Mixed Membership model in Network Data.

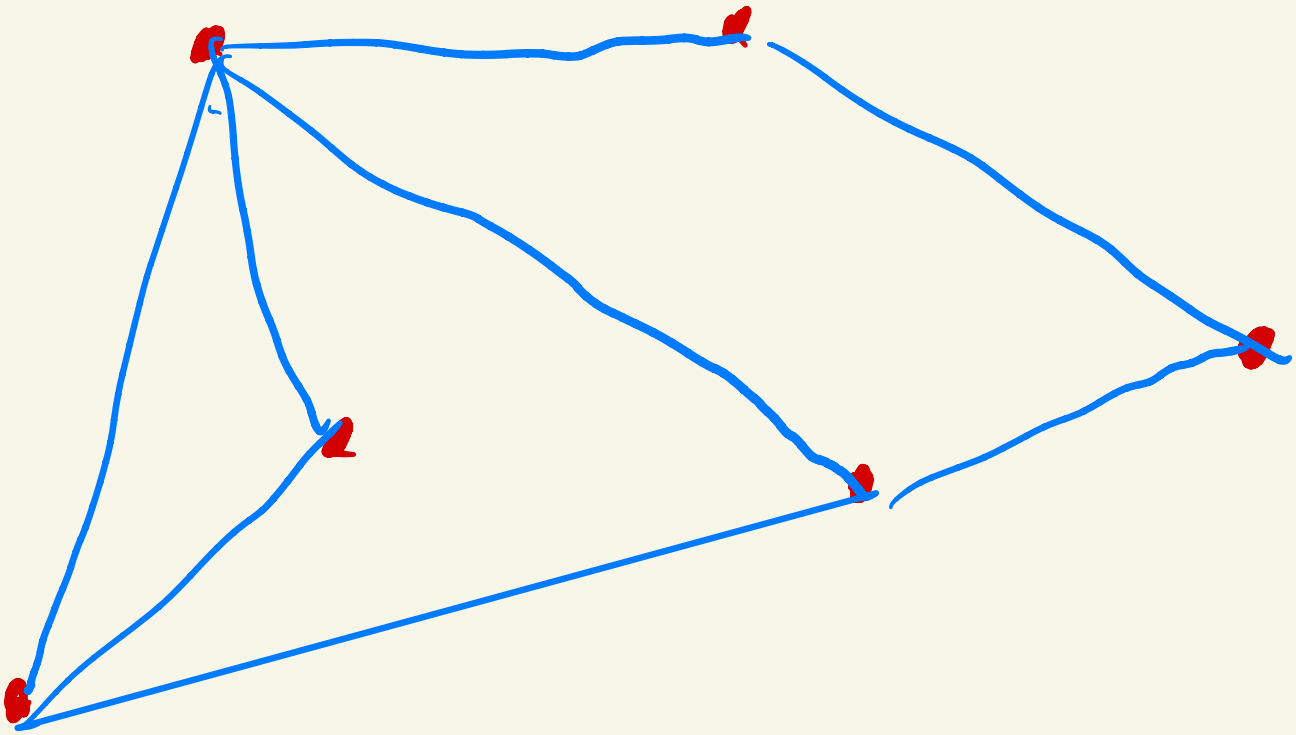
Sep 10, 2021 Bruce

Contents

1. Iroduction of Network
2. SBM
3. Mixed Membership SBM.
4. Programming Cases.

Basics of Network data.

Network : Nodes , Edges



Directed / Undirected.

Weighted / Unweighted

Attributed / Unattributed.

Simple Network.

SBM: Stochastic Block Model

0. assume K communities, n nodes

1. assign each node a community,

define a indicator matrix $M_{K \times n}$

each column of M: $\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ indicator vector

2. define a probability matrix $C_{K \times K}$

C_{ij} is probability of occurence of an edge between community i and community j.

Remark: row or col sum of C is 1.

Symmetric — undirected network

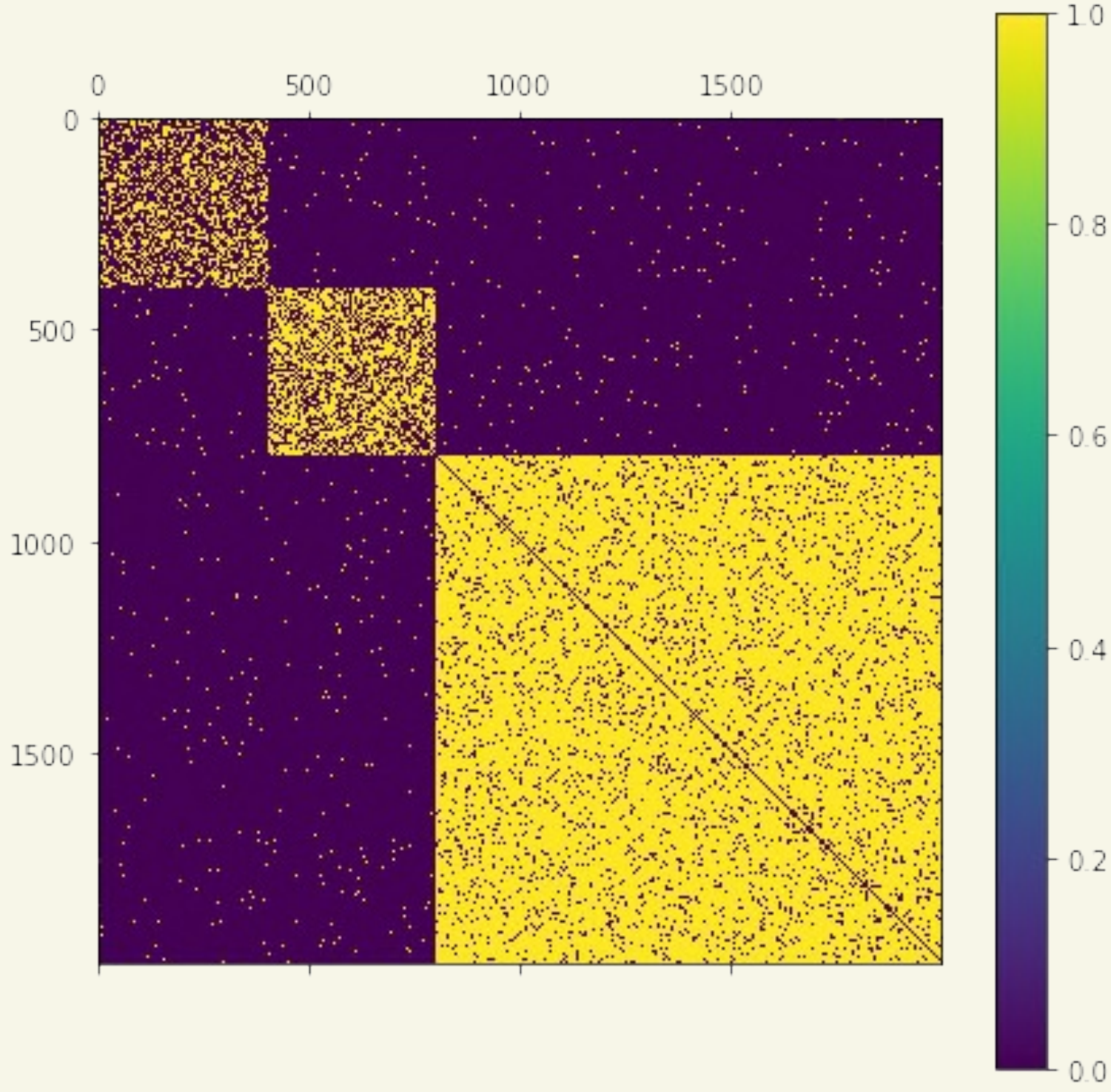
ASymmetric — directed network

Network: $P_{n \times n} = M^T_{n \times k} C_{K \times K} M_{K \times n}$

$A_{n \times n} \sim$ Bernoulli (P) adjacency matrix.

$$A_{ij} = \begin{cases} 1 & \text{if edge } (i,j) \text{ exist} \\ 0 & \text{else.} \end{cases}$$

```
model_type='SBM'
sizes = [400, 400, 1200]
probs = [[0.4, 0.01, 0.01],
         [0.01, 0.6, 0.01],
         [0.01, 0.01, 0.9]]
SBM = nx.stochastic_block_model(sizes, probs, seed=0)
A=nx.to_numpy_matrix(SBM)
```



For SBM: one node could only belong to one community.

For MMSBM: capture overlapping communities

Examples: Latent Structure Lab member

Statistics department

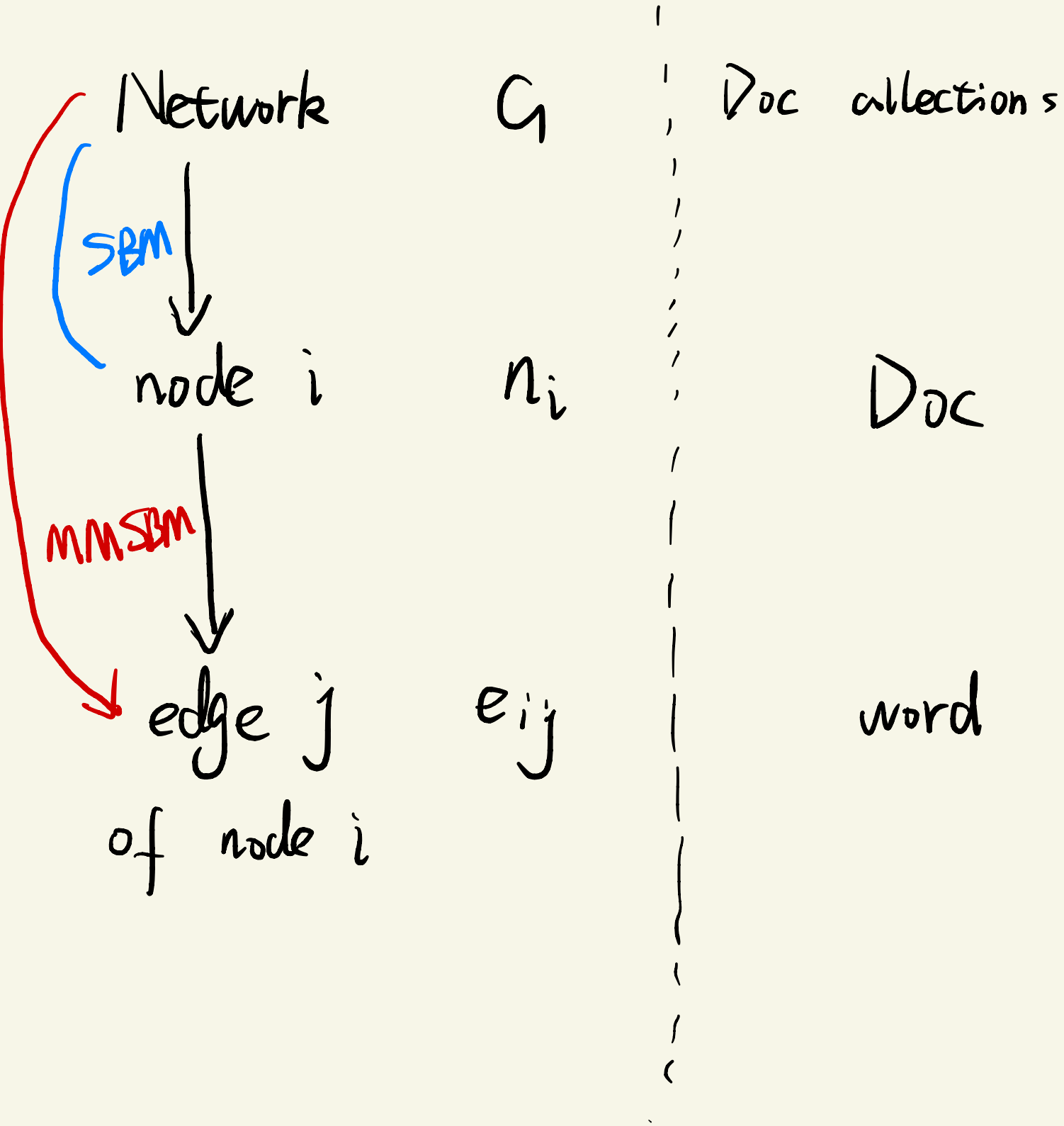
High school

.....

Multiple communities for one node.

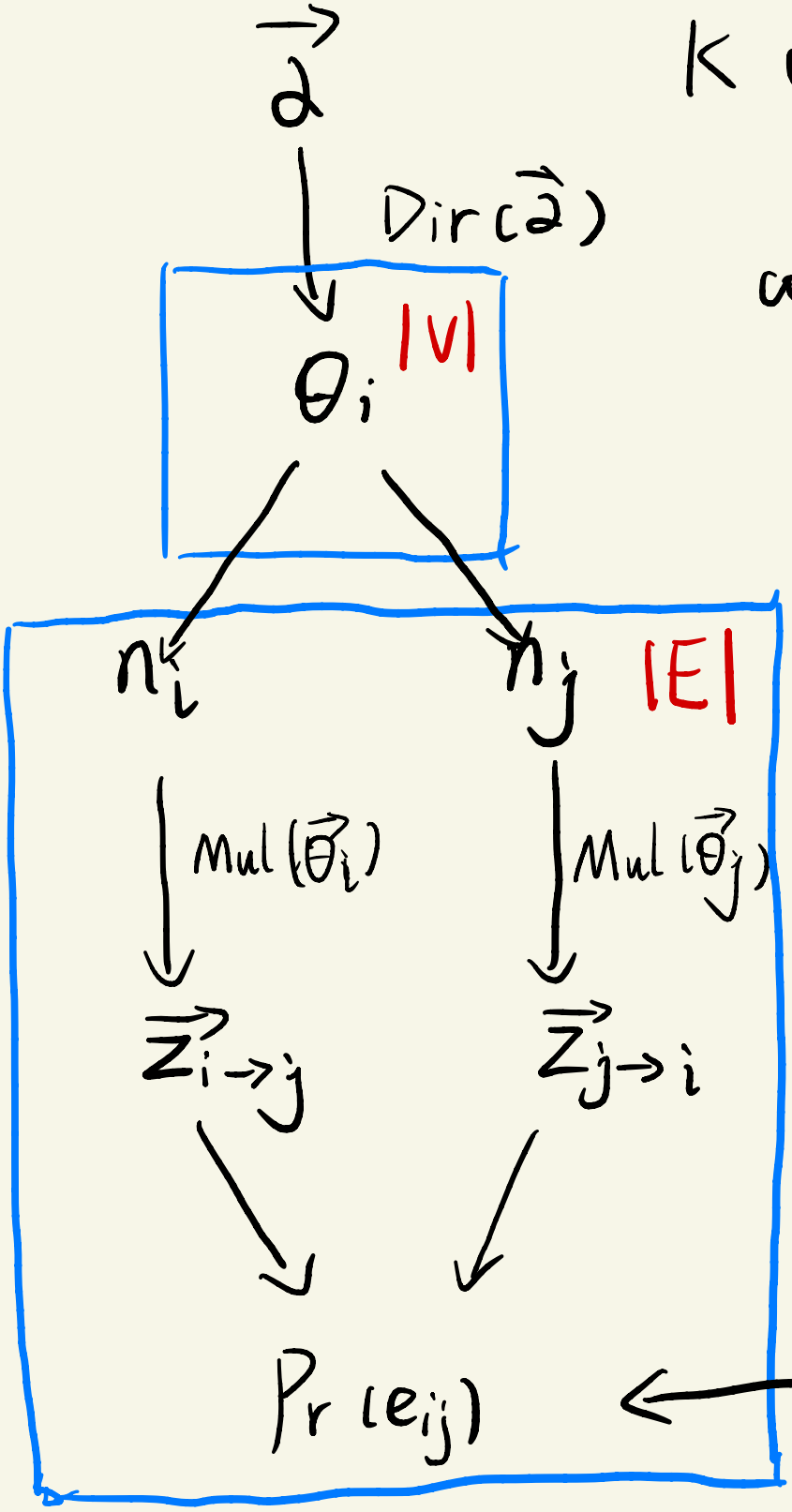
SBM — node level

MMSBM — edge level



MMSBM

$G = (V, E)$



K dim vector (communities)

communities distribution of node i (weight, portion)

$$\vec{z}_{i\to j} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}_{K\times 1}$$

community probability matrix

$$Pr(e_{ij}) = \vec{z}_{i\to j}^{\,T} \cdot C \cdot \vec{z}_{j\to i}$$

$$A_{ij} \sim Ber(Pr(e_{ij}))$$

Special point of Network data.

Dependency ↓

inherent charcteristic of network data

MMSBM Simulation