Infinite Latent Attribute Model for Network Data

Konstantina Palla, David A. Knowles, Zoubin Ghahramani
CONTRIBUTION

A Bayesian non-parametric model for networks that have rich latent structure.
A network can be represented as a graph $\mathcal{G} = (\mathcal{V}, \mathcal{Y})$

- $\mathcal{V} = \{v_1, v_2, \ldots, v_N\}$, set of $N$ vertices
- $\mathcal{Y}$ is the set of edges - links between the nodes

Adjacency binary matrix $Y \in \{0, 1\}^{N \times N}$

- $y_{ij} = 1$, node $v_i$ and $v_j$ are connected
- $y_{ij} = 0$, node $v_i$ and $v_j$ are not connected

Example

![Adjacency matrix](image)

<table>
<thead>
<tr>
<th></th>
<th>$v_1$</th>
<th>$v_2$</th>
<th>$v_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$v_1$</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>$v_2$</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$v_1$</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
Motivating example

- Imagine a friendship network in a collegiate University. Friendships arise based on features each person has; a person might belong to clusters, e.g. college, department, sport team etc.

- Need a model that
  - Recovers the latent structure
  - Has good performance at link prediction

Categories

- Latent Class
- Latent Feature
Latent Class Models

▶ Assume a number of clusters $K$
▶ Each node $v_i$ belongs to a single cluster:
  $c_i \in \{1, \ldots, K\}$
▶ The link probability between two nodes:
  $\pi_{ij} = f(c_i, c_j)$

The Infinite Relational Model (IRM) by Kemp et al. (2006)

▶ $K \to \infty$
▶ Cluster assignments drawn from the Chinese Restaurant Process (CRP; Aldous, 1985)
  $c_1, c_2, \ldots, c_N \sim CRP(\gamma)$

For object $i$
  $P(c_i = A|c_1, \ldots, c_{i-1}) = \left\{ \begin{array}{ll}
\frac{n_A}{i-1+\gamma} & \text{if } n_A > 0 \\
\frac{\gamma}{i-1+\gamma} & \text{if } A \text{ is a new cluster}
\end{array} \right.$

▶ Link probability between members of each cluster pair
  $f(a, b)|\beta \sim Beta(\beta, \beta) \quad a, b \text{ are cluster labels}$
  $y_{ij}|c, \eta \sim Bernoulli(f(c_i, c_j))$
LATENT CLASS MODELS

- Assume a number of clusters $K$
- Each node $v_i$ belongs to a single cluster: $c_i \in \{1, \ldots, K\}$
- The link probability between two nodes:
  \[ \pi_{ij} = f(c_i, c_j) \]

The Infinite Relational Model (IRM) by Kemp et al. (2006)

- $K \to \infty$
- Cluster assignments drawn from the Chinese Restaurant Process (CRP; Aldous, 1985)
  \[ c_1, c_2, \ldots c_N \sim CRP(\gamma) \]
  For object $i$
  \[ P(c_i = A|c_1, \ldots, c_{i-1}) = \begin{cases} \frac{n_A}{i-1+\gamma} & \text{if } n_A > 0 \\ \frac{i-1+\gamma}{\gamma} & \text{if } A \text{ is a new cluster} \end{cases} \]

- Link probability between members of each cluster pair
  \[ f(a, b)|\beta \sim Beta(\beta, \beta) \quad a, b \text{ are cluster labels} \]
  \[ y_{ij}|c, \eta \sim Bernoulli(f(c_i, c_j)) \]
**Latent Class Models**

- Assume a number of clusters $K$
- Each node $v_i$ belongs to a single cluster: $c_i \in \{1, \ldots, K\}$
- The link probability between two nodes:
  \[
  \pi_{ij} = f(c_i, c_j)
  \]

**The Infinite Relational Model (IRM) by Kemp et al. (2006)**

- $K \to \infty$
- Cluster assignments drawn from the **Chinese Restaurant Process** (CRP; Aldous, 1985)
  \[
  c_1, c_2, \ldots c_N \sim CRP(\gamma)
  \]
  For object $i$
  \[
  P(c_i = A|c_1, \ldots, c_{i-1}) = \begin{cases} 
  \frac{n_A}{i-1+\gamma} & \text{if } n_A > 0 \\
  \frac{\gamma}{\gamma + i-1} & \text{if } A \text{ is a new cluster}
  \end{cases}
  \]

- Link probability between members of each cluster pair
  \[
  f(a, b)|\beta \sim Beta(\beta, \beta) \quad a, b \text{ are cluster labels}
  \]
  \[
  y_{ij}|c, \eta \sim Bernoulli(f(c_i, c_j))
  \]
LATENT CLASS MODELS

- Assume a number of clusters $K$
- Each node $v_i$ belongs to a single cluster: $c_i \in \{1, \ldots, K\}$
- The link probability between two nodes:
  \[ \pi_{ij} = f(c_i, c_j) \]

The Infinite Relational Model (IRM) by Kemp et al. (2006)

- $K \rightarrow \infty$
- Cluster assignments drawn from the Chinese Restaurant Process (CRP; Aldous, 1985)
  \[ c_1, c_2, \ldots c_N \sim CRP(\gamma) \]
  For object $i$
  \[ P(c_i = A|c_1, \ldots, c_{i-1}) = \begin{cases} \frac{n_A}{i-1+\gamma} & \text{if } n_A > 0 \\ \frac{\gamma}{i-1+\gamma} & \text{if } A \text{ is a new cluster} \end{cases} \]

- Link probability between members of each cluster pair
  \[ f(a, b)|\beta \sim Beta(\beta, \beta) \quad a, b \text{ are cluster labels} \]
  \[ y_{ij}|c, \eta \sim Bernoulli(f(c_i, c_j)) \]
**Latent Class Models**

- Assume a number of clusters $K$
- Each node $v_i$ belongs to a single cluster: $c_i \in \{1, \ldots, K\}$
- The link probability between two nodes:
  \[ \pi_{ij} = f(c_i, c_j) \]

The Infinite Relational Model (IRM) by Kemp et al. (2006)

- $K \to \infty$
- Cluster assignments drawn from the Chinese Restaurant Process (CRP; Aldous, 1985)
  \[ c_1, c_2, \ldots, c_N \sim CRP(\gamma) \]
  For object $i$
  \[ P(c_i = A|c_1, \ldots, c_{i-1}) = \begin{cases} \frac{n_A}{i-1+\gamma} & \text{if } n_A > 0 \\ \frac{\gamma}{i-1+\gamma} & \text{if } A \text{ is a new cluster} \end{cases} \]

- Link probability between members of each cluster pair
  \[ f(a, b)|\beta \sim Beta(\beta, \beta) \quad a, b \text{ are cluster labels} \]
  \[ y_{ij}|c, \eta \sim Bernoulli(f(c_i, c_j)) \]
**Latent Class Models**

**Restriction**
Latent Class models allow only *one* cluster per object.

**Example**
- A person might belong to more than one cluster at the same time, e.g. *college*, *department*, *sport team* etc.
- IRM needs a new cluster for each combination; e.g. *(college, department)*, *(college, sport team)*, *(college, department, sport team)* etc.
- Exponential explosion of clusters
LATENT FEATURE MODELS

- Each object possesses a vector of $M$ features: $z_i$.
  - each object might have *multiple* features.
- The link probability between two nodes $v_i$ and $v_j$

$$\pi_{ij} = f(z_i, z_j)$$
Each object possesses a vector of $M$ features: $z_i$.

- each object might have multiple features.

The link probability between two nodes $v_i$ and $v_j$ is given by

$$\pi_{ij} = f(z_i, z_j)$$
Each object possesses a vector of $M$ features: $z_i$.

- Each object might have multiple features.

The link probability between two nodes $v_i$ and $v_j$

$$\pi_{ij} = f(z_i, z_j)$$
Latent Feature Relational Model by Miller et al., 2010

- Each object is assigned a binary feature vector \( z_i \) of length \( M \to \infty \)
- Binary feature matrix \( Z \) of dimension \( N \times M \), drawn from an Indian Buffet Process (Griffiths & Ghahramani, 2005)

\[
Z | \alpha \sim IBP(\alpha)
\]

\[
p(z_{ik} = 1 | Z_{-ik}) = \frac{m_{-i,k} + \alpha}{N + \alpha}.
\]

For \( M \to \infty \): \( m_{-i,k} = 0 \), \( p(z_{ik} = 1 | Z_{-ik}) = 0 \).
Sample number of new features for the \( i \)th row from \( \text{Poisson}(\alpha/N) \).

- \( M \times M \) weight matrix \( W \) defines the weights between each pair of features

\[
w_{kl} | \sigma_w \sim N(0, \sigma_w^2)
\]

- Generate a link \( y_{ij} \) with probability

\[
\pi_{ij} = \sigma(z_i^T W z_j) = \sigma(\sum_{kl} w_{kl} z_{ik} z_{jl})
\]
**Restriction**
Both IRM and LFRM account for a flat clustering of the objects

**Motivating Example**

- The *College* feature might be divided into many subclusters e.g. “Trinity”, “King’s” etc.
- To model this, LFRM would have to represent each subcluster with a *new* feature. Cost in interpretability and computation.

**Proposal**
Explicit representation of the partitioning of each general feature into subclusters.
The Infinite Latent Attribute Model

Model Description

▶ Use IBP to assign $N$ objects to features (upper level of clustering)
  ▶ $z_i$ indicates the features object $i$ has on.
▶ Use CRP to assign the members of each feature to subclusters (lower level of clustering)
  ▶ If $i$ has a feature $m$ on, it belongs to a specific subcluster, $c_i^m$ of that feature.

Generative Model

\[
Z|\alpha \sim \text{IBP}(\alpha)
\]

\[
c^{(m)}|\gamma \sim \text{CRP}(\gamma), \text{ where } m \in \{1, \ldots M\}
\]

\[
w_{kk'}^{(m)}|\sigma_w \sim N(0, \sigma_w^2), \text{ where } k, k' \in \{1, \ldots, K^{(m)}\}
\]

\[
\pi_{ij} = \sigma \left( \sum_m z_{im}z_{jm}w_{c_i^m c_j^m}^{(m)} + s \right).
\]
THE INFINITE LATENT ATTRIBUTE MODEL

M features

N objects

Palla, Knowles, Ghahramani 10 / 17
THE INFINITE LATENT ATTRIBUTE MODEL

M features

N objects

$W^{(2)}$
Gibbs sampling to compute joint $P(Z, C, W|Y)$

Need to integrate over $W$ when considering new subclusters, so use Algorithm 8, Neal (2000).

Prior on $W$ is non conjugate to logistic likelihood. Use slice sampling (Neal, 2003).
**Relation to Other Models**

**Infinite Relational Model**
ILA with one column of $\mathbf{Z}$ switched on $\rightarrow$ IRM

$$
\pi_{ij}^{\text{IRM}} = f(c_i, c_j)
$$

**Latent Feature Relational Model**
ILA with only one subcluster in each feature $\rightarrow$ LFRM with a diagonal weight matrix

$$
\pi_{ij}^{\text{LFRM}} = \sigma \left( \sum_k z_{ik} z_{jk} w_{kk} + s \right)
$$

**Multiplicative Attribute Graph of Kim & Leskovec (2011)**
ILA with $M$ features (all on) and $K = 2$ subclusters in each feature $\rightarrow$ MAG

$$
\pi_{ij}^{\text{MAG}} = \prod_m w_{c_i^m c_j^m}
$$
RESULTS

NIPS Coauthorship network

We used the NIPS 1-17 coauthorship dataset (Globerson et al., 2007). We kept only the 234 most connected authors, ran 10 repeats, holding out 20% of the data. ILA 500 iterations, IRM and LFRM 1000 iterations.

<table>
<thead>
<tr>
<th></th>
<th>IRM</th>
<th>LFRM</th>
<th>ILA ($M = \infty$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test error (0-1 loss)</td>
<td>0.0440 ± 0.0014</td>
<td>0.0228 ± 0.0041</td>
<td>0.0106 ± 0.0007</td>
</tr>
<tr>
<td>Test log likelihood</td>
<td>−0.0859 ± 0.0043</td>
<td>−0.0547 ± 0.0079</td>
<td>−0.0318 ± 0.0094</td>
</tr>
<tr>
<td>AUC</td>
<td>0.9565 ± 0.0037</td>
<td>0.9631 ± 0.0150</td>
<td>0.9910 ± 0.0056</td>
</tr>
</tbody>
</table>

Link prediction

The lighter the entry, the more confident the model is that the corresponding authors would collaborate.
Genes Interaction Network

We used a subset of the interaction data by Jonikas et al. (2009). We used 156 genes, ran 10 repeats, holding out 20% of the data. ILA 500 iterations, IRM and LFRM 1000 iterations.

<table>
<thead>
<tr>
<th></th>
<th>IRM</th>
<th>LFRM</th>
<th>ILA ($M = \infty$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test error (0-1 loss)</td>
<td>0.3608 ± 0.0031</td>
<td>0.2661 ± 0.0086</td>
<td>0.0735 ± 0.0047</td>
</tr>
<tr>
<td>Test log likelihood</td>
<td>−0.4669 ± 0.0097</td>
<td>−0.4223 ± 0.0147</td>
<td>−0.2654 ± 0.0447</td>
</tr>
<tr>
<td>AUC</td>
<td>0.8654 ± 0.0057</td>
<td>0.8471 ± 0.0132</td>
<td>0.9924 ± 0.0037</td>
</tr>
</tbody>
</table>

We found features with subclusters which are significantly enriched for specific biological functions taken from the Gene Ontology database.
CONCLUSION

- ILA is able to recover rich hierarchical latent structure.
- Better empirical performance on two real network datasets than existing models using flat clusterings.
Thank you very much!
Gibbs sampling to compute joint $P(Z, C, W|Y)$

For current features $M$:

$$P(z_{im}|Z_{-im}, C_{-im}, W, Y) \propto P(z_{im}|Z_{-im})P(Y|z_{im} = 1, Z_{-im}, C_{-im}, W)$$

Need to integrate over $W$. Algorithm 8, Neal (2000).

Sample the number of new features $M_*^{(i)}$ for object $i$:

$$P(M_*^{(i)}|Z_{-M^{(i)}}, Y) \propto P(Y|M_*^{(i)}, Z_{-M^{(i)}})P(M^{(i)})$$

1. Integrating over the weights is non conjugate to logistic likelihood
2. Solution: Metropolis-Hastings
   - proposal: $Q(M_*^{(i)}, C^{(m*)}, W^{(m*)})$
   - $\alpha_{MH} = \frac{P(Y|M_*^{(i)}, C^{(m*)}, W^{(m*)})}{P(Y|M^{(i)}, C^{(m)}, W^{(m)})}$

Sample the weights:

$$P(w_{kl}^{(m)}|W_{-kl}, Z, C, Y) \propto P(Y|W, Z, C)P(w_{kl}^{(m)}|W_{-kl})$$

Prior is non conjugate to logistic likelihood. Use slice sampling (Neal, 2003).
Gibbs sampling to compute joint $P(Z, C, W|Y)$

For current features $M$:

$$P(z_{im}|Z_{-im}, C_{-im}, W, Y) \propto P(z_{im}|Z_{-im})P(Y|z_{im} = 1, Z_{-im}, C_{-im}, W)$$

Need to integrate over $W$. Algorithm 8, Neal (2000).

Sample the number of new features $M_*(i)$ for object $i$:

$$P(M_*(i)|Z_{-M(i)}, Y) \propto P(Y|M_*(i), Z_{-M(i)})P(M(i))$$

1. Integrating over the weights is non conjugate to logistic likelihood
2. Solution: Metropolis-Hastings
   - proposal: $Q(M_*(i), C^{(m*)}, W^{(m*)})$
   - $\alpha_{MH} = \frac{P(Y|M_*(i), C^{(m*)}, W^{(m*)})}{P(Y|M^{(i)}, C^{(m)}, W^{(m)})}$

Sample the weights:

$$P(w_{kl}^{(m)}|W_{-kl}, Z, C, Y) \propto P(Y|W, Z, C)P(w_{kl}^{(m)}|W_{-kl})$$

Prior is non conjugate to logistic likelihood. Use slice sampling (Neal, 2003).
Gibbs sampling to compute joint $P(Z, C, W|Y)$

For current features $M$:

$$P(z_{im}|Z_{-im}, C_{-im}, W, Y) \propto P(z_{im}|Z_{-im})P(Y|z_{im} = 1, Z_{-im}, C_{-im}, W)$$

Need to integrate over $W$. Algorithm 8, Neal (2000).

Sample the number of new features $M^{(i)}_*$ for object $i$:

$$P(M^{(i)}_*|Z_{-M^{(i)}}, Y) \propto P(Y|M^{(i)}_*, Z_{-M^{(i)}})P(M^{(i)}_*)$$

1. Integrating over the weights is non conjugate to logistic likelihood
2. Solution: Metropolis-Hastings
   - proposal: $Q(M^{(i)}_*, C^{(m*)}, W^{(m*)})$
   - $\alpha_{MH} = \frac{P(Y|M^{(i)}_*, C^{(m*)}, W^{(m*)})}{P(Y|M^{(i)}, C^{(m)}, W^{(m)})}$

Sample the weights:

$$P(w^{(m)}_{kl}|W_{-kl}, Z, C, Y) \propto P(Y|W, Z, C)P(w^{(m)}_{kl}|W_{-kl})$$

Prior is non conjugate to logistic likelihood. Use slice sampling (Neal, 2003).