Topic 19: Community Detection

by Sai Zhang

Key points: .

Disclaimer: *The note is built on Prof. Jinchi Lv's lectures of the course at USC, DSO 607, High-Dimensional Statistics and Big Data Problems.*

19.1   Stochastic Block Model

Consider an undirected graph G, with nodes V and edges E. Let

- n be a positive integer: the number of **vertices**
- k be a positive integer: the number of **communities**
- $p = (p_1, \cdots, p_k)$ be a probability vector on $\{1, \cdots, k\} := [k]$: the **prior** on the k communities
- $\mathbf{W}$ be a $k \times k$ symmetric matrix with entries $W_{ij} \in [0, 1]$: the matrix of **connectivity probabilities**

then we have

> **Definition 19.1.1: Stochastic Block Model**
>
> The pair $(\mathbf{X}, G)$ is drawn under $SBM(n, p, \mathbf{W})$ if $\mathbf{X}$ is an n dimensional random vector with i.i.d. components distributed under p, and G is an $n-$vertex simple graph where vertices i and j are connected with probability W_{X_i, X_j}, **independently** of other pairs of vertices. And the **community** sets can be defined