Community Detection with Partially Observable Links and Node Attributes

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

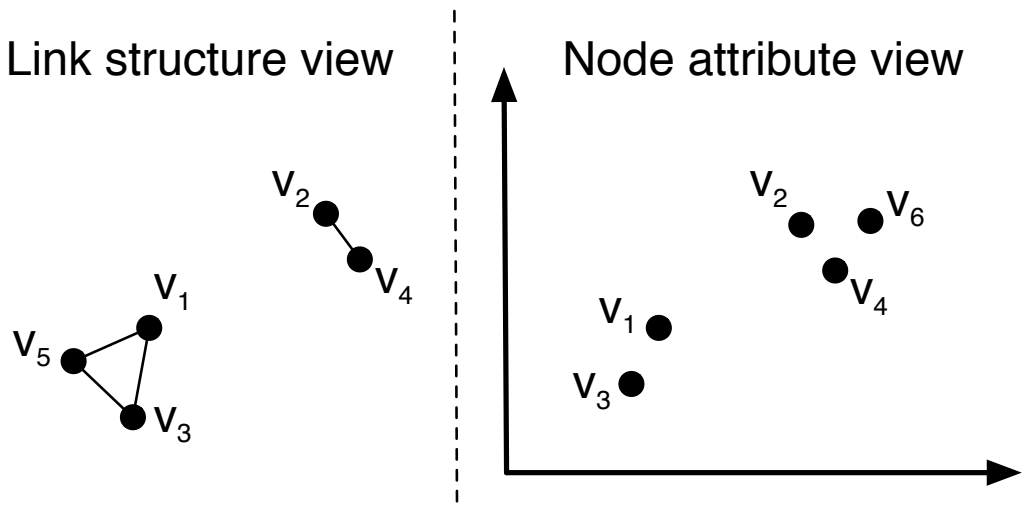
UIC UNIVERSITY OF ILLINOIS AT CHICAGO

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 does 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

Formulation

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 $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]$ ($n = |V|$), where $\mathbf{x}_i \in \mathbb{R}^D$ ($i = 1 \ldots n$) 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 $\mathbf{K}_1 \in \mathbb{R}^{n\times n}$ and $\mathbf{K}_2 \in \mathbb{R}^{n\times n}$ (assume $||\mathbf{K}_1||_F > 0$ and $||\mathbf{K}_2||_F > 0$), the alignment between $\mathbf{K}_1$ and $\mathbf{K}_2$ is defined as

$$\rho(\mathbf{K}_1, \mathbf{K}_2) = \frac{\mathbf{Tr}(\mathbf{K}_1\mathbf{K}_2)}{||\mathbf{K}_1||_F \cdot ||\mathbf{K}_2||_F} \tag{1}$$

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

Unnormalized Matrix Alignment For two matrices $\mathbf{K}_1 \in \mathbb{R}^{n\times n}$ and $\mathbf{K}_2 \in \mathbb{R}^{n\times n}$ (assume $||\mathbf{K}_1||_F > 0$ and $||\mathbf{K}_2||_F > 0$), the alignment between $\mathbf{K}_1$ and $\mathbf{K}_2$ is defined as

$$\rho(\mathbf{K}_1, \mathbf{K}_2) = \mathbf{Tr}(\mathbf{K}_1\mathbf{K}_2) \tag{1}$$

Centered Matrix Alignment For two matrices $\mathbf{K}_1 \in \mathbb{R}^{n\times n}$ and $\mathbf{K}_2 \in \mathbb{R}^{n\times n}$ (assume $||\mathbf{K}_1||_F > 0$ and $||\mathbf{K}_2||_F > 0$), the centered alignment between $\mathbf{K}_1$ and $\mathbf{K}_2$ is defined as

$$\begin{aligned}\rho(\mathbf{K}_1, \mathbf{K}_2) =& \mathbf{Tr}(\mathbf{H}\mathbf{K}_1\mathbf{H}\mathbf{H}\mathbf{K}_2\mathbf{H}) \\ =& \mathbf{Tr}(\mathbf{H}\mathbf{K}_1\mathbf{H}\mathbf{K}_2)\end{aligned}$$

where the second equation can be obtained by noting $\mathbf{H}\mathbf{H} = \mathbf{H}$ and $\mathbf{Tr}(\mathbf{AB}) = \mathbf{Tr}(\mathbf{BA})$ for arbitrary matrices $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n\times n}$. $\mathbf{H} = \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T$

Representation Learning with Kernel Matrix Alignment

- Kernel constructed from latent representations

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

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

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

Attribute-based representation

$$\min_{\mathbf{U}^g} \; f^g = -\operatorname{Tr}(\mathbf{H}\mathbf{L}^g\mathbf{H}\mathbf{K}^g) + \lambda||\mathbf{U}^g||_F^2$$

Network-based representation

$$\min_{\mathbf{U}^a} \; f^a = -\operatorname{Tr}(\mathbf{H}\mathbf{L}^a\mathbf{H}\mathbf{K}^a) + \lambda||\mathbf{U}^a||_F^2$$

Partial Co-regularization

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

- POLNA with hard constraint

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

$$
\begin{aligned}
\min_{\mathbf{U}^*} f = & -\operatorname{Tr}(\mathbf{HL}^g\mathbf{HK}^g) \\
& -\operatorname{Tr}(\mathbf{HL}^a\mathbf{HK}^a) \\
& +\lambda||\mathbf{U}^*||_F^2
\end{aligned}
$$

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

$$
\begin{aligned}
\min_{\mathbf{U}^a,\mathbf{U}^g,\mathbf{U}^*} \quad f =& f^g + f^a + f^s \\
=& -\operatorname{Tr}(\mathbf{HL}^g\mathbf{HK}^g) + \lambda||\mathbf{U}^g||_F^2 \\
& -\operatorname{Tr}(\mathbf{HL}^a\mathbf{HK}^a) + \lambda||\mathbf{U}^a||_F^2 \\
& + \alpha \sum_{i\in O^s} (||\mathbf{U}^g_{\phi^{g-1}(i)} - \mathbf{U}^*_i||_F^2 + ||\mathbf{U}^a_{\phi^{a-1}(i)} - \mathbf{U}^*_i||_F^2)
\end{aligned}
$$

Optimization

- Gradient

$$\frac{\partial f^g}{\partial \mathbf{U}_i^g} = -\sum_{j=1}^{n^g}((\mathbf{H}\mathbf{L}^g\mathbf{H})_{ij} \cdot \frac{\partial K_{ij}^g}{\partial \mathbf{U}_i^g} + (\mathbf{H}\mathbf{L}^g\mathbf{H})_{ji} \cdot \frac{\partial K_{ji}^g}{\partial \mathbf{U}_i^g}) + 2\lambda \mathbf{U}_i^g$$

$$= 4 \cdot \sum_{j=1}^{n^g}(((\mathbf{H}\mathbf{L}^g\mathbf{H}) \odot \mathbf{K}^g)_{ij} \left(\mathbf{U}_i^g - \mathbf{U}_j^g\right)) + 2\lambda \mathbf{U}_i^g$$

Optimization

- POLNA with hard constraint

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

1: Input: (Partially observable) feature matrix $\mathbf{X}$, (Partially observable) network G, λ.
2: Output: Consensus representation $\mathbf{U}^*$
3: Initialize: $\mathbf{U}_i^g = rand(0,1)$, $\mathbf{U}_i^a = rand(0,1)$, $\mathbf{U}^* = \mathbf{0}$
4: Update $\mathbf{U}^*$ with Eq (15) by L-BFGS

Optimization

- POLNA with soft constraint
  - Alternating optimization

Algorithm 1 Alternating Optimization for POLNA with Soft Consensus Constraint (POLNA-SC)

Input: (Partially observable) feature matrix $\mathbf{X}$, (Partially observable) network G, λ and α.
Output: Consensus representation $\mathbf{U}^*$.
Initialize: $\mathbf{U}_i^g = rand(0,1)$, $\mathbf{U}_i^a = rand(0,1)$, $\mathbf{U}^* = \mathbf{0}$
Set $t = 1$
while not converged **do**
    Find the optimal $\mathbf{U}^g$ by L-BFGS with Eq (16)
    if $t = 1$ **then**
        $\mathbf{U}_i^* = \mathbf{U}^g_{\phi^{g-1}(i)}, \forall i \in O^s$
    end if
    Find the optimal $\mathbf{U}^a$ by L-BFGS with Eq (17)
    $\mathbf{U}_i^* = (\mathbf{U}^g_{\phi^{g-1}(i)} + \mathbf{U}^a_{\phi^{a-1}(i)})/2, \quad \forall i \in O^s$
    $t = t + 1$
end while
$\mathbf{U}_i^* = \mathbf{U}^g_{\phi^{g-1}(i)}, \quad \forall i \in O^g/O^s$
$\mathbf{U}_i^* = \mathbf{U}^a_{\phi^{a-1}(i)}, \quad \forall i \in O^a/O^s$

Evaluation

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

Table 1: Statistics of two datasets

Statistics	Citeseer	Cora
# of instances	3312	2708
# of links	4598	5429
# of attributes	3703	1433
# of classes	6	7

- Evaluation Metrics
  - Accuracy
  - NMI

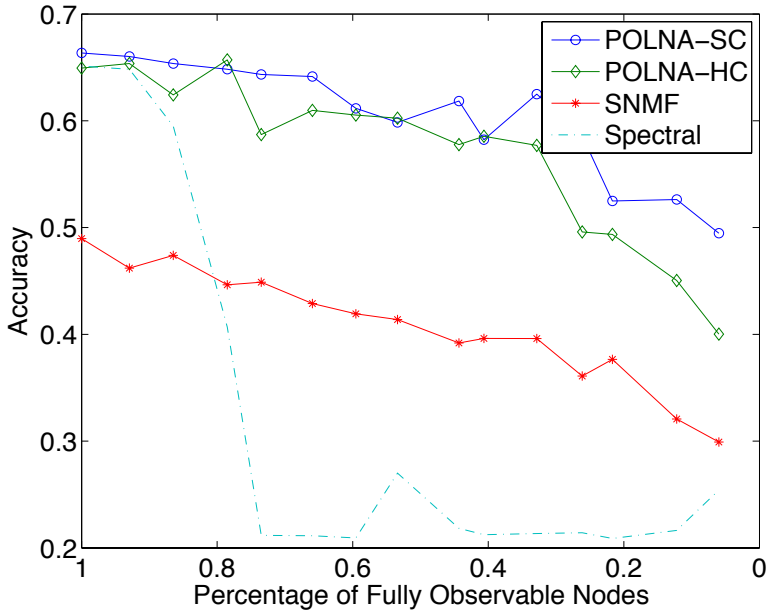
Evaluation

- 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

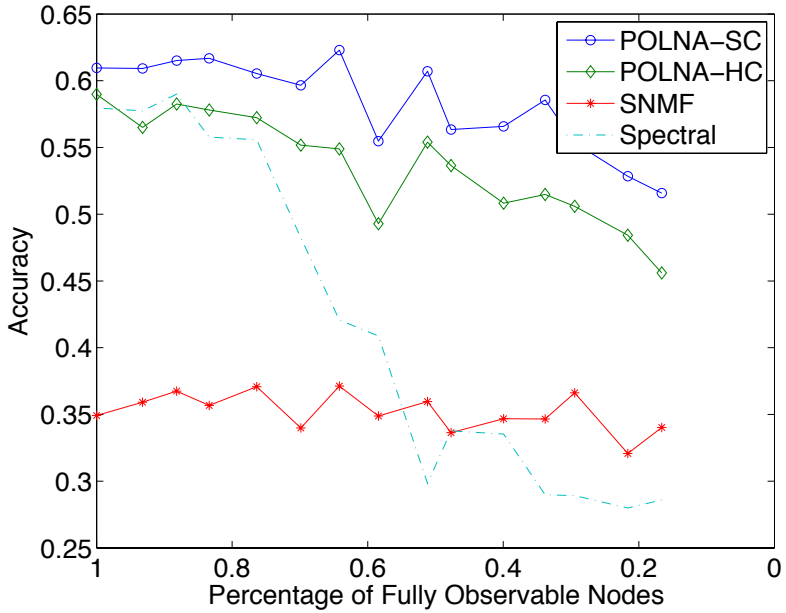
- Combined network for baselines:
  - Create kNN network from observable node attributes and combine it with observable link structure

Evaluation

- Accuracy with different observable rate



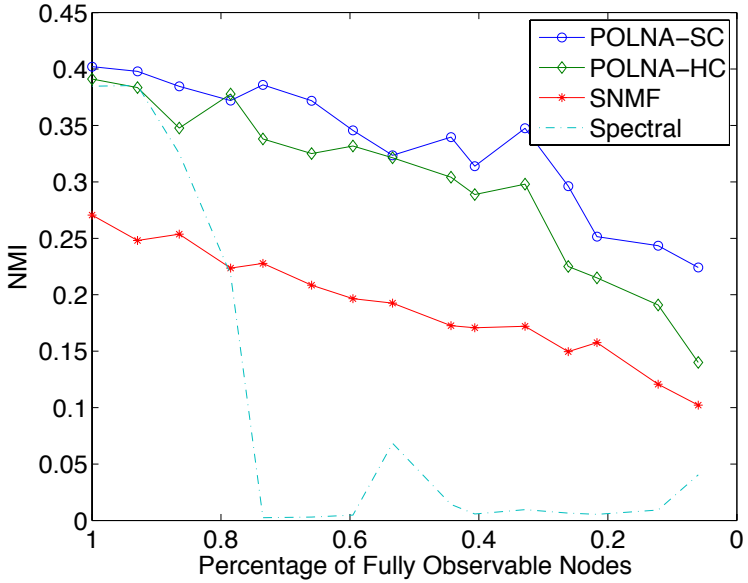
Citeseer



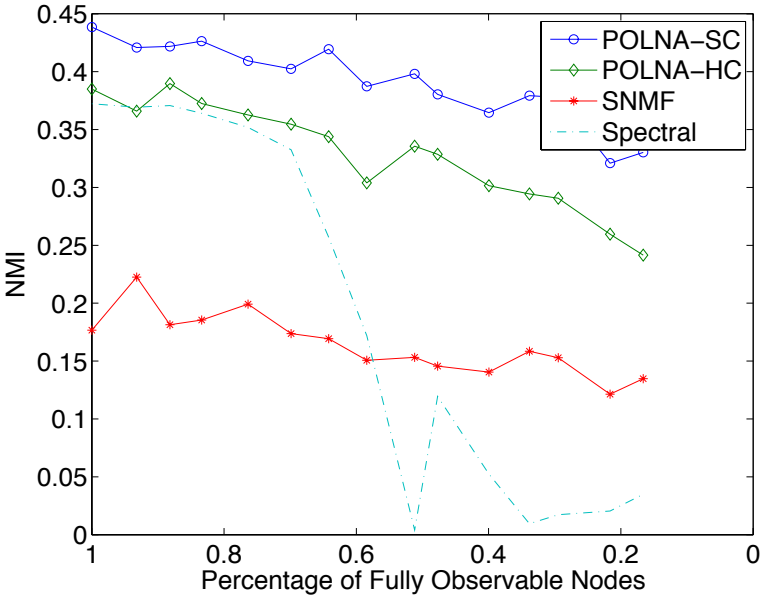
Cora

Evaluation

- NMI with different observable rate



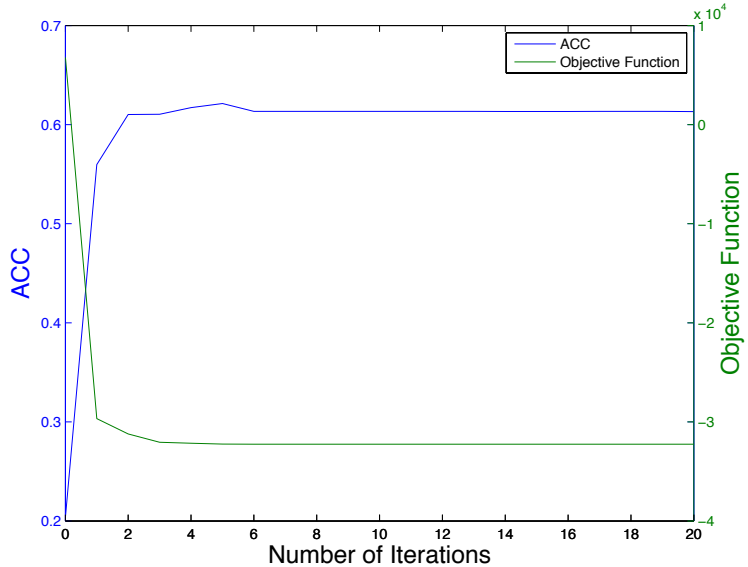
Citeseer



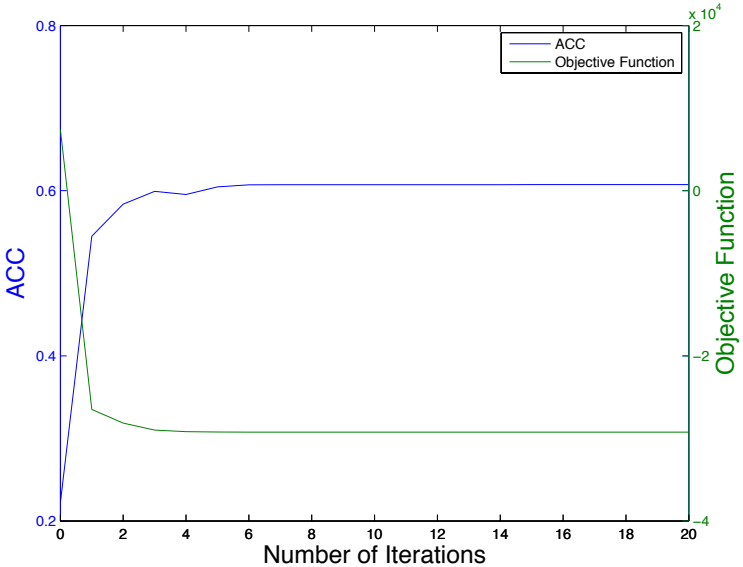
Cora

Convergence Analysis

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



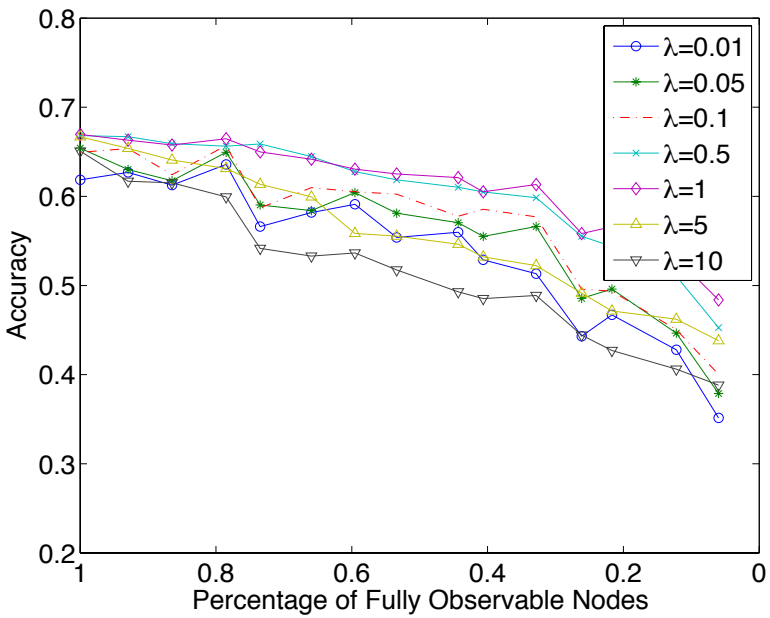
Citeseer



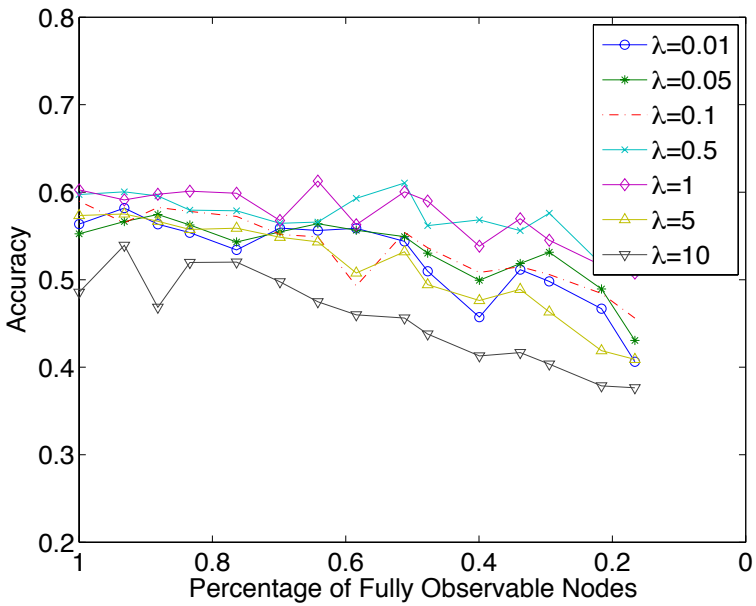
Cora

Sensitivity Analysis

- POLNA-HC w.r.t. λ



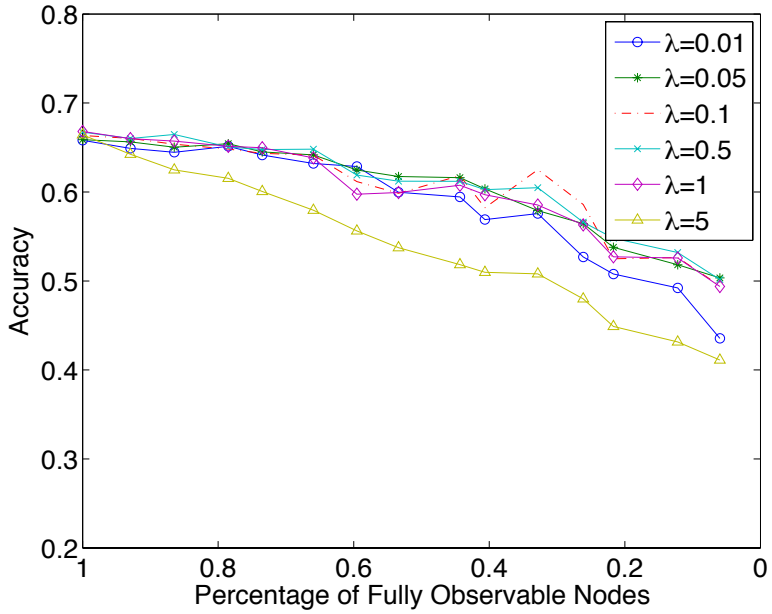
Citeseer



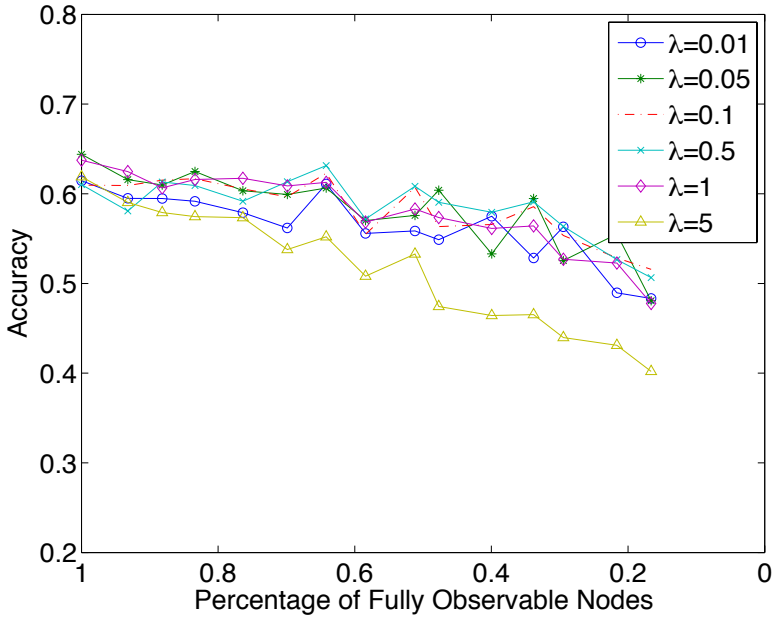
Cora

Sensitivity Analysis

- POLNA-SC w.r.t. λ



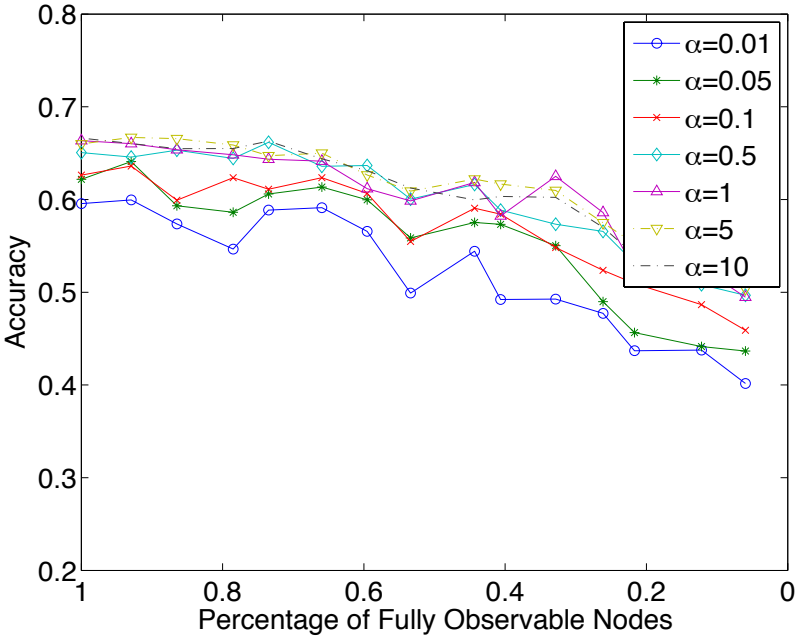
Citeseer



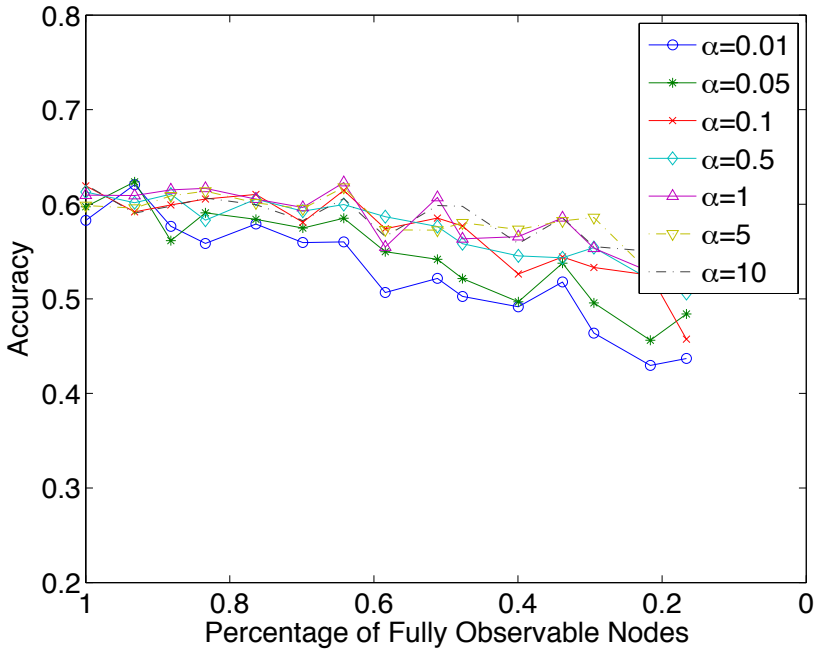
Cora

Sensitivity Analysis

- POLNA-SC w.r.t. α



Citeseer



Cora

Questions?