Report January 6, 2016

Trong Dinh Thac Do

University of Technology Sydney

trongdinhthac.do@student.uts.edu.au

January 6 2016

Overview

1. Motivation
2. Literature review
3. Proposed Model - LFRMtimes
   - LFRMtimes
   - Inferences
   - Experiments
   - Future works

Motivation

A model for:

- Link prediction or community detection in social networks.
- Objects with coupling relations along time(serial coupling relations).
- Capturing some aspects(infinite, dynamic, mixed-membership and data-driven inference).

Motivation(cont.)

- Infinite: We do not have to define the number of communities before hand. It can prevent under or over fitting problem.
- Dynamic: The number of communities can change over time.
- Mixed-membership: one node can belongs to multiple communities.
- Data-driven inference: model bases on data only.

Litarature review

IRM

Infinite Relation Model(Kemp et al. 2006) cluster nodes into different groups based on their pairwise and directional binary interactions.

- Infinite.
- Not take into account changing with time.
- One node can only belong to one community.
- Data-driven.

dIRM

Dynamic Infinite Relation Model(Ishiguro et al. 2010)

- Infinite.
- Changing with time.
- One node can only belong to one community.
- Data-driven.

Litarature review(cont.)

MMSB

Mixed-Membership Block Model(Airoldi et al. 2008)

- Not Infinite.
- Not take into account changing with time.
- One node can belong to multiple communities.
- Data-driven.

LFRM

Latent Feature Relation Model(Miller et al. 2009)

- Infinite.
- Does not take into account changing with time.
- One node can belong to multiple communities.
- Data-driven.

Litarature review(cont.)

Sticky HDP-HMM

Sticky Hierachical Dirichlet Process - Hidden Markov Model(Fox et al. 2008)

- Infinite.
- Changing with time.
- One node can only belong to one community.
- Data-driven.

A model can capture all aspects:

- Infinite.
- Changing with time.
- One node can belong to multiple communities.
- Data-driven.

LFRMtimes

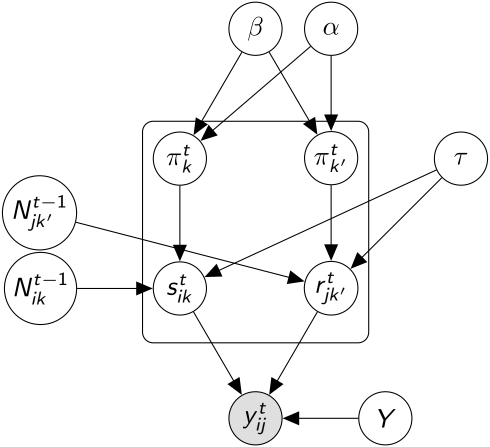

Figure: LFRMtimes. i,j $= 1:$ n, k,k' $= 1:$ K, t $= 1:$ T

LFRMtimes(cont.)

$\pi_k^t \sim BP(\alpha, \beta)$

$s_{ik}^t \sim BeP(\pi_k^t + \tau.N_{ik}^{t-1})$

$r_{jk'}^t \sim BeP(\pi_{k'}^t + \tau.N_{jk'}^{t-1})$

$w_{kk'}^t \sim \mathcal{N}(0, \sigma_{w^t}^2)$ for all k, k' which features k and k' are non-zero. $w_{kk'}^t$ is the weight that effects probability of there is link between entity i and entity j at time t.

$y_{ij}^t \sim \sigma(S_i W R_j^T) = \sigma(\sum_{k,k'} s_{ik}^t r_{jk'}^t w_{kk'}^t)$ for each observation. y_{ij}^t is the probability that there is a link between entity i and entity k at time t.

$\sigma(x) = \frac{1}{1+exp(-x)}$

LFRMtimes(cont.)

$$\pi_k^t | s_{1k}^t..s_{nk}^t \sim BP(\alpha + n + \tau, \frac{\alpha\beta + \frac{\tau}{n}.\sum_i N_{ik}\delta_i + \sum_i s_{ik}^t}{\alpha+n+\tau})$$

LFRMtimes can be described as below:

1. An Indian Buffet Restaurant will record all orders of all customers. In order word, they will keep all the history of customers, who ordered which disks.
2. When a customer come to the restaurant at time t, he will ask for the history of all the customers that are in the restaurant at that time who have the same as his disks in the past. Then he can choose:
   1. Get a Poi($\frac{\alpha\beta}{\alpha+n+\tau}$) new disk.
   2. Stay at history disk with probability $\frac{\frac{\tau}{n}.\sum_i N_{ik}\delta_i + \sum_i s_{ik}^t}{\alpha+n+\tau}$

LFRMtimes - Inferences

LFRMtimes is approximately inference via Markov Chain Monte Carlo(MCMC). (Will provide the details in the paper.)

LFRMtimes - Experiments

I got some results from Python Implementation of LFRMtimes. I will do some analysis and compare to previous model and provide next time.

LFRMtimes - Future works

- Using other inference solution to boost up the speed of algorithm like Variational Inference.
- Find out how the node information effects the latent communities of the node.

The End