Stochastic Blockmodel with Cluster Overlap, Relevance Selection, and Similarity-Based Smoothing

Joyce Jiyoung Whang[1] Piyush Rai[2] Inderjit S. Dhillon[1]

[1]The University of Texas at Austin

[2]Duke University

International Conference on Data Mining

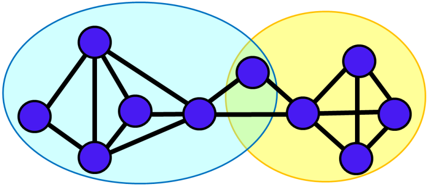
Dec. 7 - Dec. 10, 2013.

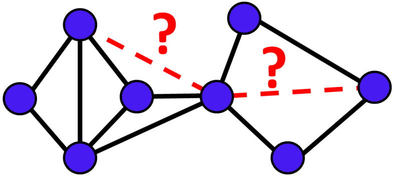
Contents

Introduction

- Stochastic Blockmodel
  - Generative model
  - Expresses objects as a low dimensional representation U_i, U_j
  - Models the link probability of a pair of objects $P(A_{ij}) = f(U_i, U_j, \boldsymbol{\theta})$
  - e.g., latent class model, mixed membership stochastic blockmodel

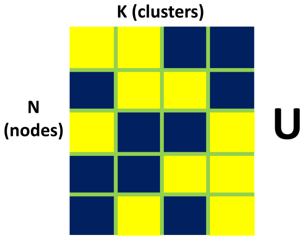
- Applications
  - Revealing structures in networks
  - (Overlapping) Clustering, Link prediction





Introduction

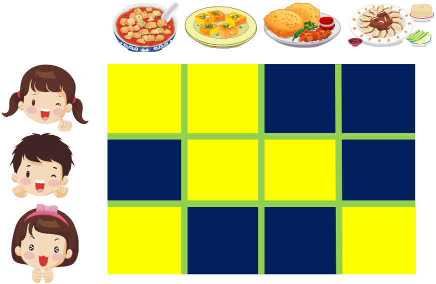
- Overlapping stochastic blockmodels
  - Objects have **hard** memberships in **multiple** clusters.



- Contributions of this paper
  - Extend the overlapping stochastic blockmodel to bipartite graphs
  - Relevance selection mechanism
  - Make use of additionally available object features
  - Nonparametric Bayesian approach

Background

- Indian Buffet Process (IBP) (Griffiths et al. 2011)
  - N objects, K clusters, overlapping clustering $\mathbf{U} \in \{0,1\}^{N \times K}$.
  - Object: customer, cluster: dish
  - The first customer selects $Poisson(\alpha)$ dishes to begin with
  - Each subsequent customer n:
    - Selects an already selected dish k with probability $\frac{m_k}{n}$
    - Selects $Poisson(\alpha/n)$ new dishes



The Proposed Model

Basic Model

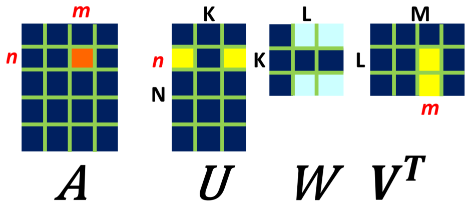
- Bipartite graph ($N \times M$ binary adjacency matrix, $|\mathcal{A}| = N$, $|\mathcal{B}| = M$)

$$\begin{aligned}
\mathbf{U} &\sim \mathcal{IBP}(\alpha_u) \\
\mathbf{V} &\sim \mathcal{IBP}(\alpha_v) \\
\mathbf{W} &\sim \mathcal{N}or(0, \sigma_w^2) \\
\mathbf{A} &\sim \mathcal{B}er(\sigma(\mathbf{UWV}^\top))
\end{aligned}$$

- $\mathcal{IBP}(\alpha)$: IBP prior distribution, $\mathcal{N}or(0, \sigma^2)$: Gaussian distribution,
- $\sigma(x) = \frac{1}{1+\exp(-x)}$, $\mathcal{B}er(p)$: Bernoulli distribution,
- $\mathbf{U} \in \{0,1\}^{N \times K}$, $\mathbf{V} \in \{0,1\}^{M \times L}$: cluster assignment matrices

$$\begin{aligned}
P(A_{nm} = 1) &= \sigma(\mathbf{u}_n \mathbf{W} \mathbf{v}_m^\top) \\
&= \sigma(\sum_{k,l} u_{nk} W_{kl} v_{ml})
\end{aligned}$$

- W_{kl}: the interaction strength between two nodes due to their memberships in cluster k and cluster l



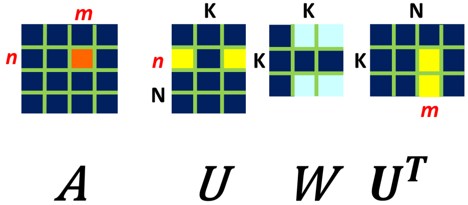
$$P(A_{nm} = 1) = \sigma(W_{12} + W_{13} + W_{32} + W_{33})$$

Basic Model

- Unipartite graph ($\mathbf{A} \in \{0,1\}^{N \times N}$)

$$\begin{array}{rcl} \mathbf{U} & \sim & \mathcal{IBP}(\alpha_u) \\ \mathbf{W} & \sim & \mathcal{N}or(0, \sigma_w^2) \\ \mathbf{A} & \sim & \mathcal{B}er(\sigma(\mathbf{UWU}^\top)) \end{array}$$

$$\begin{array}{rcl} P(A_{nm} = 1) & = & \sigma(\mathbf{u}_n \mathbf{W} \mathbf{u}_m^\top) \\ & = & \sigma(\sum_{k,l} u_{nk} W_{kl} u_{ml}) \end{array}$$

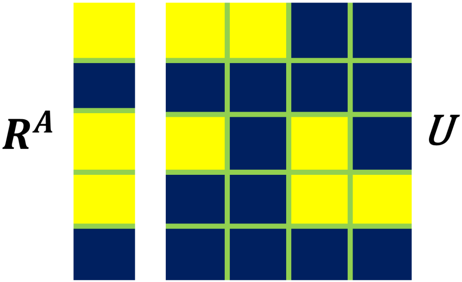


- $\mathcal{IBP}(\alpha)$: IBP prior distribution,
  $\mathcal{N}or(0, \sigma^2)$: Gaussian distribution,
- $\sigma(x) = \frac{1}{1+\exp(-x)}$,
  $\mathcal{B}er(p)$: Bernoulli distribution,
- $\mathbf{U} \in \{0,1\}^{N \times K}$: cluster assignment matrix

$$P(A_{nm} = 1) = \sigma(W_{12} + W_{13} + W_{32} + W_{33})$$

Relevance Selection Mechanism

- Motivation
  - In real-world networks, there may be some noisy objects (e.g., spammer)
  - May lead to bad parameter estimates
- Maintain two random binary vectors $\mathbf{R}^A \in \{0,1\}^{N\times 1}$, $\mathbf{R}^B \in \{0,1\}^{M\times 1}$



Relevance Selection Mechanism

- Background noise link probability $\phi \sim \mathcal{B}et(a, b)$
- If one or both objects $n \in \mathcal{A}$ and $m \in \mathcal{B}$ are irrelevant
  - A_{nm} is drawn from $\mathcal{B}er(\phi)$
- If both n and m are relevant,
  - A_{nm} is drawn from $\mathcal{B}er(p) = \mathcal{B}er(\sigma(\mathbf{u}_n \mathbf{W} \mathbf{v}_m^\top))$

$$
\begin{aligned}
\phi &\sim \mathcal{B}et(a, b) \\
R_n^A &\sim \mathcal{B}er(\rho_n^A), \quad R_m^B \sim \mathcal{B}er(\rho_m^B) \\
\mathbf{u}_n &\sim \mathcal{IBP}(\alpha_u) \quad \text{if } R_n^A = 1\text{; zeros otherwise} \\
\mathbf{v}_m &\sim \mathcal{IBP}(\alpha_v) \quad \text{if } R_m^B = 1\text{, zeros otherwise} \\
p &= \sigma(\mathbf{u}_n \mathbf{W} \mathbf{v}_m^\top) \\
A_{nm} &\sim \mathcal{B}er(p^{R_n^A R_m^B} \phi^{1 - R_n^A R_m^B})
\end{aligned}
$$

Exploiting Pairwise Similarities

- We may have access to side information
  - e.g., a similarity matrix between objects
- The IBP does not consider the pairwise similarity information.
  - Customer n chooses an existing dish regardless of the similarity of this customer with other customers.

- Two objects n and m have a high pairwise similarity $\Rightarrow$ $\mathbf{u}_n$ and $\mathbf{u}_m$ should also be similar.
  - Encourages a customer to select a dish if the customer has a high similarity with all other customers who chose that dish.
  - Let the customer select many new dishes if the customer has low similarity with previous customers.

Exploiting Pairwise Similarities

- Modify the sampling scheme in the IBP based generative model
  - The probability that object n gets membership in cluster k will be proportional to $\frac{\sum_{n' \neq n} S^A_{nn'} u_{n'k}}{\sum_{n'=1}^{n} S^A_{nn'}}$.

    $\sum_{n'=1}^{n} S^A_{nn'}$: effective total number of objects,
    $\sum_{n' \neq n} S^A_{nn'} u_{n'k}$: effective number of objects (other than n) that belong to cluster k

    - IBP: $\frac{\sum_{n' \neq n} u_{n'k}}{n} = \frac{m_k}{n}$
  - The number of new clusters for object n is given by $Poisson(\alpha / \sum_{n'=1}^{n} S^A_{nn'})$.

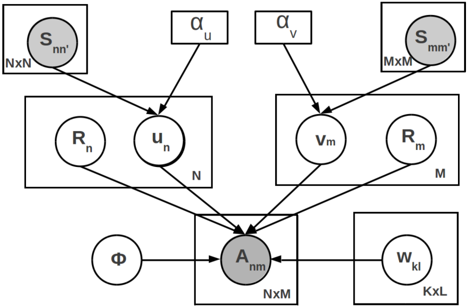
    If the object n has low similarities with the previous objects, encourage it more to get memberships in its own new clusters

    - IBP: $Poisson(\alpha/n)$

The Final Model

- **ROCS** (**R**elevance-based **O**verlapping **C**lustering with **S**imilarity-based-smoothing)

$$
\begin{aligned}
\phi &\sim \mathcal{B}et(a, b) \\
\rho_n^A &\sim \mathcal{B}et(c, d), \quad \rho_m^B \sim \mathcal{B}et(e, f) \\
R_n^A &\sim \mathcal{B}er(\rho_n^A), \quad R_m^B \sim \mathcal{B}er(\rho_m^B) \\
\mathbf{u}_n &\sim Sim\mathcal{IBP}(\alpha_u, \mathbf{S}^A) \\
\mathbf{v}_m &\sim Sim\mathcal{IBP}(\alpha_v, \mathbf{S}^B) \\
p &= \sigma(\mathbf{u}_n \mathbf{W} \mathbf{v}_m^\top) \\
A_{nm} &\sim \mathcal{B}er(p^{R_n^A R_m^B} \phi^{1 - R_n^A R_m^B})
\end{aligned}
$$



  - $Sim\mathcal{IBP}(\alpha_u, \mathbf{S}^A)$: similarity information augmented variant of the IBP

- For inference, we use MCMC (Gibbs sampling)

Experiments

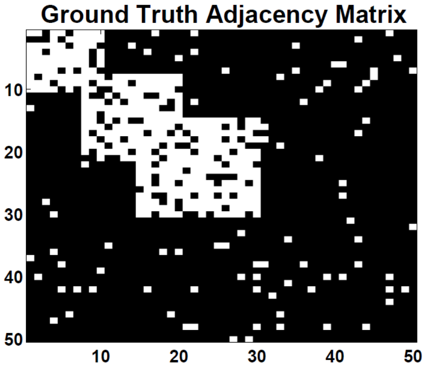
Experiments

- Tasks
  - The correct number of clusters
  - Identify relevant objects
  - Use pairwise similarity information
  - Overlapping clustering
  - Link prediction

- Baselines
  - Overlapping Clustering using Nonnegative Matrix Factorization (OCNMF) (Psorakis et al. 2011)
  - Kernelized Probabilistic Matrix Factorization (KPMF) (Zhou et al. 2012)
  - Bayesian Community Detection (BCD) (Mørup et al. 2012)
  - Latent Feature Relational Model (LFRM) (Miller et al. 2009)

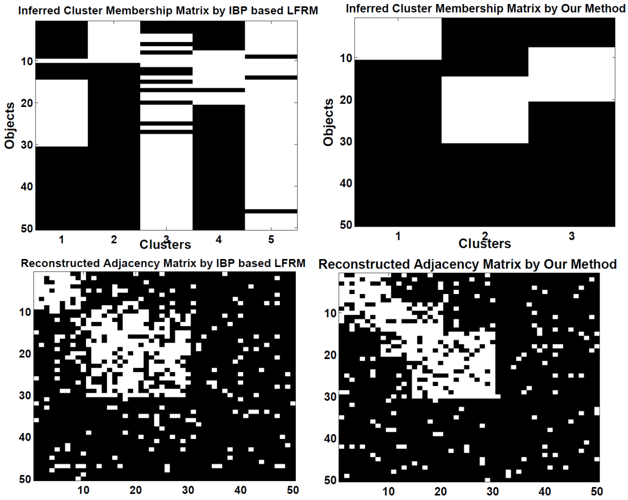
Experiments

- Synthetic Data
  - 30 relevant objects, 20 irrelevant objects
  - Three overlapping clusters



Experiments

- Overlapping clustering



Experiments

Table 1: Link Prediction on Synthetic Data

Method	0-1 Test Error (%)	AUC
OCNMF	44.82 ($\pm$12.59)	0.7164 ($\pm$0.1987)
KPMF	39.70 ($\pm$1.78)	0.6042 ($\pm$0.0517)
BCD	20.05 ($\pm$1.49)	0.8504 ($\pm$0.0197)
LFRM	9.59 ($\pm$0.36)	0.8619 ($\pm$0.0374)
ROCS	**9.05 ($\pm$0.42)**	**0.8787 ($\pm$ 0.0303)**

- Results Summary
  - ROCS perfectly identifies relevant/irrelevant objects
  - ROCS identifies the correct number of clusters
  - For link prediction task, ROCS is better than other methods in terms of both 0-1 test error and AUC score.

Experiments

- Facebook Data
  - An ego-network in Facebook (228 nodes)
  - User profile (e.g., age, gender, etc.) – select 92 features.
  - Known number of clusters: 14

Table 2: Link Prediction on Facebook Data

Method	0-1 Test Error (%)	AUC
OCNMF	36.58 (±19.74)	0.7215 (±0.1666)
KPMF	35.76 (±2.76)	0.7013 (±0.0174)
BCD	13.59 (±0.31)	0.9187 (±0.0242)
LFRM	12.38 (±2.82)	0.9156 (±0.0134)
ROCS	**11.96 (±1.44)**	**0.9388 (± 0.0156)**

- BCD overestimated the number of clusters (20-22 across multiple runs).
- LFRM and ROCS almost correctly inferred the ground truth number of clusters (13-15 across multiple runs).

Experiments

- Drug-Protein Interaction Data
  - Bipartite graph (200 drug molecules, 150 target proteins)
  - Drug-drug similarity matrix, Protein-protein similarity matrix

Table 3: Link Prediction on Drug-Protein Interaction Data

Method	0-1 Test Error (%)	AUC
KPMF	16.65 ($\pm$ 0.36)	0.8734 ($\pm$ 0.0133)
LFRM	2.75 ($\pm$ 0.04)	0.9032 ($\pm$ 0.0156)
ROCS	**2.31 ($\pm$ 0.06)**	**0.9276 ($\pm$ 0.0142)**

- OCNMF and BCD are not applicable for bipartite graphs.
- LFRM here denotes ROCS without similarity information.
- KPMF takes into account the similarity information but does not assume overlapping clustering.

Experiments

- Lazega Lawyers Data
  - Directed graph, social networks (71 partners)
  - Each entry has features (gender, office-location, age, etc.)

Table 4: Link Prediction on Lazega-Lawyers Data

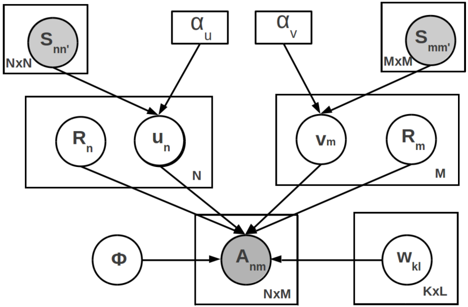
Method	0-1 Test Error (%)	AUC
OCNMF	35.36 ($\pm$20.71)	0.6388 ($\pm$0.1527)
KPMF	34.69 ($\pm$1.13)	0.7203 ($\pm$0.0229)
BCD	16.58 ($\pm$0.56)	0.7876 ($\pm$0.0168)
LFRM	14.05 ($\pm$ 2.04)	0.8025 ($\pm$ 0.0205)
ROCS	**12.98 ($\pm$ 0.32)**	**0.8248 ($\pm$ 0.01642)**

- Even weak similarity information can yield reasonable improvements in the prediction accuracy

Conclusions

Conclusions

- ROCS: a flexible model for modelling unipartite/bipartite graphs.
  - Each object can belong to multiple clusters (hard membership).
  - Nonparametric Bayesian approach.
  - Irrelevant objects can be dealt with in a principled manner.
  - Pairwise similarity between objects can be exploited to regularize the cluster memberships of objects.
  - Future work: make the model scalable.



References

- T. L. Griffiths and Z. Ghahramani. The Indian buffet process: An introduction and review. *JMLR*, 2011.
- K. Miller, T. Griffiths, and M. Jordan. Nonparametric latent feature models for link prediction. *NIPS*, 2009.
- M. Mørup and M. N. Schmidt. Bayesian community detection. *NeuralComputation*, 24(9):24342456, 2012.
- I. Psorakis, S. Roberts, M. Ebden, and B. Sheldon. Overlapping community detection using Bayesian non-negative matrix factorization. *PhysicalReviewE*, 2011.
- T. Zhou, H. Shan, A. Banerjee, and G. Sapiro. Kernelized probabilistic matrix factorization: Exploiting graphs and side information. In *SDM*, 2012.