

# Nonparametric Bayesian Matrix Factorization for Assortative Networks

Mingyuan Zhou

IROM Department, McCombs School of Business  
Department of Statistics and Data Sciences  
The University of Texas at Austin

23rd European Signal Processing Conference (EUSIPCO 2015)  
Nice, France, September 4, 2015

# Table of Contents

Introduction

Gamma process edge partition model

Example results

Improve EPM to model dissortativity

Conclusions

# Network community detection and link prediction

- ▶ We will focus on unweighted undirected relational networks, which can also be represented as binary symmetric network matrices.
- ▶ Community detection algorithms based on useful heuristics (see Fortunato, 2010 for a comprehensive review)
  - ▶ Examples:
    - ▶ Modularity maximization (Newman and Girvan, 2004)
    - ▶ Click percolation (Palla et al., 2005)
  - ▶ Restrictions:
    - ▶ Usually cannot be used to generate networks and predict missing edges (links)
    - ▶ Often need to tune the number of communities

# Network community detection and link prediction

- ▶ Generative network models
  - ▶ Model assumptions are clearly stated in a hierarchical model
  - ▶ Generate random networks
  - ▶ Detect latent communities
  - ▶ Detect community-community interactions
  - ▶ Predict missing edges (links)
  - ▶ Automatically infer the number of communities with nonparametric Bayesian priors

# Assortative and dissortative relational networks

- ▶ Assortativity:
  - ▶ Also known as Homophily.
  - ▶ A subset of nodes that are densely connected to each other but sparsely to the others are often considered to belong to the same community.
  - ▶ Example: in a social network, a community may consist of a group of closely related friends.

# Assortative and dissortative relational networks

- ▶ Dissortativity:
  - ▶ Also known as Stochastic Equivalence.
  - ▶ A subset of nodes that are sparsely connected to each other but densely connected to another subset of nodes are often considered to belong to the same community.
  - ▶ Example: in a predator-prey network, a community may consist of a group of animals that play similar roles in the ecosystem but not necessarily prey on each other.

# Probabilistic models for network analysis

- ▶ Latent class model
  - ▶ Stochastic blockmodel (Holland et al., 1983; Nowichi and Snijders, 2001)
  - ▶ Infinite relational model (Kemp et al., 2006)
  - ▶ Mixed-membership stochastic blockmodel (Airoldi et al., 2008)
- ▶ Latent factor model
  - ▶ Eigenmodel (Hoff, 2008)
  - ▶ Infinite latent feature relational model (Miller et al., 2009; Morup et al., 2011)
  - ▶ Community-affiliation graph model (Yang and Leskovec, 2012, 2014)
- ▶ Detection of disjoint or overlapping communities
- ▶ Interpretation of latent representations
- ▶ Prediction of missing edges

## Gamma process edge partition model

- ▶ Detect overlapping communities and predict missing edges
- ▶ As a latent factor model:
  - ▶ Connect each binary edge to a latent count via the Bernoulli-Poisson link
  - ▶ Factorize the latent count matrix
- ▶ As a latent class model:
  - ▶ Explicitly partition each observed edge into multiple latent communities
  - ▶ Implicitly assign a node to multiple communities based on how its edges are partitioned (overlapping communities)
- ▶ Designed to analyze assortative networks without considering community-community interactions
- ▶ Can be generalized to analyze disassortative networks by modeling community-community interactions



# Gamma process edge partition model

## ► Hierarchical model

$$b_{ij} = \mathbf{1}(m_{ij} \geq 1),$$

$$m_{ij} = \sum_{k=1}^K m_{ijk}, \quad m_{ijk} \sim \text{Po}(r_k \phi_{ik} \phi_{jk}),$$

$$\phi_{ik} \sim \text{Gam}(a_i, 1/c_i), \quad a_i \sim \text{Gam}(e_0, 1/f_0),$$

$$r_k \sim \text{Gam}(\gamma_0/K, 1/c_0), \quad \gamma_0 \sim \text{Gam}(e_1, 1/f_1).$$

## ► The Bernoulli-Poisson link

$$b_{ij} \sim \text{Bernoulli} \left[ 1 - \prod_{k=1}^K \exp(-r_k \phi_{ik} \phi_{jk}) \right].$$

## The Bernoulli-Poisson link

- ▶ Thresholding a count variable to obtain a binary variable

$$b = \mathbf{1}(m \geq 1), \quad m \sim \text{Po}(\lambda) \quad (1)$$

- ▶ Marginal likelihood of the Bernoulli-Poisson link

$$b \sim \text{Ber}\left(1 - e^{-\lambda}\right).$$

- ▶ The conditional posterior of the latent count  $m$  follows a truncated Poisson distribution, expressed as

$$(m|b, \lambda) \sim b \cdot \text{Po}_+(\lambda),$$

- ▶ Use rejection sampling to sample from the truncated Poisson distribution.
- ▶ Conceptual and computational advantages over the probit and logistic links.

## Overlapping community structures

- ▶ Edge partition model (EPM) under data augmentation

$$m_{ij} = \sum_k m_{ijk}, \quad m_{ijk} \sim \text{Po}(r_k \phi_{ik} \phi_{jk}).$$

- ▶  $m_{ijk}$  represents how often nodes  $i$  and  $j$  interact due to their affiliations with community  $k$ .
- ▶  $r_k \phi_{ik} \sum_{j \neq i} \phi_{jk}$  measures how strongly node  $i$  is affiliated with community  $k$ , and the latent count

$$m_{i \cdot k} := \sum_{j=i+1}^N m_{ijk} + \sum_{j=1}^{i-1} m_{jik} \quad (2)$$

represents how often node  $i$  is connected to the other nodes due to its affiliation with community  $k$ .

- ▶ Assign node  $i$  to multiple communities in  $\{k : m_{i \cdot k} \geq 1\}$ , or (hard) assign it to a single community using either  $\underset{k}{\operatorname{argmax}}(r_k \phi_{ik} \sum_{j \neq i} \phi_{jk})$  or  $\underset{k}{\operatorname{argmax}}(m_{i \cdot k})$ .

## Related model: community-affiliation graph model (AGM)

- ▶ A restricted version of the gamma process EPM, expressed as

$$b_{ij} \sim \text{Ber} \left[ 1 - e^{-\epsilon} \prod_k \exp(-r_k \phi_{ik} \phi_{jk}) \right],$$

where  $\epsilon \in \mathbb{R}^+$  and  $\phi_{ik} \in \{0, 1\}$ , could be considered as a nonparametric Bayesian generalization of the community-affiliation graph model (AGM) of (Yang and Leskovec, 2012, 2014).

- ▶ It is argued in AGM that all previous community detection methods would fail to detect communities with dense overlaps, due to a hidden assumption that a community's overlapping parts are less densely connected than its non-overlapping ones.
- ▶ The EPM does not make such a restrictive assumption; and beyond the AGM, it does not restrict  $\phi_{ik}$  to be binary.

## Data augmentation and marginalization

- ▶ Let  $m_{..k} := \sum_{i=1}^N \sum_{j=i+1}^N m_{ijk} = \frac{1}{2} \sum_{i=1}^N m_{i \cdot k}$
- ▶ Using the Poisson additive property, we have

$$m_{i \cdot k} \sim \text{Po} \left( r_k \phi_{ik} \sum_{j \neq i} \phi_{jk} \right), \quad m_{..k} \sim \text{Po} \left( r_k \frac{\sum_i \sum_{j \neq i} \phi_{ik} \phi_{jk}}{2} \right)$$

- ▶ Marginalizing out  $\phi_{ik}$  leads to

$$m_{ik..} \sim \text{NB} (a_i, p'_{ik}), \quad p'_{ik} := \frac{r_k \sum_{j \neq i} \phi_{jk}}{c_i + r_k \sum_{j \neq i} \phi_{jk}}.$$

- ▶ Marginalizing out  $r_k$  leads to

$$m_{..k} \sim \text{NB} (\gamma_0 / K, \tilde{p}_k), \quad \tilde{p}_k := \frac{\sum_i \sum_{j \neq i} \phi_{ik} \phi_{jk}}{2c_0 + \sum_i \sum_{j \neq i} \phi_{ik} \phi_{jk}}.$$

- ▶ Using these equations, we can develop closed-form Gibbs sampling update equations for all model parameters.

## Gibbs sampling

- ▶ **Sample**  $m_{ij}$ .  $(m_{ij}|-) \sim b_{ij} \text{Po}_+ \left( \sum_{k=1}^K r_k \phi_{ik} \phi_{jk} \right)$ .
- ▶ **Sample**  $m_{ijk}$ .  $(\{m_{ijk}\}_{k=1:K}|-) \sim \text{Mult} \left( m_{ij}; \frac{\{r_k \phi_{ik} \phi_{jk}\}_{k=1:K}}{\sum_{k'} r_{k'} \phi_{ik'} \phi_{jk'}} \right)$ .
- ▶ **Sample**  $a_i$ .

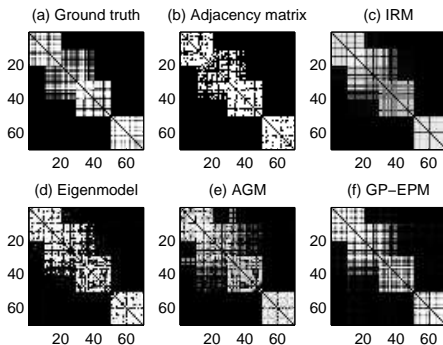
$$(\ell_{ik}|-) \sim \sum_{t=1}^{m_{i \cdot k}} \text{Ber} \left( \frac{a_i}{a_i + t - 1} \right),$$

$$(a_i|-) \sim \text{Gam} \left( e_0 + \sum_k \ell_{ik}, \frac{1}{f_0 + \sum_k \ln(1 - p'_{ik})} \right).$$

- ▶ **Sample**  $\phi_{ik}$ .  $(\phi_{ik}|-) \sim \text{Gam} \left( a_i + m_{i \cdot k}, \frac{1}{c_i + r_k \sum_{j \neq i} \phi_{jk}} \right)$ .
- ▶ **Sample**  $\gamma_0$ ,  $c_i$  and  $c_0$ .
- ▶ **Sample**  $r_k$ .  $(r_k|-) \sim \text{Gam} \left( \frac{\gamma_0}{K} + m_{..k}, \frac{1}{c_0 + \sum_i \sum_{j \neq i} \frac{1}{2} \phi_{ik} \phi_{jk}} \right)$ .

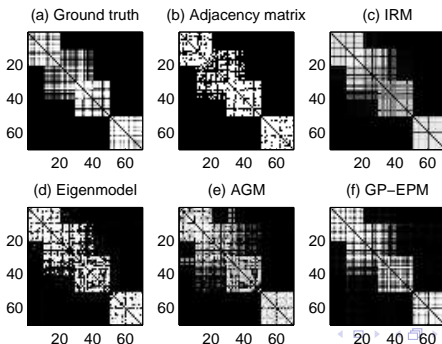
## Synthetic assortative network

- Four communities with dense intra-community connections. The 2nd community overlaps with both the 1st and 3rd ones.



**Figure:** Comparison of four algorithms' abilities to recover the ground-truth link probabilities using 80% of the pairs of nodes randomly selected from a synthetic relational network. The number of features for the Eigenmodel is set as  $K = 4$ .

- ▶ The infinite relational model (IRM) accurately captures the community structures but produces cartoonish blocks
- ▶ The Eigenmodel somewhat overfits the data
- ▶ The AGM produces some undesired artifacts
- ▶ The gamma process EPM (GP-EPM) provides a reconstruction that looks most similar to the ground truth.





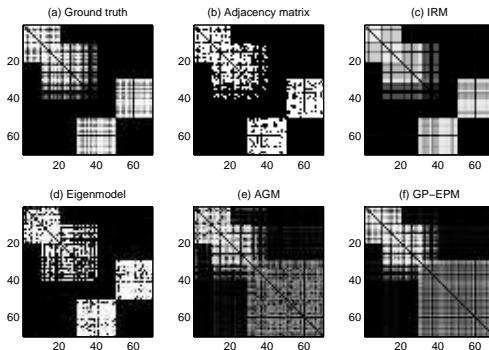
- Both the GP-EPM and Eigenmodel perform well and clearly outperform the IRM and AGM in missing link prediction, measured by both the area under the ROC curve and the area under the precision-recall (PR) curve.

**Table:** Comparison of four algorithms' abilities to predict missing edges of a synthetic assortative network. The number of features for the Eigenmodel is set as  $K = 4$ .

Model	AUC-ROC	AUC-PR
IRM	$0.9680 \pm 0.0073$	$0.8636 \pm 0.0448$
Eigenmodel	<b><math>0.9746 \pm 0.0066</math></b>	<b><math>0.9073 \pm 0.0236</math></b>
AGM	$0.9291 \pm 0.0184$	$0.8166 \pm 0.0470$
GP-EPM	<b><math>0.9746 \pm 0.0056</math></b>	<b><math>0.9042 \pm 0.0270</math></b>

## Synthetic dissortative network

- Four communities with dense intra-community or inter-community connections.



**Figure:** Comparison of four algorithms' abilities to recover the ground-truth link probabilities using 80% of the pairs of nodes randomly selected from a synthetic relational network that exhibits clear dissortativity.

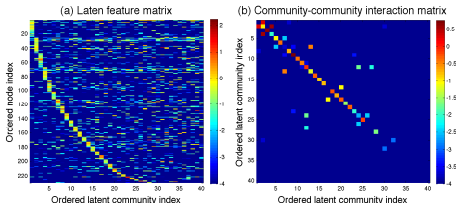
# EPM for dissortative networks (Zhou, AISTATS 2015)

- EPM that captures community-community interactions

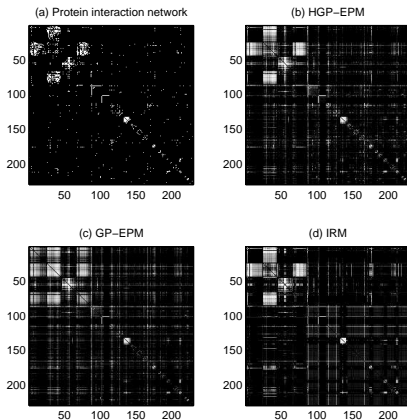
$$b_{ij} = \mathbf{1}(m_{ij} \geq 1), \quad m_{ij} = \sum_{k_1=1}^K \sum_{k_2=1}^K m_{ik_1 k_2 j},$$

$$m_{ik_1 k_2 j} \sim \text{Po}(\phi_{ik_1} \lambda_{k_1 k_2} \phi_{jk_2}),$$

- Use a relational hierarchical gamma process to support  $K = \text{infty}$
- The inferred latent feature matrix  $\{\phi_k\}$  and community-community interaction rate matrix  $\{\lambda_{k_1 k_2}\}$  for the improved EPM model on Protein230



## EPM for dissortative networks (Zhou, AISTATS 2015)



**Figure:** Comparison of three models on estimating the link probabilities for the Protein230 network using 80% of its node pairs.

# Conclusions

- ▶ The gamma process edge partition model (GP-EPM) provides an efficient and effective solution to model assortative relational networks
- ▶ The GP-EPM has limited ability to model dissortativity in relational networks.
- ▶ As in (Zhou, AISTATS 2015), to model dissortativity in relational networks, one may modify the model to capture community-community interactions.