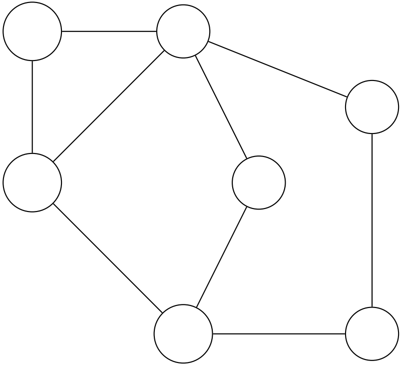
Modeling Content Spreading On Networks Through Network Science And Graph Neural Networks

Abhinav Chand

Kansas State University

Joint Mathematics Meetings
January 10, 2025

Model of Content Spreading

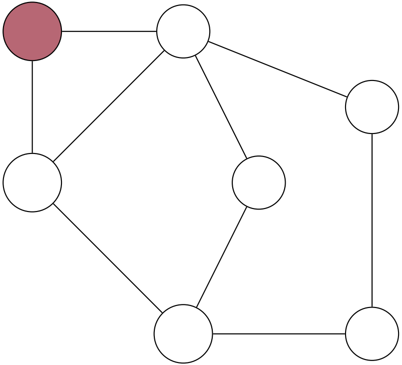


H.Z. Brooks, M.Porter (2024)

An "Opinion Reproduction Number" for Infodemics in a Bounded-Confidence Content-Spreading Process on Networks. arxiv:2403.01066

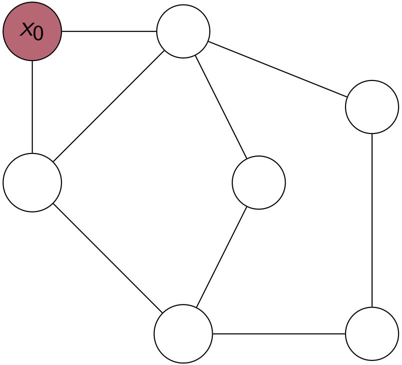
- Undirected network

Model of Content Spreading



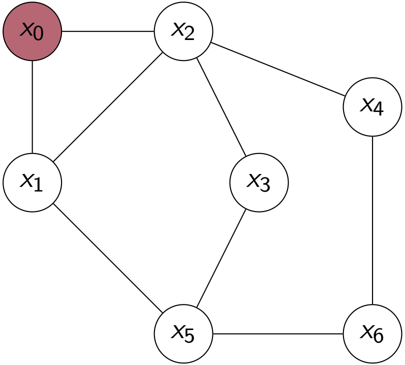
- Undirected network
- Choose source node 0 uniformly at random

Model of Content Spreading



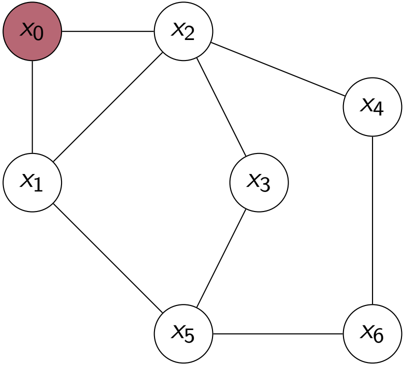
- Undirected network
- Choose source node 0 uniformly at random
- Content $x_0 \in [0, 1]$ starts at node 0

Model of Content Spreading



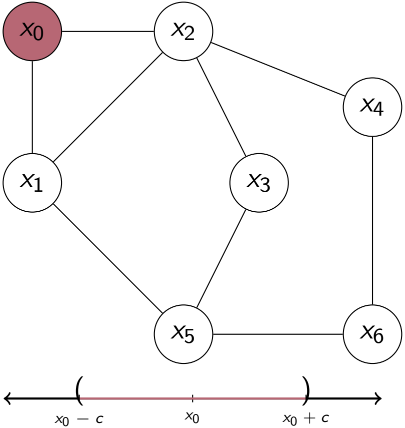
- Undirected network
- Choose source node 0 uniformly at random
- Content $x_0 \in [0, 1]$ starts at node 0
- Initialize other nodes with opinion states $x_i \in [0, 1]$ chosen from some distribution

Model of Content Spreading



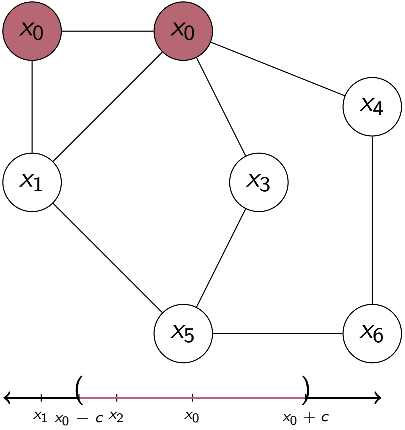
- Node 0 is active

Model of Content Spreading



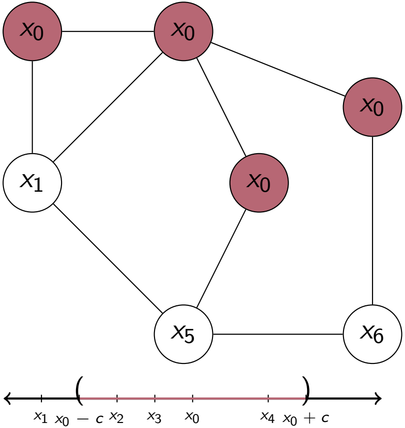
- Node 0 is active
- Choose a receptiveness parameter $c \in [0, \frac{1}{2}]$.

Model of Content Spreading



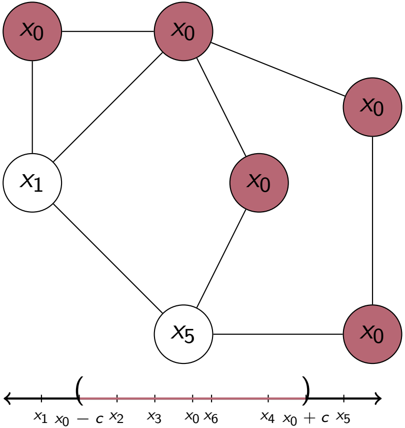
- Node 0 is active
- Choose a receptiveness parameter $c \in [0, \frac{1}{2}]$.
- For a neighbor i of 0, if $|x_i - x_0| < c$, activate it and x_i adopts the state x_0.

Model of Content Spreading



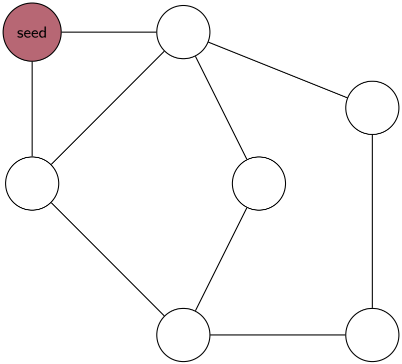
- Node 0 is active
- Choose a receptiveness parameter $c \in [0, \frac{1}{2}]$.
- For a neighbor i of 0, if $|x_i - x_0| < c$, activate it and x_i adopts the state x_0.
- The newly activated nodes try to influence their neighbors.

Model of Content Spreading



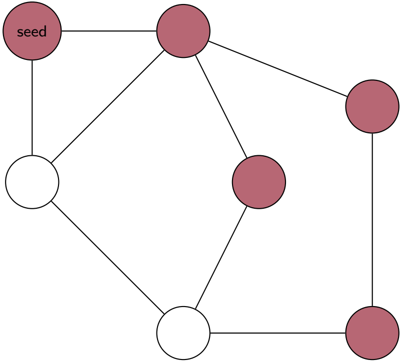
- Node 0 is active
- Choose a receptiveness parameter $c \in [0, \frac{1}{2}]$.
- For a neighbor i of 0, if $|x_i - x_0| < c$, activate it and x_i adopts the state x_0.
- The newly activated nodes try to influence their neighbors.

Site percolation is a special case



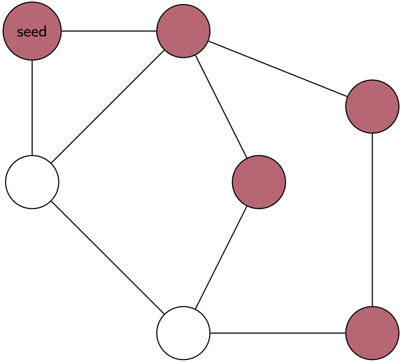
- Seed node is open.

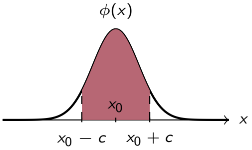
Site percolation is a special case



- Seed node is open.
- Every node is open with probability p and closed with probability $1 - p$.

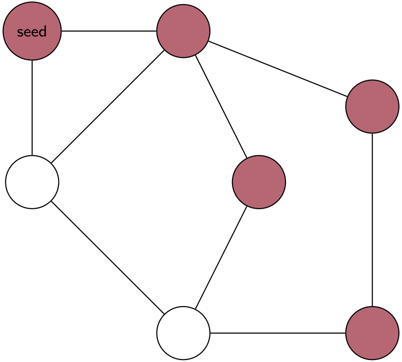
Site percolation is a special case

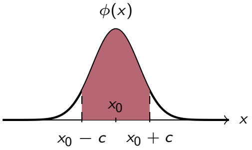




- Seed node is open.
- Every node is open with probability p and closed with probability $1 - p$.
- The probability $p = \int_{x_0-c}^{x_0+c} \phi(x)dx$.

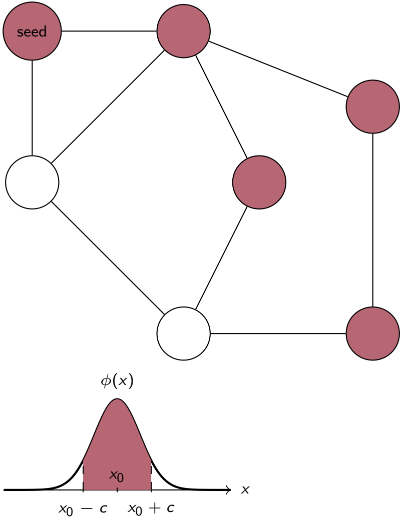
Site percolation is a special case





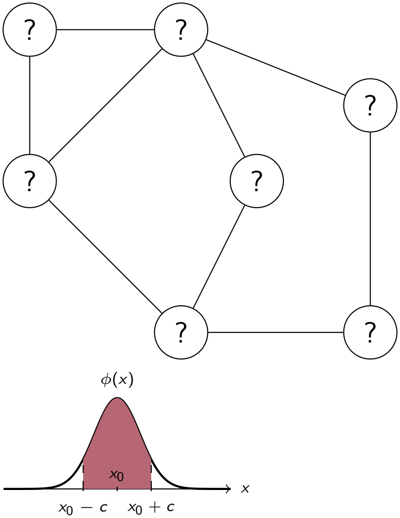
- Seed node is open.
- Every node is open with probability p and closed with probability $1 - p$.
- The probability $p = \int_{x_0-c}^{x_0+c} \phi(x)dx$.
- p is called occupation probability.

Site percolation is a special case



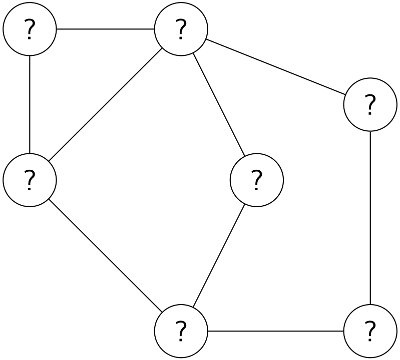
- Seed node is open.
- Every node is open with probability p and closed with probability $1 - p$.
- The probability $p = \int_{x_0-c}^{x_0+c} \phi(x)dx$.
- p is called occupation probability.
- Influenced nodes are nodes that are contained in the "open" cluster containing the seed.

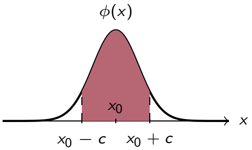
Two Questions



- Content state x_0, receptiveness parameter c and the distribution ϕ is known.

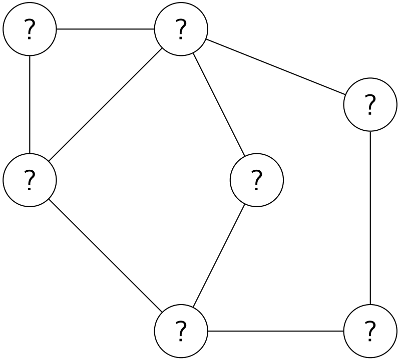
Two Questions

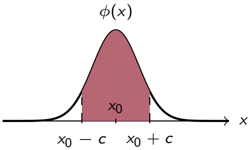




- Content state x_0, receptiveness parameter c and the distribution ϕ is known.
- **Influence Maximization:** What is the best seed to start the diffusion to maximize influence?
  $\mathbb{E}$[size of cluster containing seed]

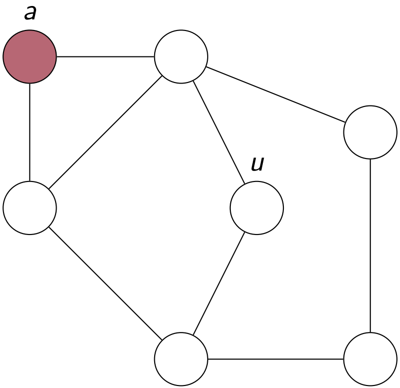
Two Questions





- Content state x_0, receptiveness parameter c and the distribution ϕ is known.
- **Influence Maximization:** What is the best seed to start the diffusion to maximize influence?
  $\mathbb{E}$[size of cluster containing seed]
- **Influence Computation:** Given a seed what is the probability that a node is influenced?
  $\mathbb{P}_{\text{seed}}(v \text{ is influenced})$ for $v \in V$.
  (Note: This does not refer to occupation probability)

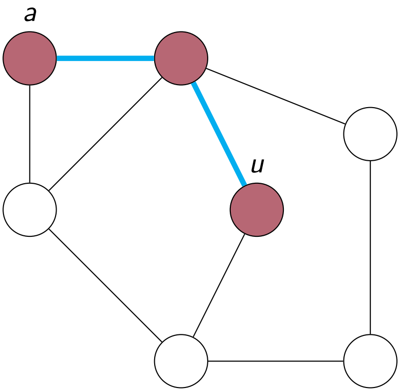
Example



- a is the seed node. What is $\mathbb{P}_a(u \text{ influenced})$?
- Assume we know the "openness/occupation" probability p.

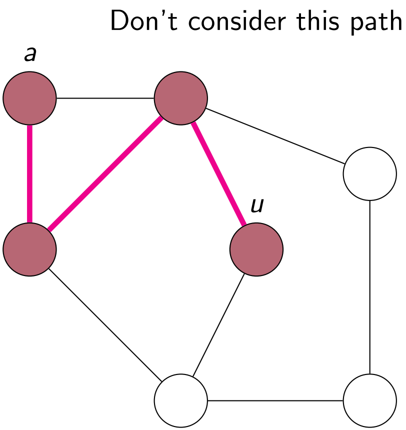
Example

Consider this path



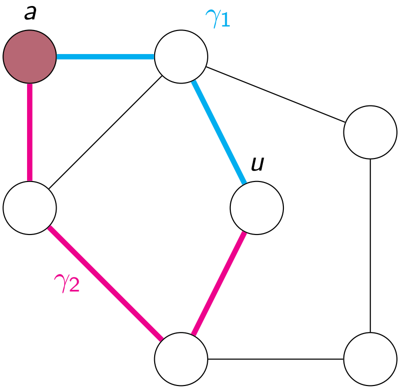
- a is the seed node. What is $\mathbb{P}_a(u \text{ influenced})$?
- Assume we know the "openness/occupation" probability p.

Example



- a is the seed node. What is $\mathbb{P}_a(u \text{ influenced})$?
- Assume we know the "openness/occupation" probability p.

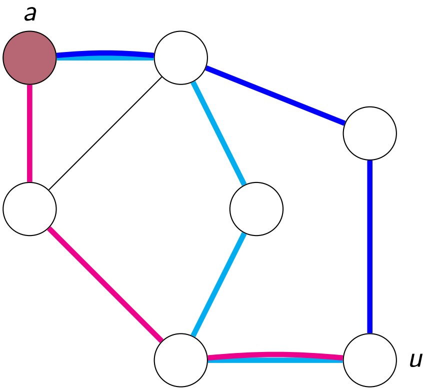
Example



- a is the seed node. What is $\mathbb{P}_a(u \text{ influenced})$?
- Assume we know the "openness/occupation" probability p.
- Two paths γ_1 and γ_2. So

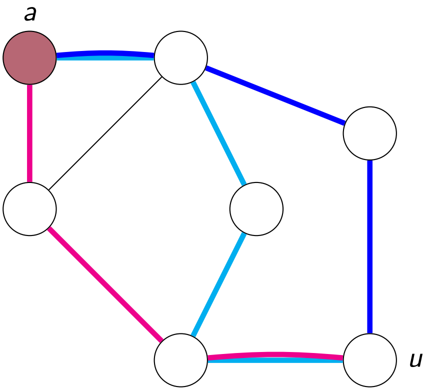
$$\begin{aligned}\mathbb{P}_a(u) &= \mathbb{P}(\gamma_1 \cup \gamma_2) \\ &= \mathbb{P}(\gamma_1) + \mathbb{P}(\gamma_2) - \mathbb{P}(\gamma_1 \cap \gamma_2) \\ &= p^2 + p^3 - p^5\end{aligned}$$

General Case



- Difficult: have to consider all chord-less paths and their intersections.

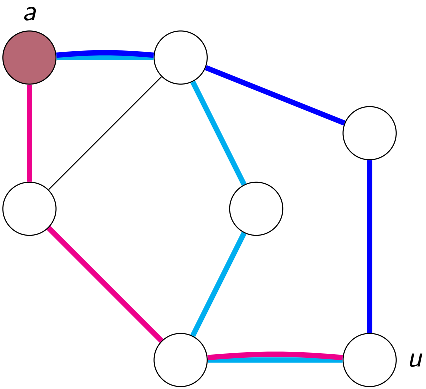
General Case



- Difficult: have to consider all chord-less paths and their intersections.
- Monstrous calculations (Bonferroni type):

$$\mathbb{P}(\gamma_1 \cup \gamma_2 \cup \gamma_3) = \mathbb{P}(\gamma_1) + \mathbb{P}(\gamma_2) + \mathbb{P}(\gamma_3) - \mathbb{P}(\gamma_1 \cap \gamma_2) - \mathbb{P}(\gamma_1 \cap \gamma_3) - \mathbb{P}(\gamma_2 \cap \gamma_3) + \mathbb{P}(\gamma_1 \cap \gamma_2 \cap \gamma_3)$$

General Case

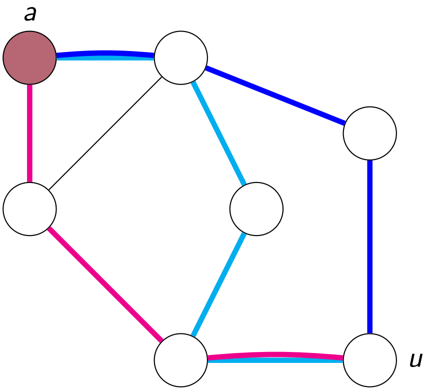


- Difficult: have to consider all chord-less paths and their intersections.
- Monstrous calculations (Bonferroni type):

$$\mathbb{P}(\gamma_1 \cup \gamma_2 \cup \gamma_3) = \mathbb{P}(\gamma_1) + \mathbb{P}(\gamma_2) + \mathbb{P}(\gamma_3) - \mathbb{P}(\gamma_1 \cap \gamma_2) - \mathbb{P}(\gamma_1 \cap \gamma_3) - \mathbb{P}(\gamma_2 \cap \gamma_3) + \mathbb{P}(\gamma_1 \cap \gamma_2 \cap \gamma_3)$$

- Paths might have intersections which implies correlations between events.

General Case

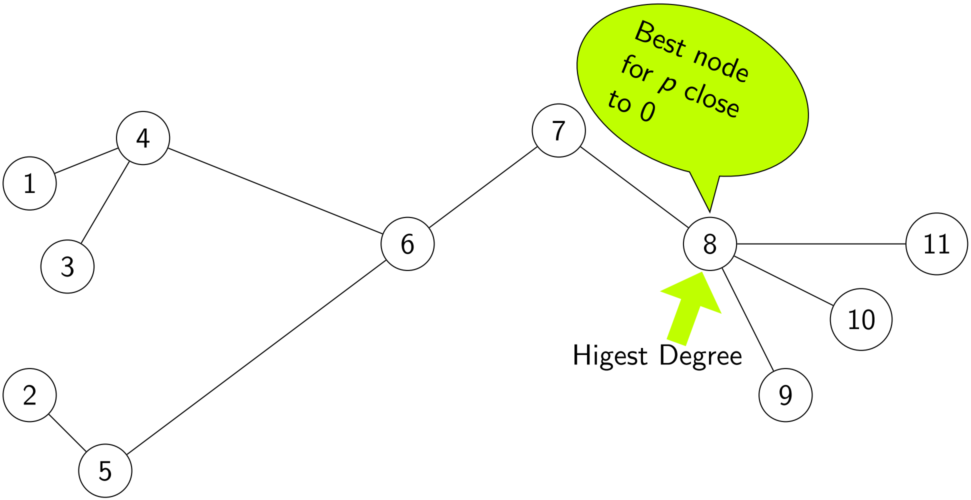


- Difficult: have to consider all chord-less paths and their intersections.
- Monstrous calculations (Bonferroni type):

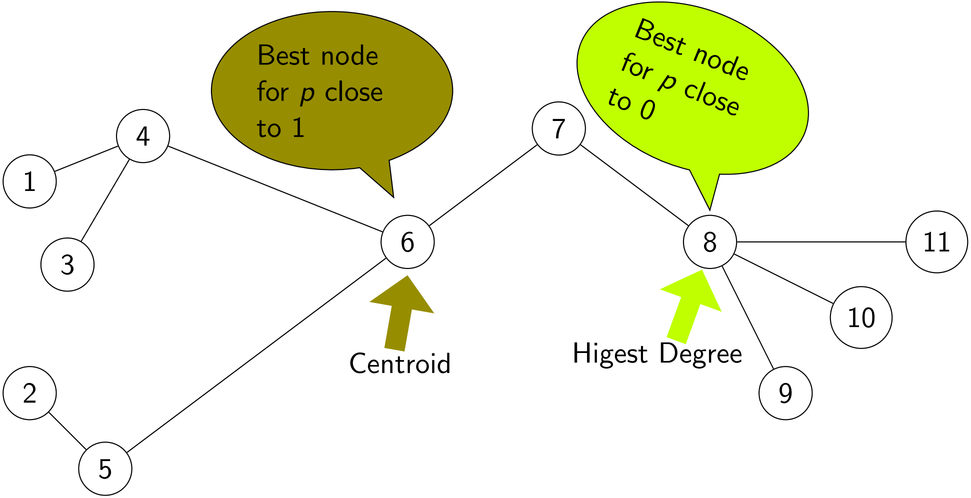
$$\mathbb{P}(\gamma_1 \cup \gamma_2 \cup \gamma_3) = \mathbb{P}(\gamma_1) + \mathbb{P}(\gamma_2) + \mathbb{P}(\gamma_3) - \mathbb{P}(\gamma_1 \cap \gamma_2) - \mathbb{P}(\gamma_1 \cap \gamma_3) - \mathbb{P}(\gamma_2 \cap \gamma_3) + \mathbb{P}(\gamma_1 \cap \gamma_2 \cap \gamma_3)$$

- Paths might have intersections which implies correlations between events.
- Influence maximization is NP-hard, and influence computation is #P-hard.

Trees are more tractable



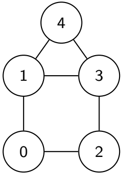
Trees are more tractable



Towards Deep Learning: Node2Vec for input embeddings

Goal

Train a model to learn influence probabilities.

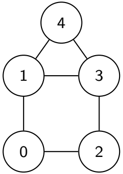


Graph

Towards Deep Learning: Node2Vec for input embeddings

Goal

Train a model to learn influence probabilities.



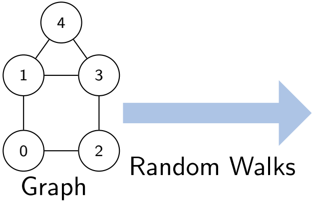
Graph

Random Walks

Towards Deep Learning: Node2Vec for input embeddings

Goal

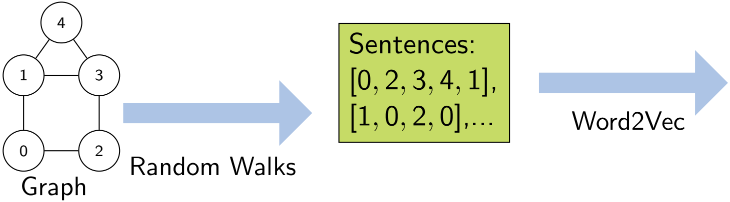
Train a model to learn influence probabilities.



Sentences:
$[0, 2, 3, 4, 1]$,
$[1, 0, 2, 0]$,...

Goal

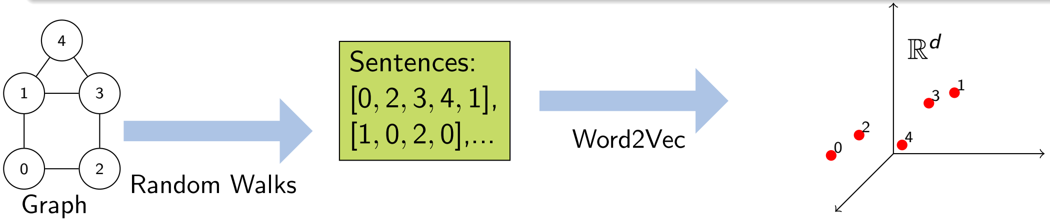
Train a model to learn influence probabilities.



Towards Deep Learning: Node2Vec for input embeddings

Goal

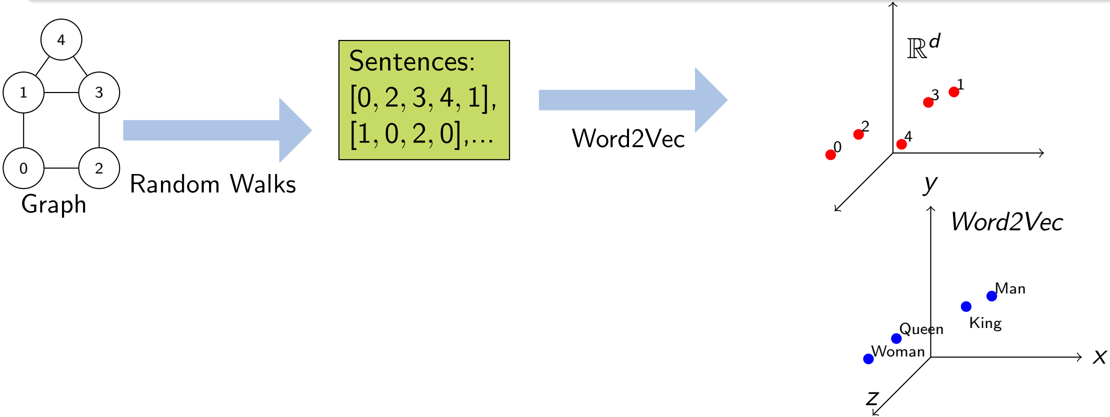
Train a model to learn influence probabilities.



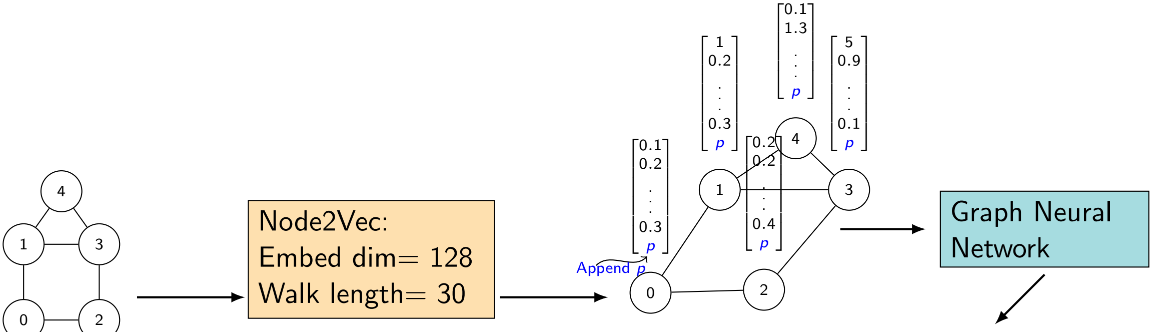
Towards Deep Learning: Node2Vec for input embeddings

Goal

Train a model to learn influence probabilities.

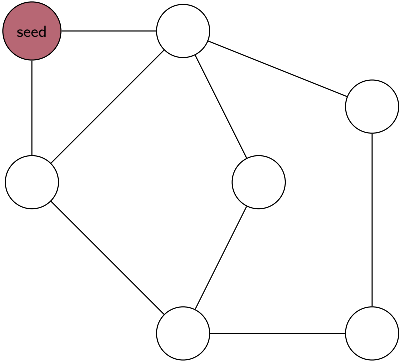


Model Architecture



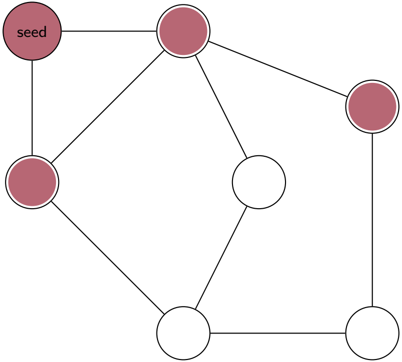
$$
\begin{array}{c}
\mathbb{P}_0 \\ \mathbb{P}_1 \\ \mathbb{P}_2 \\ \mathbb{P}_3 \\ \mathbb{P}_4
\end{array}
\begin{bmatrix}
1 & 0.2 & 0.2 & 0.08 & 0.04 \\
0.2 & 1 & 0.04 & 0.2 & 0.2 \\
0.2 & 0.04 & 1 & 0.2 & 0.04 \\
0.8 & 0.2 & 0.2 & 1 & 0.04 \\
0.04 & 0.1 & 0.04 & 0.04 & 1
\end{bmatrix}
$$

Synthetic Dataset



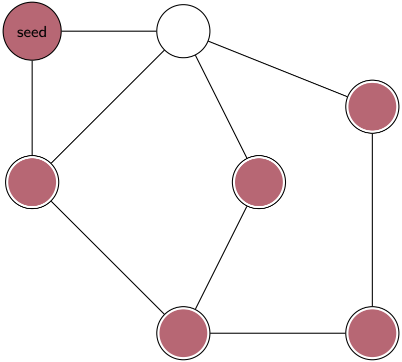
- Monte Carlo Simulations

Synthetic Dataset



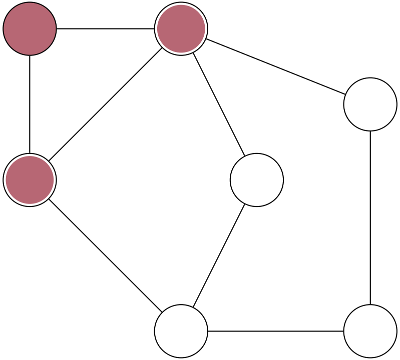
- Monte Carlo Simulations

Synthetic Dataset



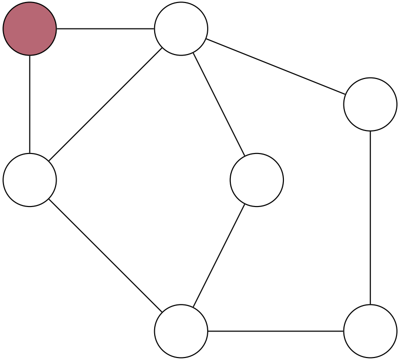
- Monte Carlo Simulations

Synthetic Dataset



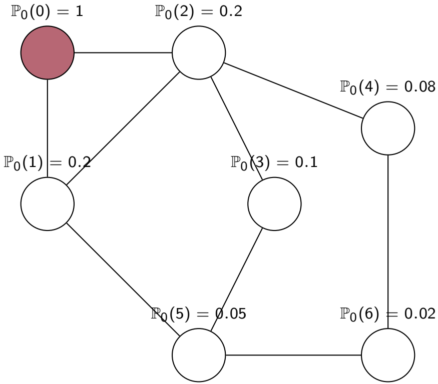
- Monte Carlo Simulations

Synthetic Dataset



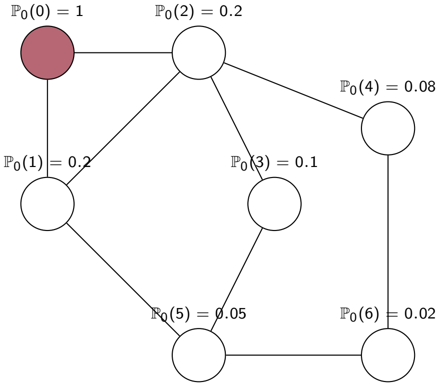
- Monte Carlo Simulations
- For each random graph, run MC starting from each node for 9 randomly chosen probabilities $p \in [0, 1]$.
  Complexity: $O(|V|^2) \cdot |V|$

Synthetic Dataset



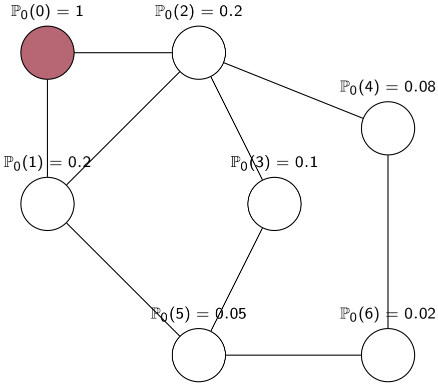
- Monte Carlo Simulations
- For each random graph, run MC starting from each node for 9 randomly chosen probabilities $p \in [0, 1]$.
  Complexity: $O(|V|^2) \cdot |V|$
- Example: $p = 0.2$
  Get first row of matrix: For $p = 0.2$, $[\mathbb{P}_0(0), \mathbb{P}_0(1), ..., \mathbb{P}_0(6)]$

Synthetic Dataset



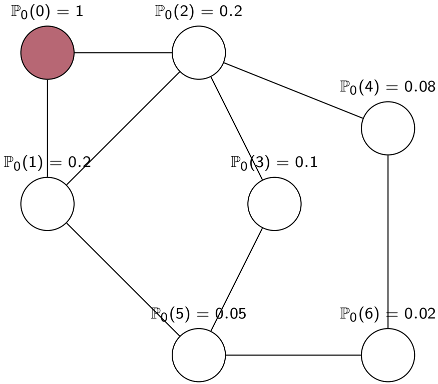
- Monte Carlo Simulations
- For each random graph, run MC starting from each node for 9 randomly chosen probabilities $p \in [0, 1]$.
  Complexity: $O(|V|^2) \cdot |V|$
- Example: $p = 0.2$
  Get first row of matrix: For $p = 0.2$,
  $[\mathbb{P}_0(0), \mathbb{P}_0(1), ..., \mathbb{P}_0(6)]$
- 1000 random graphs of sizes 8-50 (random geometric, erdos-renyi, etc.).

Synthetic Dataset



- Monte Carlo Simulations
- For each random graph, run MC starting from each node for 9 randomly chosen probabilities $p \in [0, 1]$.
  Complexity: $O(|V|^2) \cdot |V|$
- Example: $p = 0.2$
  Get first row of matrix: For $p = 0.2$,
  $[\mathbb{P}_0(0), \mathbb{P}_0(1), ..., \mathbb{P}_0(6)]$
- 1000 random graphs of sizes 8-50 (random geometric, erdos-renyi, etc.).
- $\sim$11,000 data points.

Synthetic Dataset



- Monte Carlo Simulations
- For each random graph, run MC starting from each node for 9 randomly chosen probabilities $p \in [0, 1]$.
  Complexity: $O(|V|^2) \cdot |V|$
- Example: $p = 0.2$
  Get first row of matrix: For $p = 0.2$, $[\mathbb{P}_0(0), \mathbb{P}_0(1), ..., \mathbb{P}_0(6)]$
- 1000 random graphs of sizes 8-50 (random geometric, erdos-renyi, etc.).
- $\sim$11,000 data points.
- 90-10 train-test split.

GNN Parameters and Results

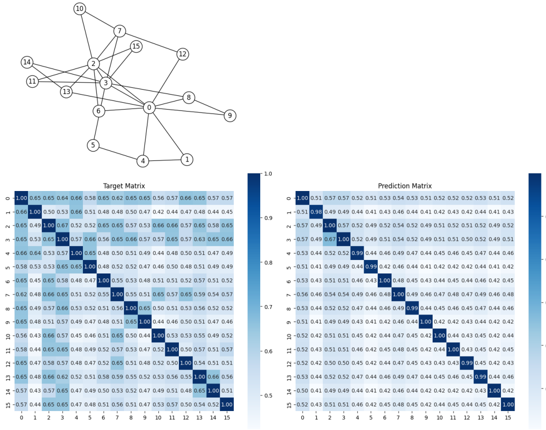


Figure: Sample graph with target and predicion matrices for $p = 0.65$

- Loss Function: Mean Squared Error (MSE)
- Graph neural network parameters:
  Number of layers: 3 (dropout 0.1)
  Hidden Dimension: 256
  Batch: 4
- Training time $\sim$15 minutes (GPU). Early stopping after 23 epochs
- Train Loss: MSE 0.0063
  Test Loss: MSE 0.020

Future Work

- Interpret the graph neural network model. What is the model doing?

Future Work

- Interpret the graph neural network model. What is the model doing?
- Find approximation algorithms for influence computation and influence maximization.

Future Work

- Interpret the graph neural network model. What is the model doing?
- Find approximation algorithms for influence computation and influence maximization.
- Apply the methods from this work to other models in opinion dynamics and epidemic modeling.

Thank You!