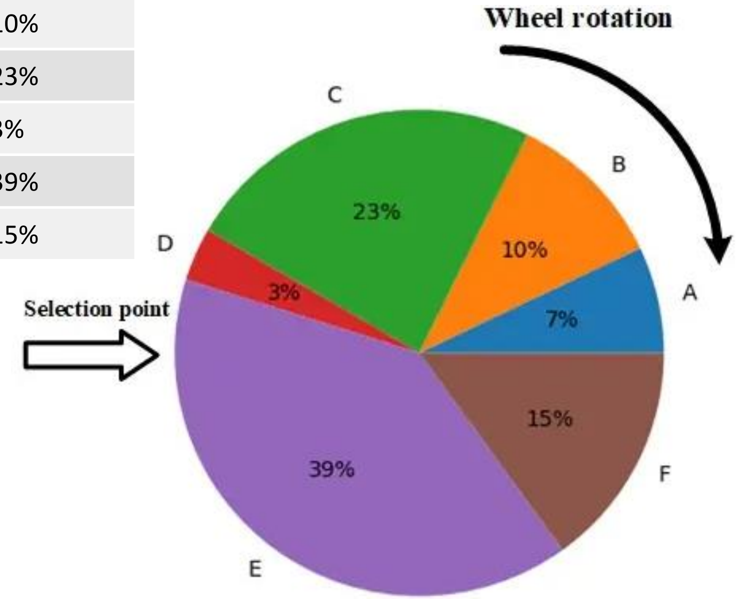
General Evolutionary Optimization Framework



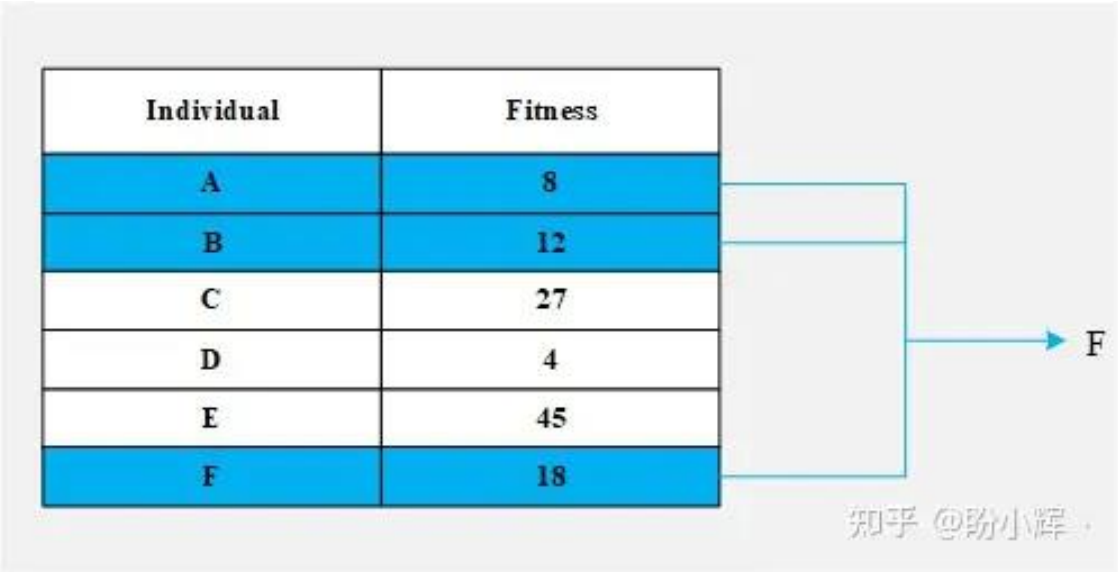
Evolutionary optimization for IEM

Parent selection

Solution	Fitness	proportion
A	8	7%
B	12	10%
C	27	23%
D	4	3%
E	45	39%
F	17	15%



roulette wheel selection



Tournament selection

Evolutionary optimization for IEM

Crossover

单点交叉（Single-point crossover）

parent1	1	2	3	4	5	6	7	8
parent2	a	b	c	d	e	f	g	h

两点交叉（Two-points crossover）

parent1	1	2	3	4	5	6	7	8
parent2	a	b	c	d	e	f	g	h

多点交叉（Multi-point crossover）

parent1	1	2	3	4	5	6	7	8
parent2	a	b	c	d	e	f	g	h

均匀交叉（Uniform crossover）

parent1	1	2	3	4	5	6	7	8
parent2	a	b	c	d	e	f	g	h

顺序交叉（Order crossover, OX）

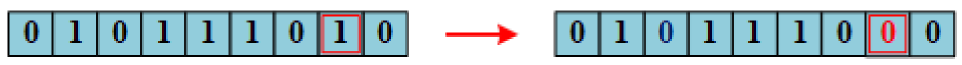
parent1	1	2	3	4	5	6	7	8
parent2	3	5	8	1	7	4	2	6

位置交叉（Position-based crossover , PBX）

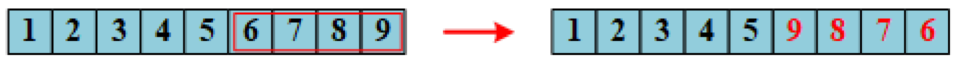
parent1	1	2	3	4	5	6	7	8
parent2	3	5	8	1	7	4	2	6

Evolutionary optimization for IEM

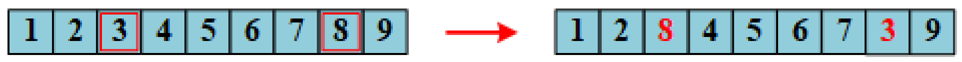
Mutation



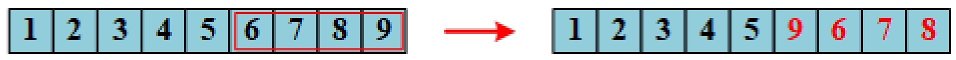
Flip bit mutation



Inversion mutation



Swap\Exchange mutation



Scramble mutation

Evolutionary optimization for IEM

Selection strategy

Elite selection	The larger the fitness, the better
Non-Elite selection	Not entirely dependent on the fitness

Evolutionary optimization for IEM

An Evolutionary Algorithm for IEM

$$x = \{x_1, \cdots, x_{|V|}, x_{|V|+1}, \cdots, x_{|V|+|V|}\}$$

$$x_i \in \{False, True\}$$

$$\max f(x) = \begin{cases} \hat{\Phi}(x) & \text{if } \sum x \leq k, \\ -\sum x & \text{otherwise.} \end{cases}$$

Generate a population P_t with N solutions randomly

↓

Generate N new solutions By conducting the two-points crossover and bit-flip mutation on solutions from P_t based on the binary tournament selection

↓

Fitness Evaluation

↓

Sort solutions based on f in descending order and reserve the first N solutions as the next population P_{t+1}

→ Stop? → return the best

(Stop? → back to "Generate N new solutions")

Outline

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

Summary

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