Project 1

Information Exposure Maximization

Outline

- Overview
- Problem Description
- IEM Example

Outline

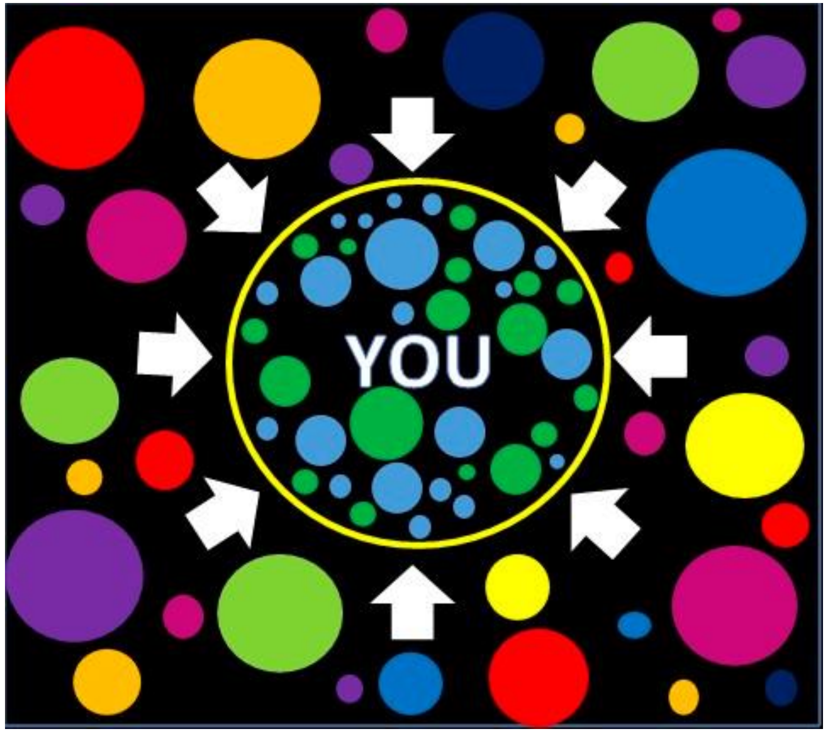
- **Overview**

- Problem Description

- IEM Example

Overview

Information Exposure Maximization (IEM) is an important algorithmic problem that is proposed to solve the echo chamber effect on social media.



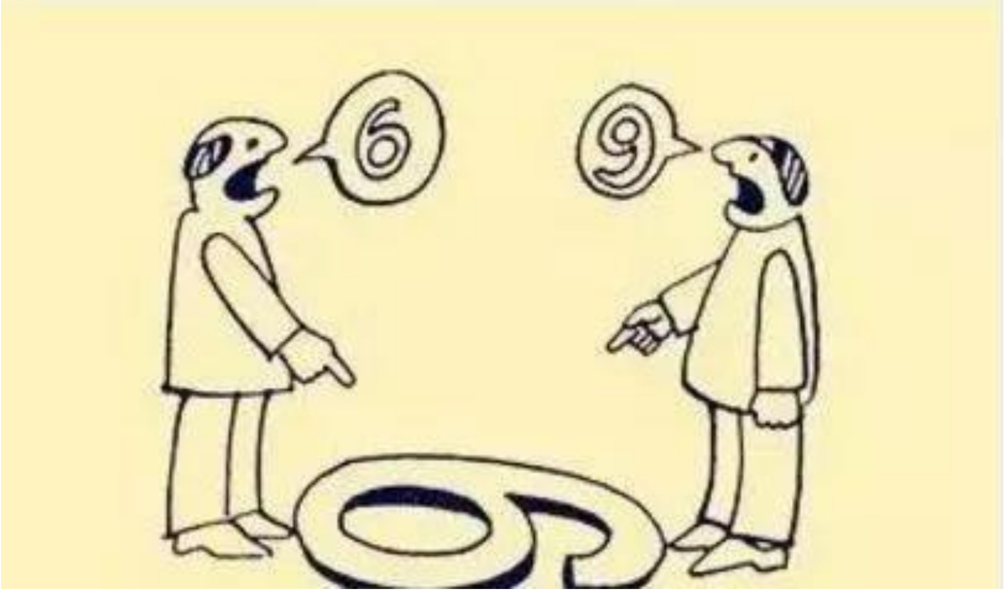
Users tend to gather in groups whose members think alike, and that polarisation is greater when content feeds cannot be easily tweaked.

[M. Cinelli et al. PNAS 2021]

Overview

Assume that there are two viewpoints in a social network,

the IEM problem is **to select two campaigns**, each containing a set of users who hold one of these viewpoints, to maximize the expected number of users that are **either reached by both campaigns** or **remain oblivious to both campaigns**.



Overview

TASK: **Two search algorithms** to solve the IEM problem

- One heuristic algorithm
- One evolutionary algorithm or one simulated annealing algorithm

Grading Rules

- Project report
- Code evaluation (15 points)
  - Objective evaluation (2.0 points)
  - Heuristic algorithm (6.5 points)
  - Evolutionary algorithm or simulated annealing algorithm (6.5 points)

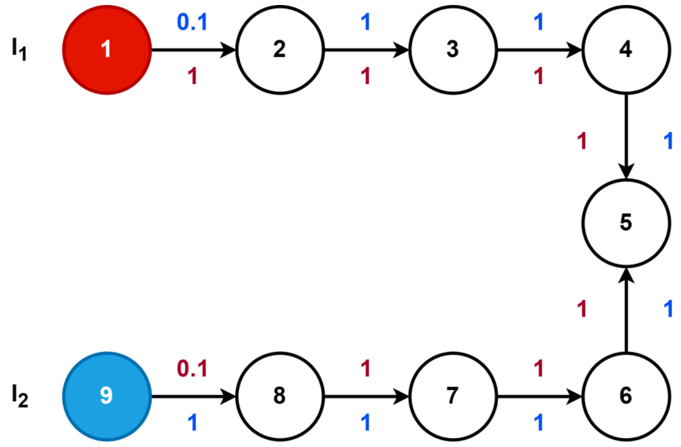
Outline

- Overview
- **Problem Description**
- IEM Example

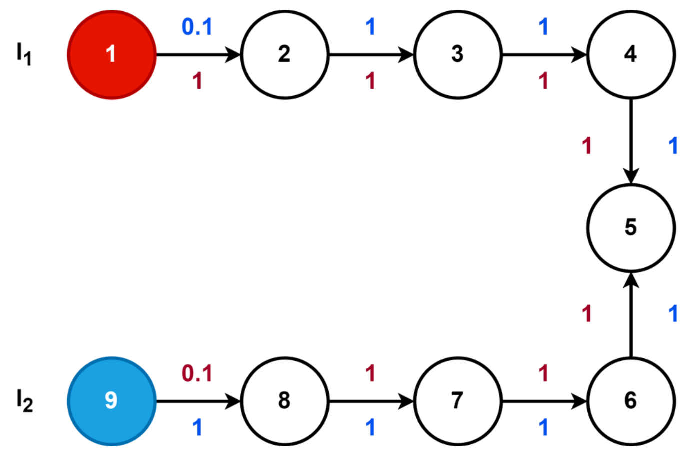
Assumption

The following information is available

- A social network $G = (V, E)$ with two campaigns
- Two seed sets
  - each support one of the two viewpoints



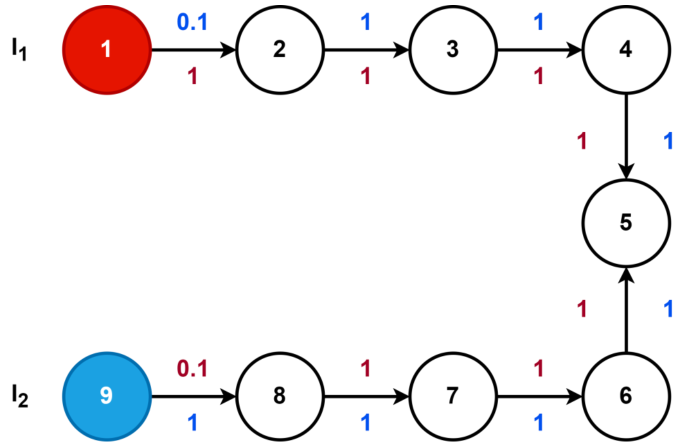
Assumption



The following information is available

- A social network $G = (V, E)$ with two campaigns
- Two seed sets
  - each support one of the two viewpoints
- Heterogeneous propagation
  - two viewpoints are propagated independently of each other
- Two functions $p: E \rightarrow [0,1]$
  - associate probability $p_{(u,v)}$ with edge (u, v) capturing the influence u exerts over v

Assumption



The following information is available

- A social network $G = (V, E)$ with two campaigns
- Two seed sets
  - each support one of the two viewpoints
- Heterogeneous propagation
  - two viewpoints are propagated independently of each other
- Two functions $p: E \rightarrow [0,1]$
  - associate probability $p_{(u,v)}$ with edge (u, v) capturing the influence u exerts over v
- A diffusion model
  - describe how influence propagates to a node from its neighbors

Influence Diffusion of Each Campaign

The famous Independent Cascade (IC) model is used.

- Each node $v \in V$ has two possible states, ***inactive*** and ***active***
  - *active*: adopts new information being propagated through the network
  - *inactive*: has not adopted new information yet
    - Ever been attempted to be activated but NOT
    - Never been attempted to be activated
  - Can ONLY switch from *inactive* to *active*

Influence Diffusion of Each Campaign

The famous Independent Cascade (IC) model is used.

- Each node $v \in V$ has two possible states, ***inactive*** and ***active***
  - *active*: adopts new information being propagated through the network
  - *inactive*: has not adopted new information yet
    - Ever been attempted to be activated but NOT
    - Never been attempted to be activated
  - Can ONLY switch from *inactive* to *active*

Exposed Node Set

Influence Diffusion of Each Campaign

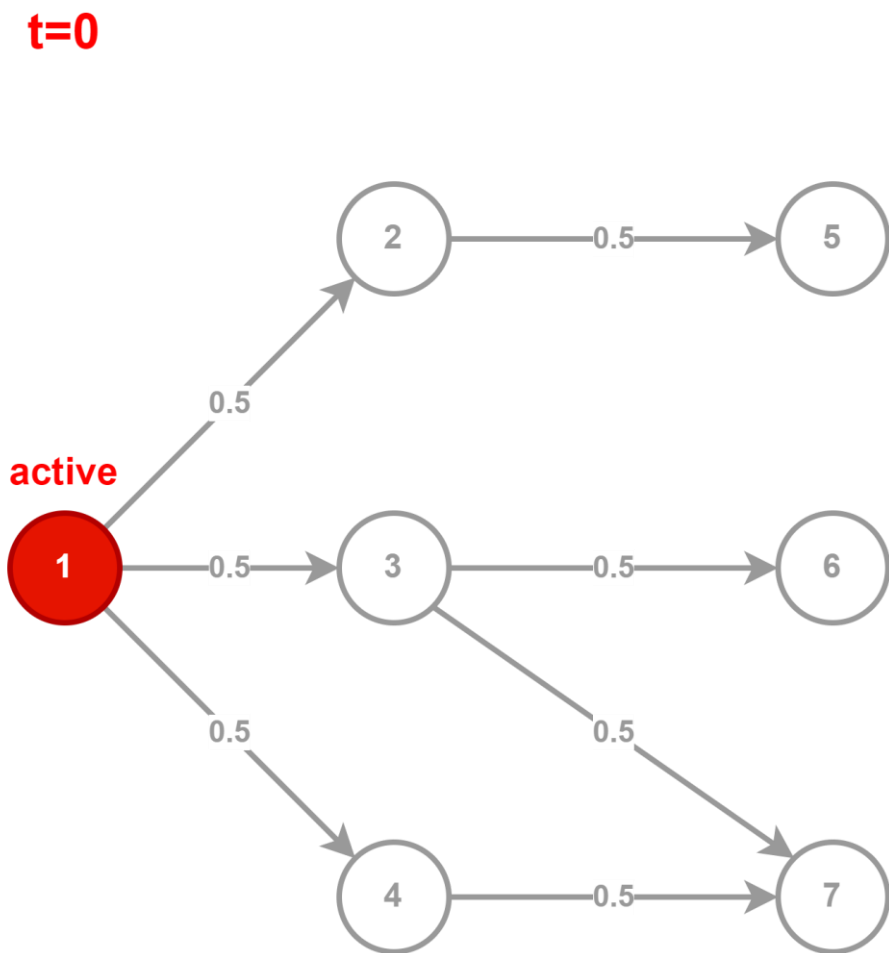
Information proceeds in discrete time steps, with $t = 0,1,2,\cdots$

- Let U be the set of **active** nodes in time 0
- Let $r(U)$ be the **FINAL exposed node set**

Influence Diffusion of Each Campaign (Case)

- Seed set $U = \{1\}$

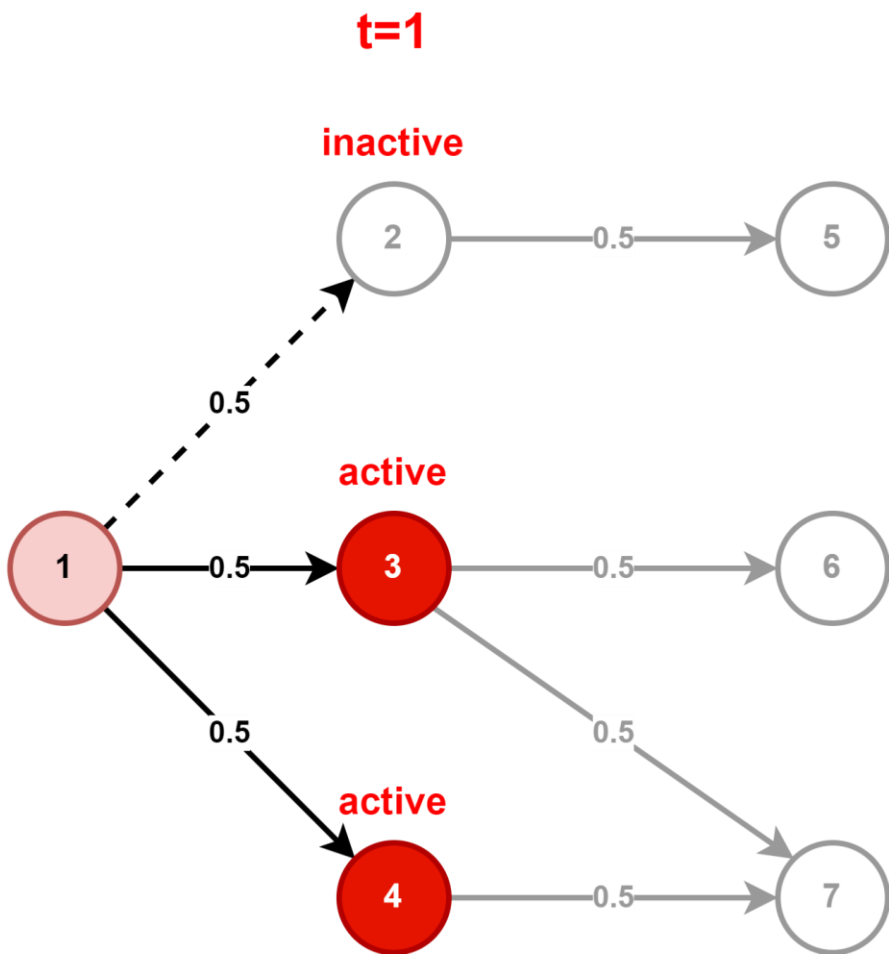
- Time step $t = 0$
  - Active node set $\{1\}$
  - Exposed node set $\{1\}$



Influence Diffusion of Each Campaign (Case)

- Seed set $U = \{1\}$

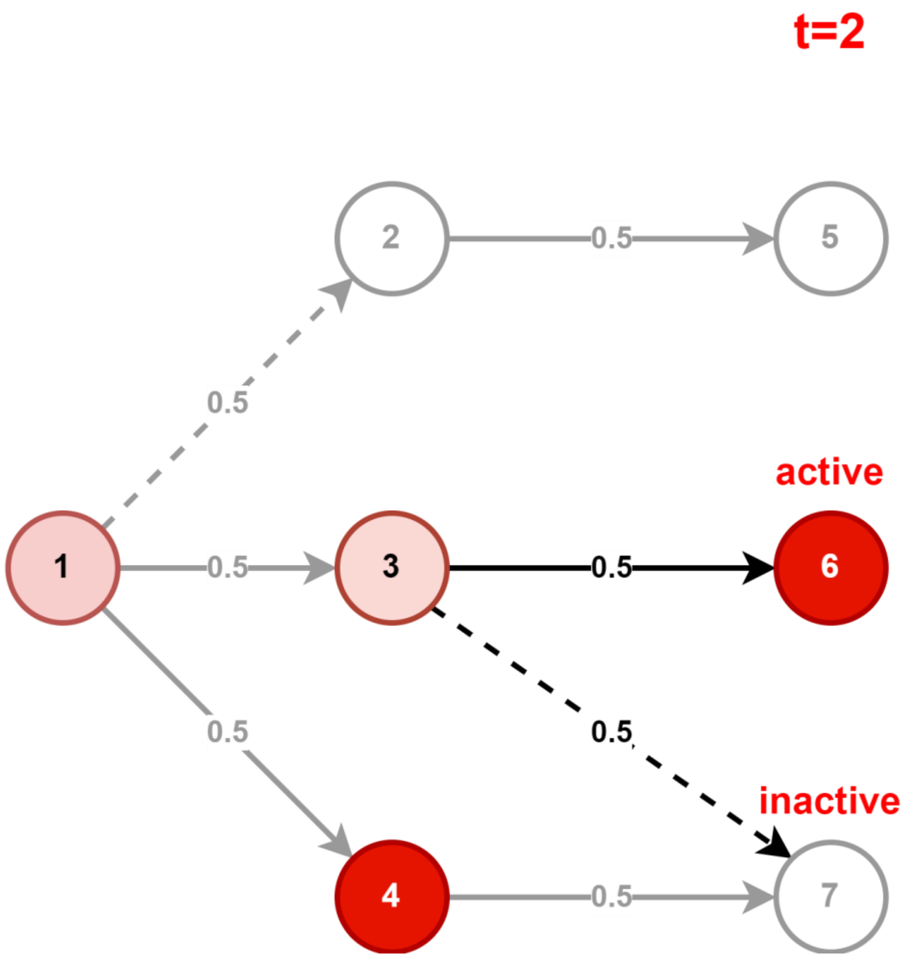
- Time step $t = 1$
  - Active node set $\{1, 3, 4\}$
  - Exposed node set $\{1, 2, 3, 4\}$



Influence Diffusion of Each Campaign (Case)

- Seed set $U = \{1\}$

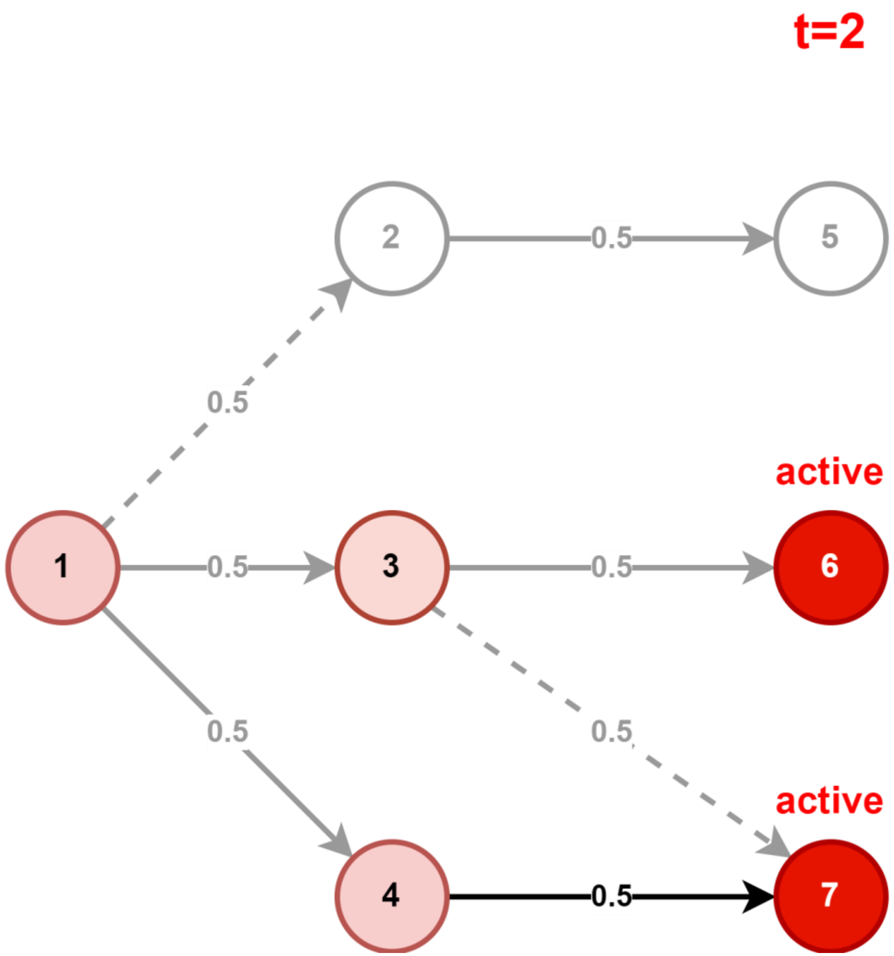
- Time step $t = 2$
  - Active node set $\{1, 3, 4, 6\}$
  - Exposed node set $\{1, 2, 3, 4, 6, 7\}$



Influence Diffusion of Each Campaign (Case)

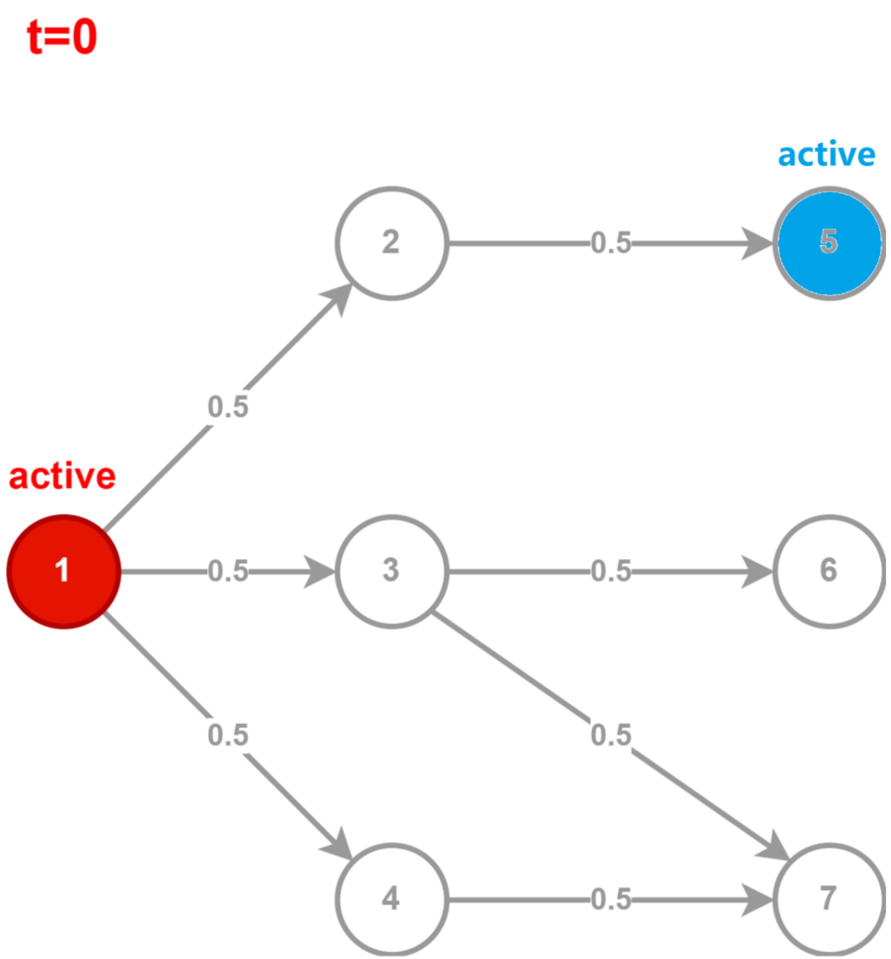
- Seed set $U = \{1\}$

- Time step $t = 2$
  - Active node set $\{1, 3, 4, 6, 7\}$
  - Exposed node set $\{1, 2, 3, 4, 6, 7\}$



Influence Exposure of Two Campaigns (Case)

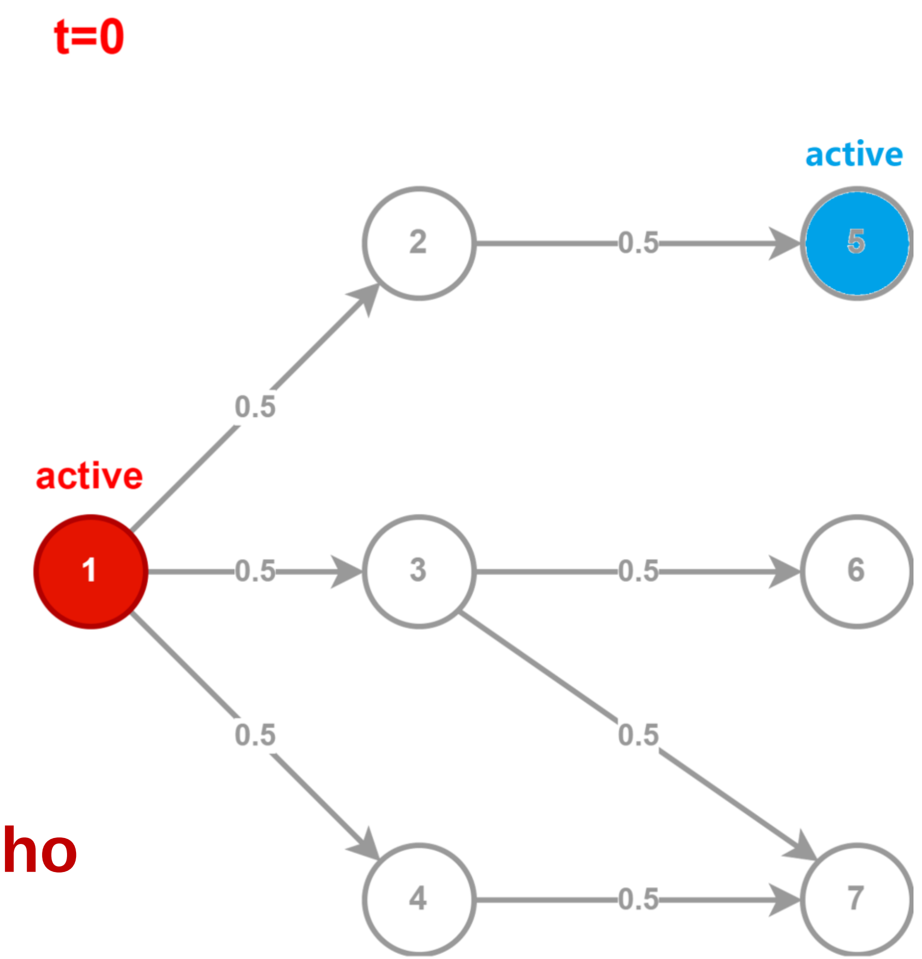
- Seed set $U_1 = \{1\}$
  - Exposed node set $\{1, 2, 3, 4, 6, 7\}$
- Seed set $U_2 = \{5\}$
  - Exposed node set $\{5\}$



Influence Exposure of Two Campaigns (Case)

- Seed set $U_1 = \{1\}$
  - Exposed node set $\{1, 2, 3, 4, 6, 7\}$
- Seed set $U_2 = \{5\}$
  - Exposed node set $\{5\}$

NO node can break through the echo chamber in this propagation



Balanced Seed Set

Campaign c_i, $i \in \{1,2\}$

- Seed set $U_i = I_i \cup S_i$
  - Initial seed set $I_i \subseteq V$: available
  - Balanced seed set $S_i \subseteq V$: **to be found**

Balanced Information Exposure

The expected number of nodes that are either **reached by both campaigns** or **remain oblivious to both campaigns**

$$A \triangle B = (A \setminus B) \cup (B \setminus A)$$

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

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

Problem Description

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 $\boldsymbol{S_1}$ and $\boldsymbol{S_2}$**, where $|S_1| + |S_2| \leq k$, and **maximize the balanced information exposure**, i.e.,

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

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

$$S_1, S_2 \subseteq V$$

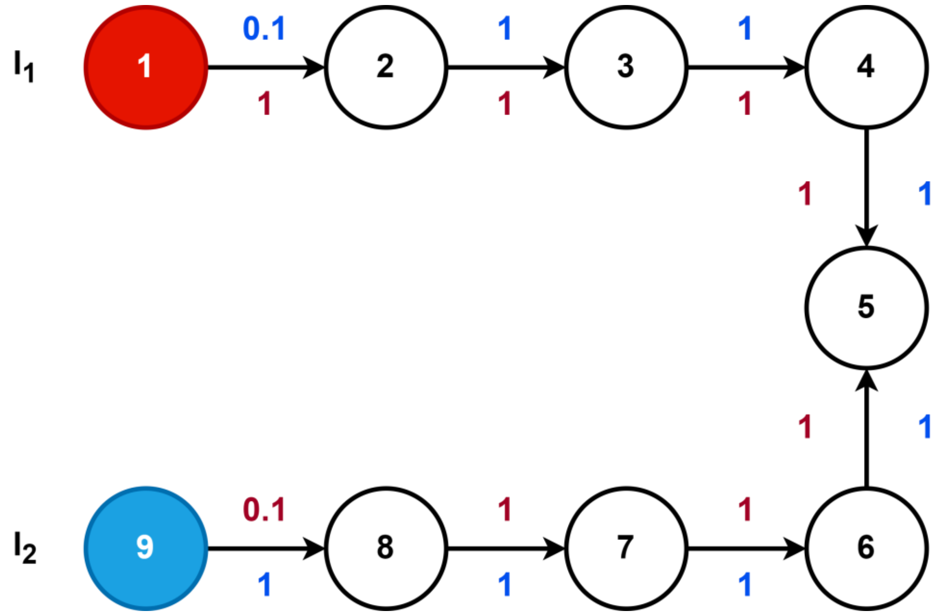
Outline

- Overview
- Problem Description
- **IEM Example**

IEM Example

Campaigns c_1 (Red), c_2 (blue)

- Initial Seed Set $I_1 = \{1\}, I_2 = \{9\}$
- Balanced Seed Set:
  - Solution 1: $S_1 = \{8\}, S_2 = \{2\}$
  - Solution 2: $S_1 = \{9\}, S_2 = \{1\}$

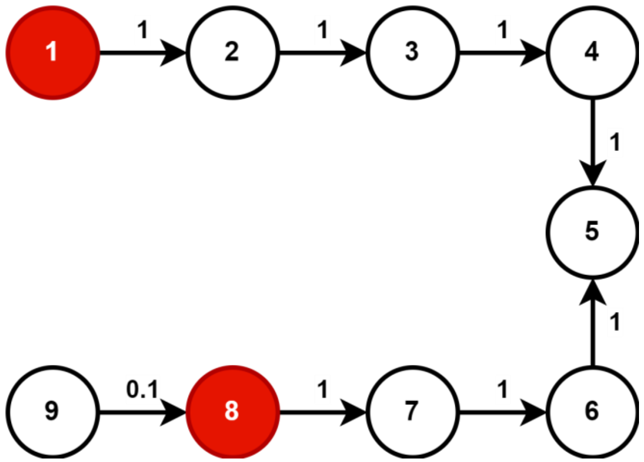


IEM Solution 1

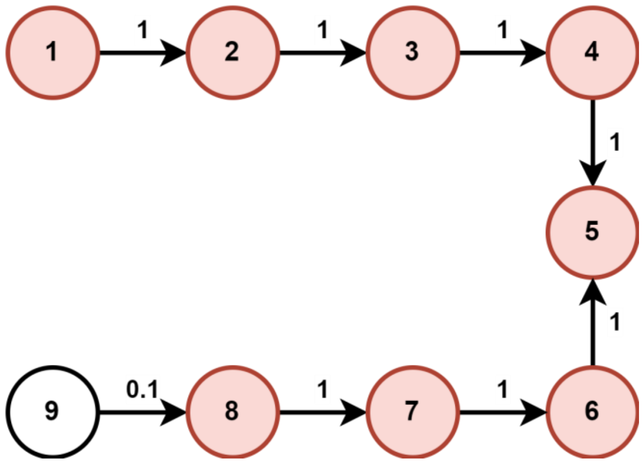
Campaigns c_1

- Initial Seed Set $I_1 = \{1\}$
- Balanced Seed Set $S_1 = \{8\}$

  - Exposed Node Set $\{1, 2, 3, 4, 5, 6, 7, 8\}$



$t = 0$



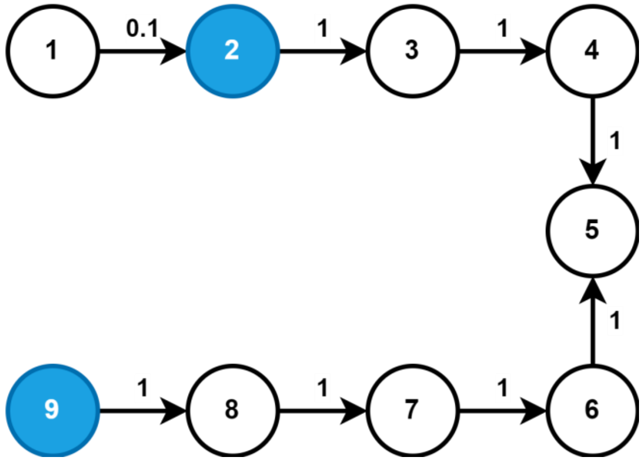
$t = 4$

IEM Solution 1

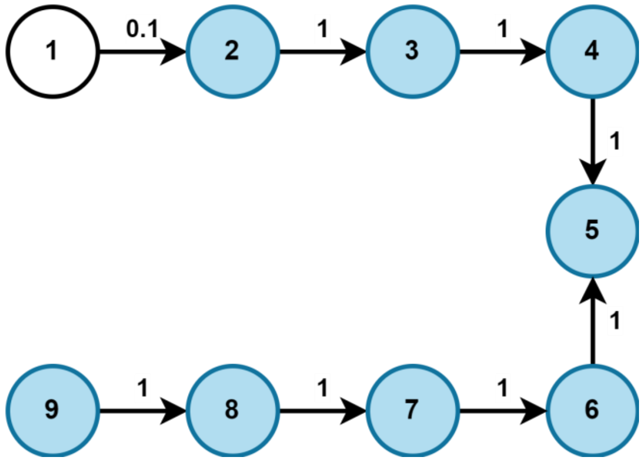
Campaigns c_2

- Initial Seed Set $I_2 = \{9\}$
- Balanced Seed Set $S_2 = \{2\}$

  - Exposed Node Set $\{2, 3, 4, 5, 6, 7, 8, 9\}$



$t = 0$



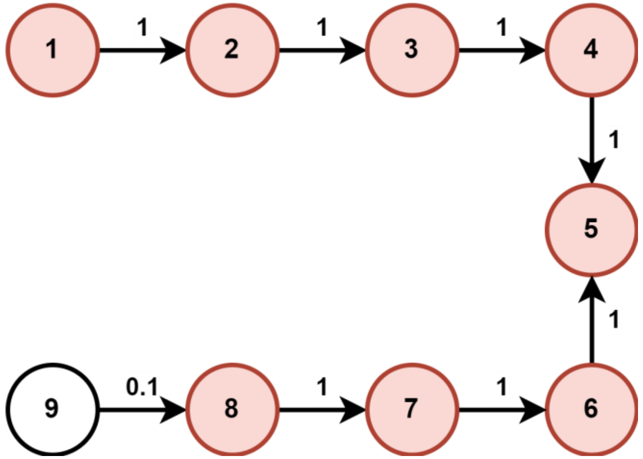
$t = 4$

IEM Solution 1

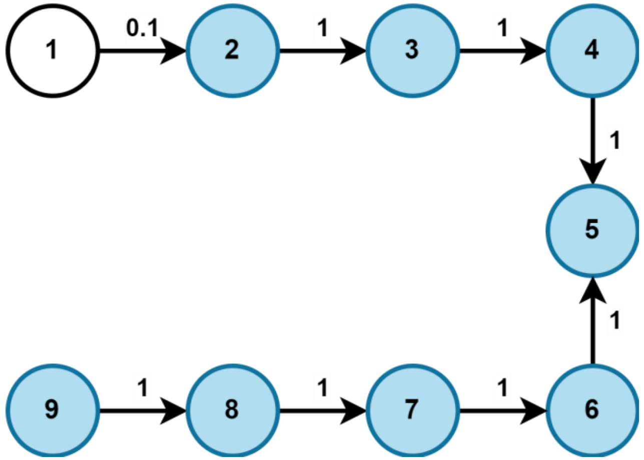
Balanced Exposed Node

- Initial Seed Set $I_1 = \{1\}, I_2 = \{9\}$
- Balanced Seed Set $S_1 = \{8\}, S_2 = \{2\}$

- $\Phi(S_1, S_2) = \left|V \setminus \left(r_1(I_1 \cup S_1) \triangle r_2(I_2 \cup S_2)\right)\right| = |\{2, 3, 4, 5, 6, 7, 8\}| = 7$



$\boldsymbol{r_1 = \{1, 2, 3, 4, 5, 6, 7, 8\}}$

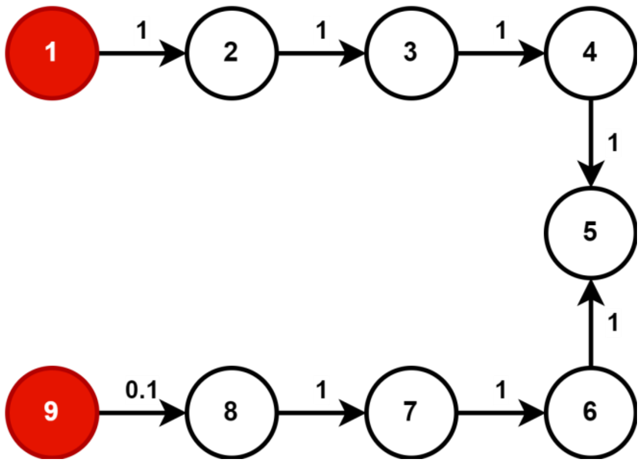


$\boldsymbol{r_2 = \{2, 3, 4, 5, 6, 7, 8, 9\}}$

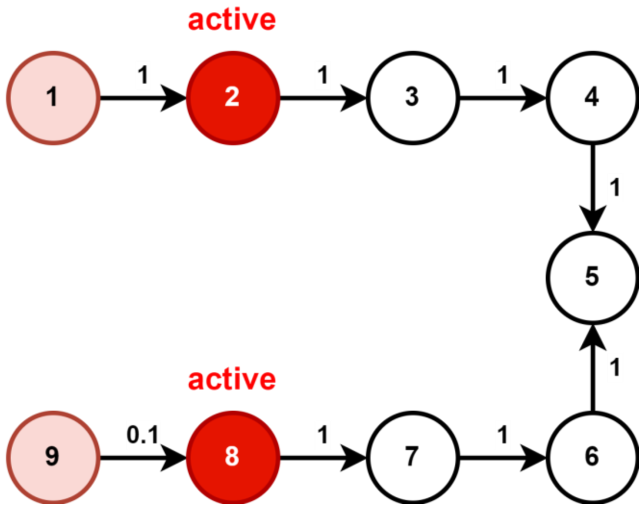
IEM Solution 2

Campaigns c_1

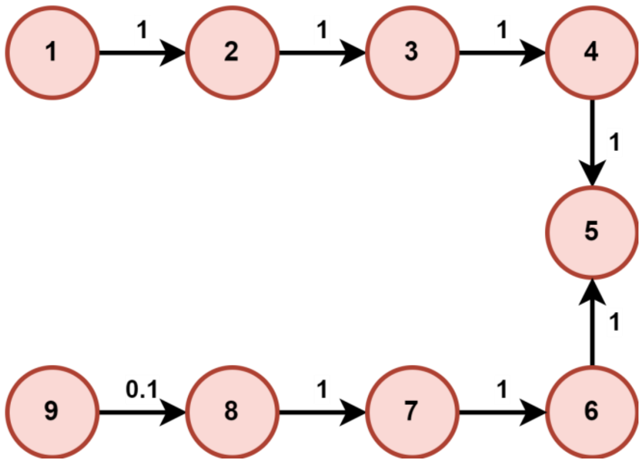
- Initial Seed Set $I_1 = \{1\}$
- Balanced Seed Set $S_1 = \{9\}$
- Case 1: node 8 activate with probability 0.1
    - Exposed Node Set $\{1, 2, 3, 4, 5, 6, 7, 8, 9\}$



$t = 0$



$t = 1$

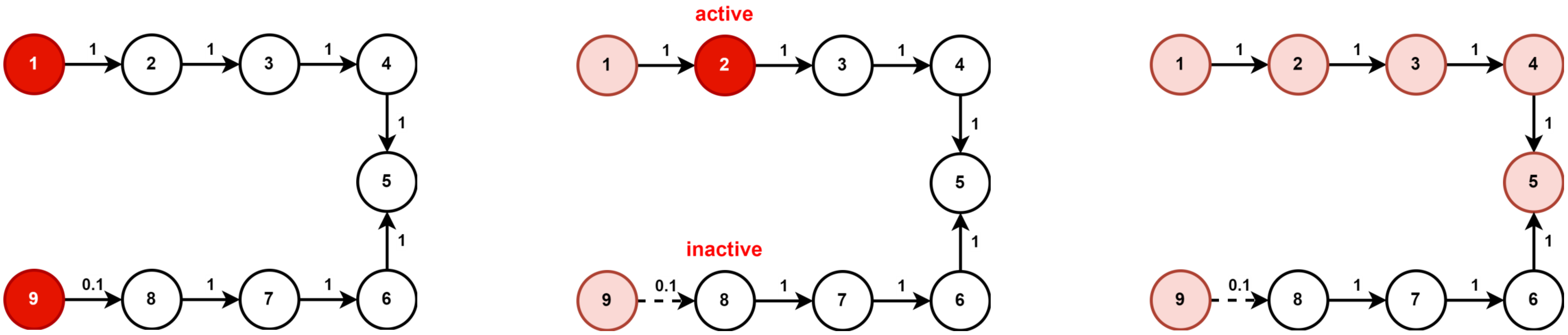


$t = 4$

IEM Solution 2

Campaigns c_1

- Initial Seed Set $I_1 = \{1\}$
- Balanced Seed Set $S_1 = \{9\}$
- Case 2: node 8 inactivate with probability 0.9
    - Exposed Node Set $\{1, 2, 3, 4, 5, 8, 9\}$

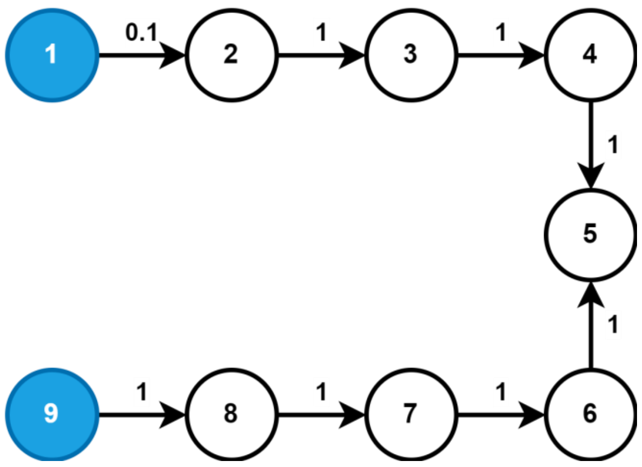


$t = 0$ $t = 1$ $t = 4$

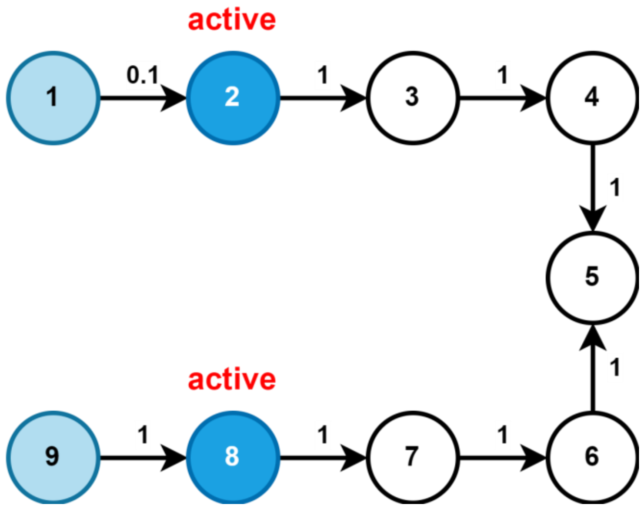
IEM Solution 2

Campaigns c_2

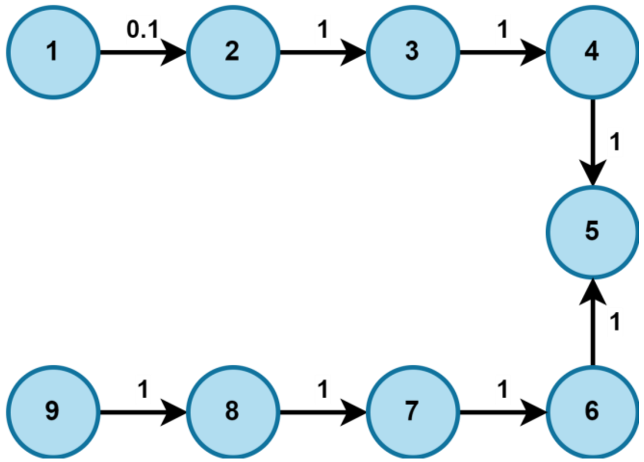
- Initial Seed Set $I_2 = \{9\}$
- Balanced Seed Set $S_2 = \{1\}$
- Case 1: node 2 activate with probability 0.1
  - Exposed Node Set $\{1, 2, 3, 4, 5, 6, 7, 8, 9\}$



$t = 0$



$t = 1$

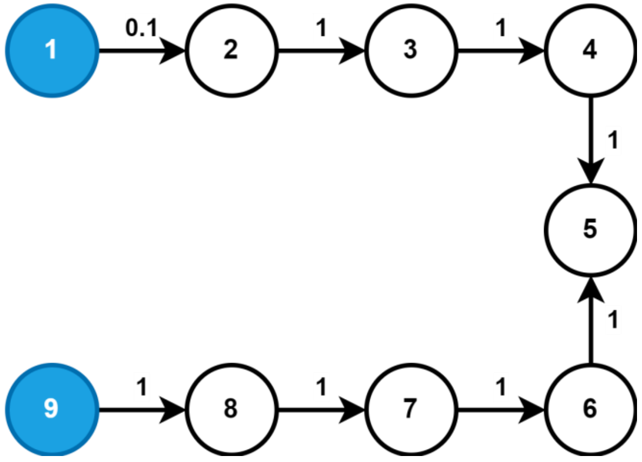


$t = 4$

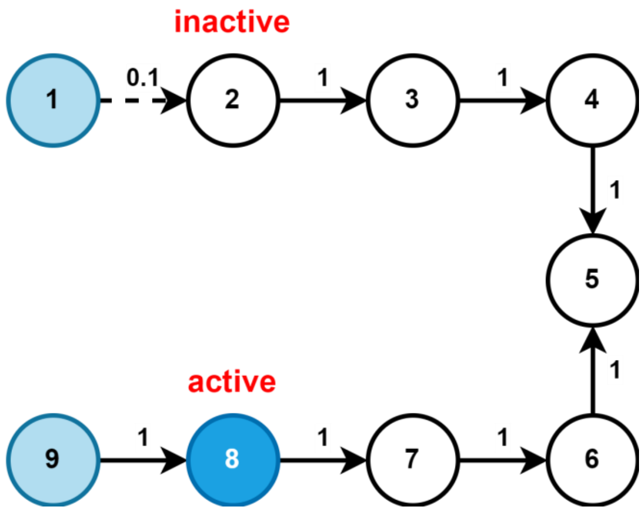
IEM Solution 2

Campaigns c_2

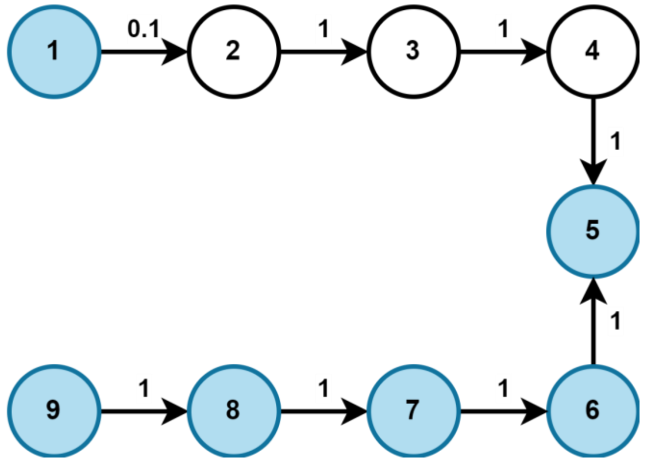
- Initial Seed Set $I_2 = \{9\}$
- Balanced Seed Set $S_2 = \{1\}$
- Case 2: node 2 inactivate with probability 0.9
  - Exposed Node Set $\{1, 2, 5, 6, 7, 8, 9\}$



$t = 0$



$t = 1$

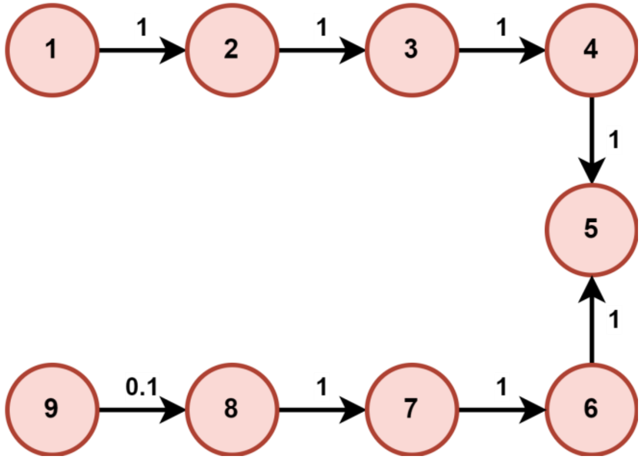


$t = 4$

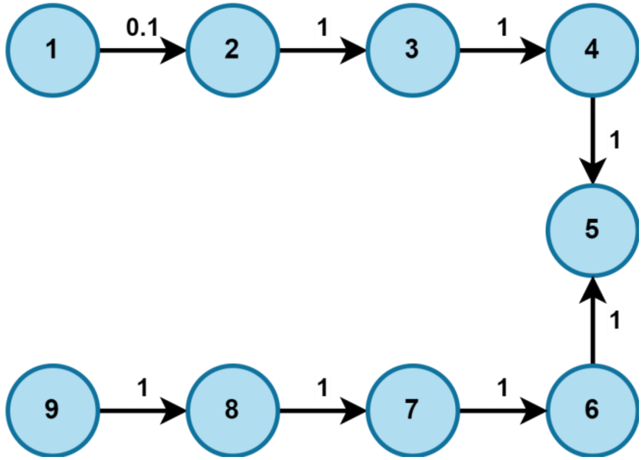
IEM Solution 2

Balanced Exposed Node

- Initial Seed Set $I_1 = \{1\}, I_2 = \{9\}$
- Balanced Seed Set $S_1 = \{9\}, S_2 = \{1\}$
  - Probability $p^{(1)} = p_1 \times p_2 = 0.01$
  - $\Phi^{(1)}(S_1, S_2) = \left|V \setminus \left(r_1(I_1 \cup S_1) \triangle r_2(I_2 \cup S_2)\right)\right| = |\{1,2,3,4,5,6,7,8,9\}| = 9$



$\boldsymbol{p_1 = 0.1, r_1 = \{1,2,3,4,5,6,7,8,9\}}$

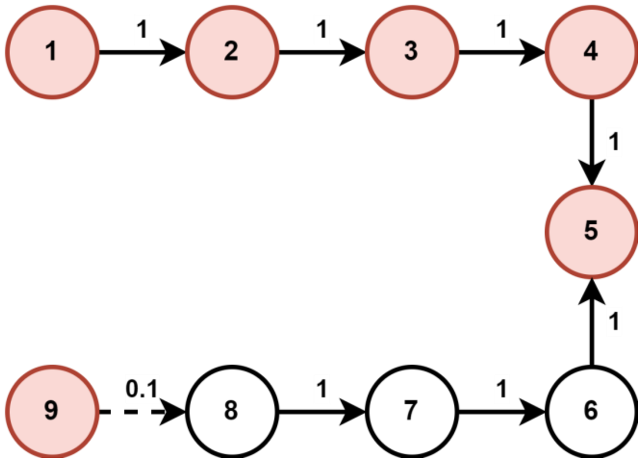


$\boldsymbol{p_2 = 0.1, r_2 = \{1,2,3,4,5,6,7,8,9\}}$

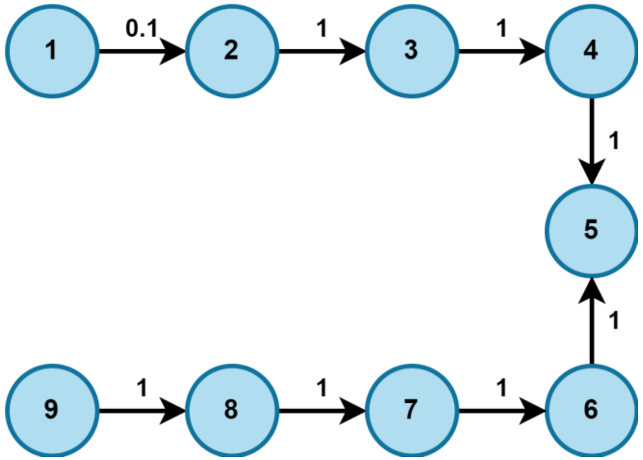
IEM Solution 2

Balanced Exposed Node

- Initial Seed Set $I_1 = \{1\}, I_2 = \{9\}$
- Balanced Seed Set $S_1 = \{9\}, S_2 = \{1\}$
  - Probability $p^{(2)} = p_1 \times p_2 = 0.09$
  - $\Phi^{(2)}(S_1, S_2) = \left|V \setminus \left(r_1(I_1 \cup S_1) \triangle r_2(I_2 \cup S_2)\right)\right| = |\{1,2,3,4,5,8,9\}| = 7$



$\boldsymbol{p_1 = 0.9, r_1 = \{1,2,3,4,5,8,9\}}$

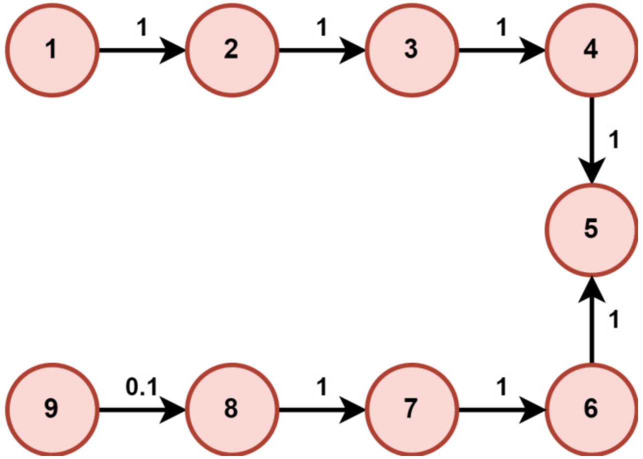


$\boldsymbol{p_2 = 0.1, r_2 = \{1,2,3,4,5,6,7,8,9\}}$

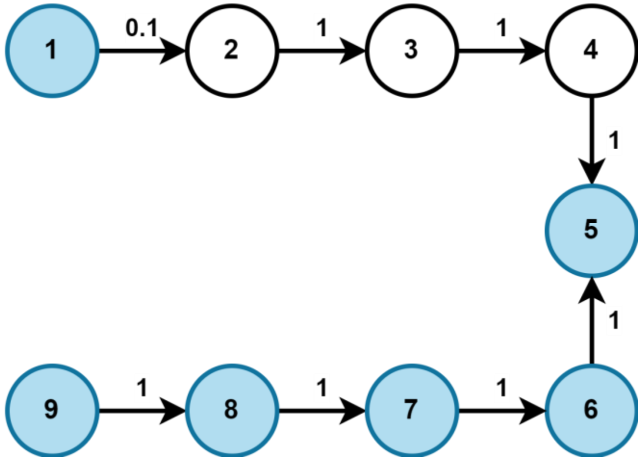
IEM Solution 2

Balanced Exposed Node

- Initial Seed Set $I_1 = \{1\}, I_2 = \{9\}$
- Balanced Seed Set $S_1 = \{9\}, S_2 = \{1\}$
  - Probability $p^{(3)} = p_1 \times p_2 = 0.09$
  - $\Phi^{(3)}(S_1, S_2) = \left|V \setminus \left(r_1(I_1 \cup S_1) \triangle r_2(I_2 \cup S_2)\right)\right| = |\{1,2,5,6,7,8,9\}| = 7$



$\boldsymbol{p_1 = 0.1, r_1 = \{1,2,3,4,5,6,7,8,9\}}$

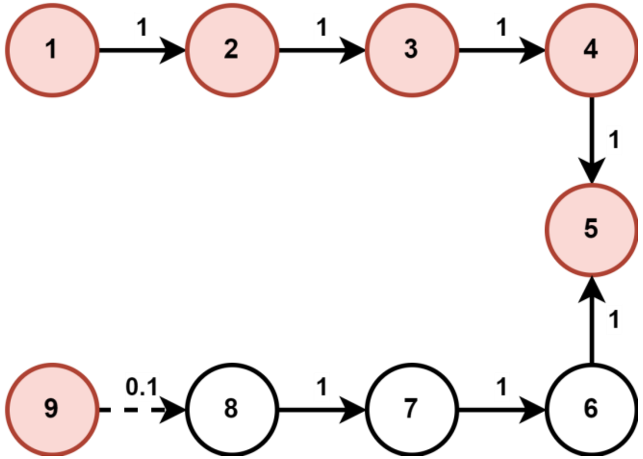


$\boldsymbol{p_2 = 0.9, r_2 = \{1,2,5,6,7,8,9\}}$

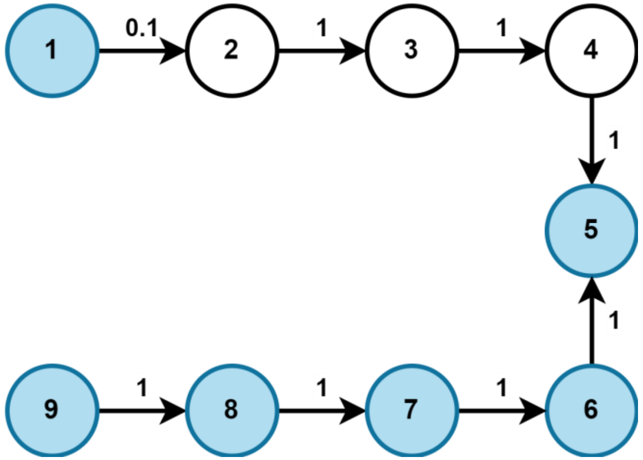
IEM Solution 2

Balanced Exposed Node

- Initial Seed Set $I_1 = \{1\}, I_2 = \{9\}$
- Balanced Seed Set $S_1 = \{9\}, S_2 = \{1\}$
  - Probability $p^{(4)} = p_1 \times p_2 = 0.81$
  - $\Phi^{(4)}(S_1, S_2) = \left|V \setminus \left(r_1(I_1 \cup S_1) \triangle r_2(I_2 \cup S_2)\right)\right| = |\{1,2,5,8,9\}| = 5$



$\boldsymbol{p_1 = 0.9, r_1 = \{1,2,3,4,5,8,9\}}$



$\boldsymbol{p_2 = 0.9, r_2 = \{1,2,5,6,7,8,9\}}$

IEM Solution 2

Balanced Exposed Node

- Initial Seed Set $I_1 = \{1\}, I_2 = \{9\}$
- Balanced Seed Set $S_1 = \{9\}, S_2 = \{1\}$

$$\Phi(S_1, S_2) = \mathbb{E}\left[\left|V \setminus \left(r_1(I_1 \cup S_1) \triangle r_2(I_2 \cup S_2)\right)\right|\right] = \sum_{i=1}^{4} p^{(i)} \times \Phi^{(i)}(S_1, S_2)$$

$$= 0.01 \times 9 + 0.09 \times 7 + 0.09 \times 7 + 0.81 \times 5 = 5.4$$

Project1: Information Exposure Maximization

- Information Exposure Maximization is computationally complex
  - Computing the balanced information exposure for a given solution is NP-hard.
  - Finding an optimal solution of IEM is NP-hard.
- Three tasks
  - An objective estimator
  - Two search algorithms

Thank you!

For more information, please refer to **Project1.PDF**