Influence Maximization

Overview

- Influence Maximization Problem (IMP) is the problem of finding a small subset of nodes (referred to as seed set) in a social network that could maximize the spread of influence.

- The influence spread is the expected number of nodes that are influenced by the nodes in the seed set in a cascade manner.

Problem Description

- Social network and influence spread
  - A social network is modeled as a directed graph $G = (V, E)$ with nodes in V modeling the individual in the network and each edge $(u, v) \in E$ is associated with a weight $w(u, v) \in [0,1]$ which indicates the probability that u influences v.
  - Let $S \subseteq V$ to be the subset of nodes selected to initiate the influence diffusion, which is called the seed set.
  - The stochastic diffusion models specify the random process of influence cascade from S, of which the output is a random set of nodes influenced by S. The expected number of influenced nodes $\sigma(S)$ is the influence spread of S.
- Goal of IMP
  - To find a seed set S that maximizes $\sigma(S)$, subject to $|S| = k$.

Problem Description

- Inputs:
  - An directed graph $G = (V, E)$
  - A predefined seed set cardinality k
  - A predefined stochastic diffusion model
- Output:
  - A size-k seed set S^* with the maximal $\sigma(S)$ for any size-k seed set $S \subseteq V$

Stochastic Diffusion Models

- Given the seed set S, the diffusion process unfolds in discrete rounds
  - At round 0, all nodes in S are active and the others are inactive.
  - In the subsequent rounds, the newly activated nodes will try to activate their neighbors.
  - Once a node becomes active, it will remain active till the end.
  - The process stops when no more nodes get activated.
- Two fundamental diffusion models are specified in this project.
  - Linear Threshold (LT)
  - Independent Cascade (IC)

Stochastic Diffusion Models

- Independent Cascade (IC) model
    - When a node u gets activated, initially or by another node, it has a single chance to activate each inactive neighbor v with the probability proportional to the edge weight $w(u, v)$.
    - Afterwards, the activated nodes remain its active state but they have no contribution in later activations.
    - The weight of the edge (u, v) is calculated as $w(u, v) = \frac{1}{d_{in}(v)}$, where $d_{in}(v)$ denotes the in-degree of node v.

Stochastic Diffusion Models

- Linear Threshold (LT) model
  - At the beginning, each node v selects a random threshold θ_v uniformly at random in range [0,1].
  - If round $t \geq 1$, an inactive node v becomes activated if $\sum_{activated\ neighbors\ u} w(u,v) \geq \theta_v$.
  - The weight of the edge (u,v) is calculated as $w(u,v) = \frac{1}{d_{in}(v)}$, where $d_{in}(v)$ denotes the in-degree of node v.