Generative Models for Graphs

STAT-S 675

Fall 2021

Introduction

- Let $G = (V, E)$ be an undirected and hollow graph with $|V| = n$ and adjacency matrix A
    - $A \in \mathbb{R}^{n \times n}$ is symmetric with zero diagonals
- Suppose $G \sim F(\theta)$
    - What kind of $F(\theta)$ make sense here?
    - Given F and observed G, how can we estimate θ?

Bernoulli Graphs

$$A_{ij} = \begin{cases} 1 & \exists \text{ edge between } i \text{ and } j \\ 0 & \text{else} \end{cases}$$

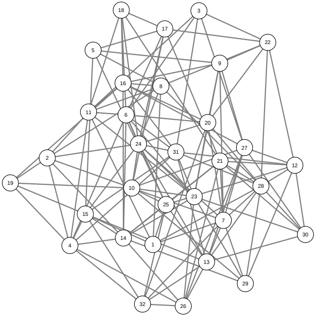
$A_{ji} = A_{ij}$ and $A_{ii} = 0 \ \forall i, j \in [n]$.

$A \sim \text{BernoulliGraph}(P)$ iff:

1. $P \in [0,1]^{n \times n}$ describes edge probabilities between pairs of vertices.
2. $A_{ij} \overset{\text{ind}}{\sim} \text{Bernoulli}(P_{ij})$ for each $i < j$.

Example 1: If every entry $P_{ij} = \theta$, then $A \sim \text{BernoulliGraph}(P)$ is an Erdos-Renyi graph.

For this model, $A_{ij} \overset{\text{iid}}{\sim} \text{Bernoulli}(\theta)$.

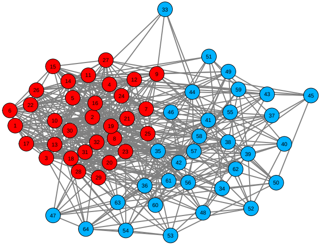


Block Models

Suppose each vertex $v_1, ..., v_n$ has hidden labels $z_1, ..., z_n \in [K]$, and each P_{ij} depends on labels z_i and z_j.
Then $A \sim \text{BernoulliGraph}(P)$ is a *block model*.

Example 2: Stochastic Block Model with two communities

- $z_1, ..., z_n \in \{1, 2\}$
- $P_{ij} = \begin{cases} p & z_i = z_j = 1 \\ q & z_i = z_j = 2 \\ r & z_i \neq z_j \end{cases}$
- To make this an assortative SBM, set $pq > r^2$.
- In this example, $p = 1/2$, $q = 1/4$, and $r = 1/8$.



Block Models

Erdos-Renyi Model (1959)

- $P_{ij} = \theta$ (not a block model)
- 1 parameter θ

Stochastic Block Model (Lorrain and White, 1971)

- $P_{ij} = \theta_{z_i z_j}$
- $K(K+1)/2$ parameters θ_{kl}

Degree Corrected Block Model (Karrer and Newman, 2011)

- $P_{ij} = \theta_{z_i z_j} \omega_i \omega_j$
- $K(K+1)/2 + n$ parameters θ_{kl}, ω_i

Popularity Adjusted Block Model (Sengupta and Chen, 2017)

- $P_{ij} = \lambda_{i z_j} \lambda_{j z_i}$
- Kn parameters λ_{ik}

Community Detection and Parameter Estimation

$$L(P) = \prod_{i<j} P_{ij}^{A_{ij}} (1 - P_{ij})^{1-A_{ij}}$$

- Erdos-Renyi: $L(\theta) = \prod_{i<j} \theta^{A_{ij}} (1-\theta)^{1-A_{ij}}$
  $\implies \hat{\theta} = \frac{\sum_{i<j} A_{ij}}{n(n-1)/2}$
- SBM: $L(\vec{z}, \{\theta_{kl}\}) = \prod_{i<j} \prod_{k,l}^{K} \theta_{kl}^{A_{ij} z_{ik} z_{jl}} (1-\theta_{kl})^{(1-A_{ij}) z_{ik} z_{jl}}$
  Computing MLEs for $\vec{z}$ and $\{\theta_{kl}\}$ is NP-hard
- DCBM, PABM ...

Expectation Maximization for SBM

$$\ell(\vec{z}, \{\theta_{kl}\}) = \sum_{i,j} \sum_{k,l}^{K} A_{ij} z_{ik} z_{jl} \log \theta_{kl} + (1 - A_{ij}) z_{ik} z_{jl} \log(1 - \theta_{kl})$$

- Mean field approximation: Assume the labels z_{ik} are independent
- E-step: $E[z_{ik}] = \pi_{ik}$
  $\propto \exp\left(\sum_{j \neq i} \sum_{l} \pi_{jl} (A_{ij} \log \theta_{kl} + (1 - A_{ij}) \log(1 - \theta_{kl})) \right)$
- M-step: $\hat{\theta}_{kl} = \frac{\sum_{i<j} A_{ij} \pi_{ik} \pi_{jl}}{\sum_{i<j} \pi_{ik} \pi_{jl}}$
- Similar types of approaches for DCBM and PABM
- Mean field approximation may or may not be correct

Implementation

- Demo