Community Detection with Partially Observable Links and Node Attributes

Xiaokai Wei, Bokai Cao, Weixiang Shao, Chun-Ta Lu and Philip S. Yu
Partially Observable Network

- Networks are partially observable w.r.t. links and attributes under various scenarios:
  - Online social network: news users typically have no links and some users do not fill up their profile information. Besides, prudent users might set strict privacy setting which limits the visibility of their profile information and connections.
  - Bioinformatics: It is more costly to obtain attributes/interaction links for certain genes.
  - Terrorist network: some terrorists’ attribute and linkage information could be difficult to obtain.
Introduction

- Link and node attributes provide complementary information about nodes

- Challenge: nodes with only observable links and nodes with only observable attributes are not directly comparable
Introduction

- **POLNA:**
  - community detection with Partially Observable Links and Node Attributes

- **Idea:**
  - Learn a latent representation for each node via kernel matrix alignment based on either link structure or node content
  - Link-based representations and content-based representations of fully observable nodes interact with each other by co-regularization
Definition 1 Information Network An information network $G = (V, E, X)$ consists of $V$, the set of nodes, $E \subseteq V \times V$, the set of links, and the feature matrix $X = [x_1, x_2, \ldots, x_n]$ ($n = |V|$), where $x_i \in \mathbb{R}^D$ ($i = 1 \ldots n$) is the feature vector of node $v_i$.

Definition 2 Partially Observable Information Network In a partially observable information network $G = (V, E, X)$, each node belongs to one or two (overlapping) of the following sets: the set of nodes with observable links (denoted as $O^g$) and nodes with observable attributes (denoted as $O^a$). We further denote the set of nodes with both observable links and attributes as $O^s = O^g \cap O^a$. Note that $V = O^g \cup O^a$ since we assume each node has at least one source of information.

Let the number of nodes with observable links or observable attributes be $n^g = |O^g|$ and $n^a = |O^a|$, respectively. For more concise notations in the following sections, we assign local indices $1 \sim n^g$ to the nodes in $O^g$ and $1 \sim n^a$ to the nodes in $O^a$. A mapping function $\phi^g(i)$ (or $\phi^a(i)$) maps the local index $i \in \{1, \ldots, n^g\}$ (or $i \in \{1, \ldots, n^a\}$) in $O^g$ (or $O^a$) to its global index.
Matrix Alignment

- Measuring the similarity between two (vectorized) matrices

Matrix Alignment For two matrices $K_1 \in \mathbb{R}^{n \times n}$ and $K_2 \in \mathbb{R}^{n \times n}$ (assume $\|K_1\|_F > 0$ and $\|K_2\|_F > 0$), the alignment between $K_1$ and $K_2$ is defined as

$$\rho(K_1, K_2) = \frac{\text{Tr}(K_1 K_2)}{\|K_1\|_F \cdot \|K_2\|_F}$$  \hspace{1cm} (1)$$

where $\text{Tr}(\cdot)$ is the trace of a matrix.

Unnormalized Matrix Alignment For two matrices $K_1 \in \mathbb{R}^{n \times n}$ and $K_2 \in \mathbb{R}^{n \times n}$ (assume $\|K_1\|_F > 0$ and $\|K_2\|_F > 0$), the alignment between $K_1$ and $K_2$ is defined as

$$\rho(K_1, K_2) = \text{Tr}(K_1 K_2)$$  \hspace{1cm} (1)$$

Centered Matrix Alignment For two matrices $K_1 \in \mathbb{R}^{n \times n}$ and $K_2 \in \mathbb{R}^{n \times n}$ (assume $\|K_1\|_F > 0$ and $\|K_2\|_F > 0$), the centered alignment between $K_1$ and $K_2$ is defined as

$$\rho(K_1, K_2) = \text{Tr}(HK_1 HK_2 H)$$

$$= \text{Tr}(HK_1 HK_2)$$

where the second equation can be obtained by noting $HH = H$ and $\text{Tr}(AB) = \text{Tr}(BA)$ for arbitrary matrices $A, B \in \mathbb{R}^{n \times n}$.  \hspace{1cm} $H = I - \frac{1}{n} 11^T$
Kernel constructed from latent representations

Suppose \( K^g \in \mathbb{R}^{n^g \times n^g} \) and \( K^a \in \mathbb{R}^{n^a \times n^a} \) are kernel matrices computed from the latent representation \( U^g \) and \( U^a \) with Gaussian Kernel, respectively.

\[
K^g_{ij} = e^{-||U^g_i - U^g_j||^2}
\]
\[
K^a_{ij} = e^{-||U^a_i - U^a_j||^2}
\]

Learn latent representations that best align with the network structure and attribute kernel

Attribute-based representation

\[
\min_{U^g} f^g = - \text{Tr}(HL^gHK^g) + \lambda ||U^g||_F^2
\]

Network-based representation

\[
\min_{U^a} f^a = - \text{Tr}(HL^aHK^a) + \lambda ||U^a||_F^2
\]
Partial Co-regularization

- Learn a consensus from the link-based representation and attribute-based representation

- POLNA with hard constraint

\[
\begin{align*}
\min_{U^a, U^g, U^*} f &= f^g + f^a + f^\lambda \\
&= - \text{Tr}(HL^g HK^g) - \text{Tr}(HL^a HK^a) + \lambda \|U^*\|^2_F \\
\text{s.t. } U^g_{\phi^g-1}(i) &= U^a_{\phi^a-1}(i) = U^*_i
\end{align*}
\]

\[
\min_{U^*} f = - \text{Tr}(HL^g HK^g) - \text{Tr}(HL^a HK^a) + \lambda \|U^*\|^2_F
\]
**Partial Co-regularization**

- **POLNA with soft constraint**
  
  - Enforcing the link representation and attribute representation to be exactly the same might be too strict, since links and attributes could contain certain information reflecting their unique characteristics.

\[
\min_{\mathbf{U}^a, \mathbf{U}^g, \mathbf{U}^*} \quad f = f^g + f^a + f^s \\
= - \Tr(\mathbf{H} \mathbf{L}^g \mathbf{H} \mathbf{K}^g) + \lambda \left\| \mathbf{U}^g \right\|_F^2 \\
- \Tr(\mathbf{H} \mathbf{L}^a \mathbf{H} \mathbf{K}^a) + \lambda \left\| \mathbf{U}^a \right\|_F^2 \\
+ \alpha \sum_{i \in O^s} \left( \left\| \mathbf{U}_{\phi^g}(i) - \mathbf{U}_{\phi^a}(i) \right\|_F^2 + \left\| \mathbf{U}_{\phi^a}(i) - \mathbf{U}_{\phi^g}(i) \right\|_F^2 \right)
\]
Optimization

- **Gradient**

\[
\frac{\partial f^g}{\partial U^g_i} = - \sum_{j=1}^{n^g} ((H L^g H)_{ij} \cdot \frac{\partial K_{ij}^g}{\partial U^g_i} + (H L^g H)_{ji} \cdot \frac{\partial K_{ji}^g}{\partial U^g_i}) + 2\lambda U^g_i \\
= 4 \cdot \sum_{j=1}^{n^g} (((H L^g H) \odot K^g)_{ij} (U^g_i - U^g_j)) + 2\lambda U^g_i
\]
Optimization

- POLNA with hard constraint

**Algorithm 1** Optimization for POLNA with Hard Consensus Constraint (POLNA-HC)

1: Input: (Partially observable) feature matrix $X$, (Partially observable) network $G$, $\lambda$.
2: Output: Consensus representation $U^*$
3: Initialize: $U^g_i = \text{rand}(0, 1)$, $U^a_i = \text{rand}(0, 1)$, $U^* = 0$
4: Update $U^*$ with Eq (15) by L-BFGS
Algorithm 1 Alternating Optimization for POLNA with Soft Consensus Constraint (POLNA-SC)

1: Input: (Partially observable) feature matrix $X$, (Partially observable) network $G$, $\lambda$ and $\alpha$.
2: Output: Consensus representation $U^*$.
3: Initialize: $U^g_i = \text{rand}(0, 1)$, $U^a_i = \text{rand}(0, 1)$, $U^* = 0$
4: Set $t = 1$
5: while not converged do
6:     Find the optimal $U^g$ by L-BFGS with Eq (16)
7:     if $t = 1$ then
8:         $U^*_i = U^g_{\phi^g-1(i)}$, $\forall i \in O^s$
9:     end if
10:    Find the optimal $U^a$ by L-BFGS with Eq (17)
11:    $U^*_i = (U^g_{\phi^g-1(i)} + U^a_{\phi^a-1(i)})/2$, $\forall i \in O^s$
12:    $t = t + 1$
13: end while
14: $U^*_i = U^g_{\phi^g-1(i)}$, $\forall i \in O^g/O^s$
15: $U^*_i = U^a_{\phi^a-1(i)}$, $\forall i \in O^a/O^s$
Evaluation

- **Datasets**
  - Citeseer: citation network
  - Cora: citation network

- **Evaluation Metrics**
  - Accuracy
  - NMI

<table>
<thead>
<tr>
<th>Table 1: Statistics of two datasets</th>
<th>Citeseer</th>
<th>Cora</th>
</tr>
</thead>
<tbody>
<tr>
<td># of instances</td>
<td>3312</td>
<td>2708</td>
</tr>
<tr>
<td># of links</td>
<td>4598</td>
<td>5429</td>
</tr>
<tr>
<td># of attributes</td>
<td>3703</td>
<td>1433</td>
</tr>
<tr>
<td># of classes</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>
Evaluation

- **Methods:**
  - NMF: Non-negative Matrix Factorization on combined network
  - Spectral Clustering on combined network
  - POLNA-HC: POLNA with hard constraint
  - POLNA-SC: POLNA with soft constraint

- **Combined network for baselines:**
  - Create kNN network from observable node attributes and combine it with observable link structure
Evaluation

- **Accuracy with different observable rate**

![Graphs showing accuracy for different observable rates for Citeseer and Cora datasets.](image_url)

- Citeseer
- Cora

- POLNA-SC
- POLNA-HC
- SNMF
- Spectral
Evaluation

- NMI with different observable rate

![Graphs showing NMI with different observable rate for Citeseer and Cora datasets.](image-url)
Convergence Analysis

- Accuracy and objective function w.r.t. # iterations

![Graphs showing convergence analysis for Citeseer and Cora datasets.](image)
Sensitivity Analysis

- **POLNA-HC w.r.t. λ**

**Citeeseer**

**Cora**
Sensitivity Analysis

- **POLNA-SC w.r.t. $\lambda$**

![Graphs showing sensitivity analysis for POLNA-SC with respect to $\lambda$ on Citeseer and Cora datasets.](image_url)
Sensitivity Analysis

- **POLNA-SC w.r.t. $\alpha$**

![Graphs showing sensitivity analysis for different datasets and values of $\alpha$.](image-url)
Questions?