Transfer Learning for Community Detection

Zexi Huang

Supervisor: Asst. Prof. Sinno Jialin Pan

Yingcai Honors College, University of Electronic Science and Technology of China
Computational Intelligence Lab, Nanyang Technological University

June 6, 2018

Outline

Background

- Community detection:
  - Objective: to find densely connected components in networks that in reality represent cohesive communities [1].
  - Applications: setting up efficient and effective recommendation systems [2], handling navigation queries [3], establishing dedicated mirror server [4], etc.
- Community detection in multiplex networks:
  - Multiplex network: the same set of nodes are connected by multiple types of edges [5]. Also known as multilayer network [6] or multidimensional network [7].
  - Objectives: to find a community partition for all layers [8, 9, 10] or a subset of layers [11].

Contribution

- First to formulate the multiplex community detection problem as to to refine community detection results in some layer with transferred knowledge from other layers.
- Design a representation-based multiplex community detection framework, and implement it with an extended symmetric non-negative matrix factorization approach for learning representations and k-means for clustering representations.
- Our algorithm outperforms the other representation-based community detection algorithms, especially when the target layer that receives transferred knowledge is noisy (e.g. much sparser than other layers).

Framework

Problem

Given a N-layer multiplex network $G(V, E_1, ..., E_N)$ with adjacency matrices $\mathbf{A}_1, ..., \mathbf{A}_N$, find distinct community partition in some layer K', $\mathtt{C}_{K'}$, so that both transferred knowledge from other layers and layer-specific features of K'th layer are exploited.

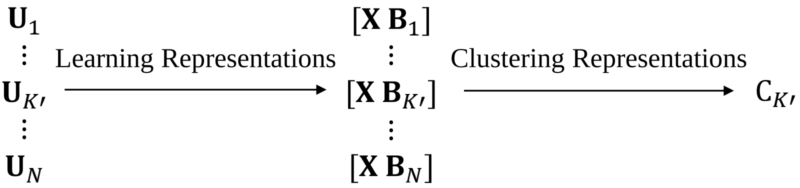

Figure 1: The overview of our framework.

Learning Representations

Stochastic Block Model [12]

Utility matrix:

$$\mathbf{U} = \mathbf{A} \tag{1}$$

Learning objective:

$$\min_{\mathbf{H} \geqslant 0} L = ||\mathbf{A} - \mathbf{H}\mathbf{H}^T||_F^2 \tag{2}$$

Objective

$$\begin{aligned} \min_{\mathbf{X} \geqslant 0, \mathbf{B}_1, \ldots, \mathbf{B}_N \geqslant 0} L &= \sum_{K=1}^{N} ||\mathbf{A}_K - [\mathbf{X}\ \mathbf{B}_K][\mathbf{X}\ \mathbf{B}_K]^T||_F^2 \\ &= \sum_{K=1}^{N} ||\mathbf{A}_K - \mathbf{X}\mathbf{X}^T - \mathbf{B}_K\mathbf{B}_K^T||_F^2 \end{aligned} \tag{3}$$

Learning Representations

Theorem 1

The sum of reconstruction error, L, can't be minimized to zero.

Proof

We only need to prove that $||\mathbf{A}_K - \mathbf{X}\mathbf{X}^T - \mathbf{B}_K\mathbf{B}_K^T||_F^2 > 0$, which is equivalent to $\mathbf{A}_K \neq \mathbf{X}\mathbf{X}^T + \mathbf{B}_K$. We prove this by showing that $\mathbf{X}\mathbf{X}^T + \mathbf{B}_K\mathbf{B}_K^T$ is positive semi-definite and that $\mathbf{A}_K$ is not. First, for any non-zero vector $s \in \mathbb{R}^{n\times 1}$, we have

$$\begin{aligned} s^T(\mathbf{X}\mathbf{X}^T + \mathbf{B}_K\mathbf{B}_K^T)s &= s^T\mathbf{X}\mathbf{X}^T s + s^T\mathbf{B}_K\mathbf{B}_K^T s \\ &= g^T g + h^T h \\ &= ||g||^2 + ||h||^2 \geqslant 0 \end{aligned} \tag{4}$$

where $g = \mathbf{X}^T s, h = \mathbf{B}_K^T s$. This shows that $\mathbf{X}\mathbf{X}^T + \mathbf{B}_K\mathbf{B}_K^T$ is positive semi-definite.

Learning Representations

Proof (cont'd)

Then, for undirected networks with no self-loops, we have $\mathbf{A}_K$ to be symmetric and

$$\sum_{i=1}^{n} \lambda_i = \mathrm{Tr}(\mathbf{A_K}) = \sum_{i=1}^{n} \mathbf{A}_{Kii} = 0 \tag{5}$$

where $\lambda_1, ..., \lambda_n$ are the n eigenvalues of $\mathbf{A}_K$. Therefore, either there is at least some $\lambda_i < 0$ or $\lambda_1 = ... = \lambda_n = 0$. However, since $\mathbf{A}_K$ is symmetric, $\lambda_1 = ... = \lambda_n = 0$ directly leads to $\mathbf{A}_K = 0$. Thus, $\lambda_i < 0$ holds, which means $\mathbf{A}_K$ can't be positive semi-definite.

Learning Representations

Objective

$$\min_{\mathbf{X}\geqslant 0,\mathbf{B}_1,\ldots,\mathbf{B}_N\geqslant 0} L = \sum_{K=1}^{N} ||\mathbf{A}_K - \mathbf{X}\mathbf{X}^T - \mathbf{B}_K\mathbf{B}_K^T||_F^2 \tag{6}$$

Karush-Kuhn-Tucker Conditions

Introduce the multipliers $\mathbf{\Lambda}, \mathbf{\Gamma}_K$ and the Lagrangian

$$\mathcal{L}(\mathbf{X}, \mathbf{B}_K) = L - \mathrm{Tr}(\mathbf{\Lambda}\mathbf{X}^T) - \sum_{K=1}^{N} \mathrm{Tr}(\mathbf{\Gamma}_K\mathbf{B}_K^T) \tag{7}$$

From the stationarity of Karush-Kuhn-Tucker conditions we have

$$\frac{\partial \mathcal{L}}{\partial \mathbf{X}} = \frac{\partial L}{\partial \mathbf{X}} - \mathbf{\Lambda} = 0, \frac{\partial \mathcal{L}}{\partial \mathbf{B}_K} = \frac{\partial L}{\partial \mathbf{B}_K} - \mathbf{\Gamma}_K = 0 \tag{8}$$

Learning Representations

Karush-Kuhn-Tucker Conditions (cont'd)

which leads to

$$\mathbf{\Lambda} = 4N\mathbf{X}\mathbf{X}^T\mathbf{X} + 4\sum_{K=1}^{N}(\mathbf{B}_K\mathbf{B}_K^T - \mathbf{A}_K)\mathbf{X} \tag{9}$$

$$\mathbf{\Gamma}_K = 4\mathbf{B}_K\mathbf{B}_K^T\mathbf{B}_K + 4(\mathbf{X}\mathbf{X}^T - \mathbf{A}_K)\mathbf{B}_K \tag{10}$$

Then, from the complementary slackness of Karush-Kuhn-Tucker conditions

$$\mathbf{\Lambda}_{ij}\mathbf{X}_{ij} = 0, \mathbf{\Gamma}_{Kij}\mathbf{B}_{Kij} = 0 \tag{11}$$

Substituting Eq.9 and Eq.10 into Eq.11, we have

Learning Representations

Karush-Kuhn-Tucker Conditions (cont'd)

$$(-\sum_{K=1}^{N}(\mathbf{A}_K - \mathbf{B}_K\mathbf{B}_K^T)\mathbf{X} + N\mathbf{X}\mathbf{X}^T\mathbf{X})_{ij}\mathbf{X}_{ij} = 0 \quad (12)$$

$$(-(\mathbf{A}_K - \mathbf{X}\mathbf{X}^T)\mathbf{B}_K + \mathbf{B}_K\mathbf{B}_K^T\mathbf{B}_K)_{ij}\mathbf{B}_{Kij} = 0 \quad (13)$$

Alternating Multiplicative Update Rules

$$\mathbf{X}_{ij} \leftarrow \mathbf{X}_{ij}\sqrt{\frac{((\sum_{K=1}^{N}(\mathbf{A}_K - \mathbf{B}_K\mathbf{B}_K^T))^+\mathbf{X})_{ij}}{((\sum_{K=1}^{N}(\mathbf{A}_K - \mathbf{B}_K\mathbf{B}_K^T))^-\mathbf{X})_{ij} + (N\mathbf{X}\mathbf{X}^T\mathbf{X})_{ij}}} \quad (14)$$

$$\mathbf{B}_{Kij} \leftarrow \mathbf{B}_{Kij}\sqrt{\frac{((\mathbf{A}_K - \mathbf{X}\mathbf{X}^T)^+\mathbf{B}_K)_{ij}}{((\mathbf{A}_K - \mathbf{X}\mathbf{X}^T)^-\mathbf{B}_K)_{ij} + (\mathbf{B}_K\mathbf{B}_K^T\mathbf{B}_K)_{ij}}} \quad (15)$$

Learning Representations

Theorem 2

The limiting solutions of the update rules in Eq.14 and Eq.15 satisfy Karush-Kuhn-Tucker conditions in Eq.12 and Eq.13.

Proof

We only need to prove that Eq.14 satisfies Eq.12 at convergence. At convergence, we have $\mathbf{X}^{\infty} = \mathbf{X}^{t+1} = \mathbf{X}^{t} = \mathbf{X}$. Then, from Eq.14

$$\mathbf{X}_{ij} = \mathbf{X}_{ij}\sqrt{\frac{((\sum_{K=1}^{N}(\mathbf{A}_K - \mathbf{B}_K\mathbf{B}_K^T))^{+}\mathbf{X})_{ij}}{((\sum_{K=1}^{N}(\mathbf{A}_K - \mathbf{B}_K\mathbf{B}_K^T))^{-}\mathbf{X})_{ij} + (N\mathbf{X}\mathbf{X}^T\mathbf{X})_{ij}}} \quad (16)$$

Square both sides, multiply it by the denominator, and transpose it, we get

$$(((\sum_{K=1}^{N}(\mathbf{A}_K-\mathbf{B}_K\mathbf{B}_K^T))^{-}-(\sum_{K=1}^{N}(\mathbf{A}_K-\mathbf{B}_K\mathbf{B}_K^T))^{+})\mathbf{X}+N\mathbf{X}\mathbf{X}^T\mathbf{X})_{ij}\mathbf{X}_{ij}^2 = 0 \quad (17)$$

Learning Representations

Proof (cont'd)

Note that

$$(\sum_{K=1}^{N}(\mathbf{A}_K - \mathbf{B}_K\mathbf{B}_K^T))^+ - (\sum_{K=1}^{N}(\mathbf{A}_K - \mathbf{B}_K\mathbf{B}_K^T))^- = \sum_{K=1}^{N}(\mathbf{A}_K - \mathbf{B}_K\mathbf{B}_K^T) \tag{18}$$

Therefore, Eq.17 reduces to

$$(-\sum_{K=1}^{N}(\mathbf{A}_K - \mathbf{B}_K\mathbf{B}_K^T)\mathbf{X} + N\mathbf{X}\mathbf{X}^T\mathbf{X})_{ij}\mathbf{X}_{ij}^2 = 0 \tag{19}$$

Eq.19 is equivalent to Eq.12, the respective Karush-Kuhn-Tucker condition, since both require either the first term $(-\sum_{K=1}^{N}(\mathbf{A}_K - \mathbf{B}_K\mathbf{B}_K^T)\mathbf{X} + N\mathbf{X}\mathbf{X}^T\mathbf{X})_{ij}$ to be zero or the second terms $\mathbf{X}_{ij}$, $\mathbf{X}_{ij}^2$ to be zero.

Learning Representations

Algorithm 1 $learn\_representation(\mathbf{A}_1, ..., \mathbf{A}_N)$

Input: Adjacency matrices of a N-layers multiplex network $\mathbf{A}_1, ..., \mathbf{A}_N$

Output: A common feature matrix $\mathbf{X}$ and layer-specific feature matrices $\mathbf{B}_1, ..., \mathbf{B}_N$

$\mathbf{X} = 0$, $\mathbf{B}_1, ..., \mathbf{B}_N = 0$

for $t = 1 \to T$ **do**

$$\mathbf{X}_{ij} \leftarrow \mathbf{X}_{ij}\sqrt{\frac{((\sum_{K=1}^{N}(\mathbf{A}_K - \mathbf{B}_K\mathbf{B}_K^T))^+\mathbf{X})_{ij}}{((\sum_{K=1}^{N}(\mathbf{A}_K - \mathbf{B}_K\mathbf{B}_K^T))^-\mathbf{X})_{ij} + (N\mathbf{X}\mathbf{X}^T\mathbf{X})_{ij}}}$$

for $K = 1 \to N$ **do**

$$\mathbf{B}_{Kij} \leftarrow \mathbf{B}_{Kij}\sqrt{\frac{((\mathbf{A}_K - \mathbf{X}\mathbf{X}^T)^+\mathbf{B}_K)_{ij}}{((\mathbf{A}_K - \mathbf{X}\mathbf{X}^T)^-\mathbf{B}_K)_{ij} + (\mathbf{B}_K\mathbf{B}_K^T\mathbf{B}_K)_{ij}}}$$

end for

end for

return $\mathbf{X}$, $\mathbf{B}_1, ..., \mathbf{B}_N$

Clustering Representations

Algorithm 2 $cluster\_representation(\mathbf{X}, \mathbf{B}_K)$

Input: The common feature matrix $\mathbf{X}$ and layer-specific feature matrices $\mathbf{B}_K$ of Kth layer.

Output: The community partition for Kth layer.

1: $\mathtt{C}_K \leftarrow k\text{-}means([\mathbf{X}\ \mathbf{B}_K])$

2: **return** $\mathtt{C}_K$

Settings

- Baselines: Monoplex-CD [13], Monoplex-Newton [14].
- Datasets:
  - Multiplex social networks: AUCS [15], Seventhgrade [16] and Physician [17].
  - Multiplex biological networks: Bos [18, 19] and Celegans [20, 21].
  - Multiplex online networks: ff-tw-yt [22].
- Metric: modularity [23].
- Environment: MATLAB 9.0.0.341360 (R2016a).

Convergence Analysis

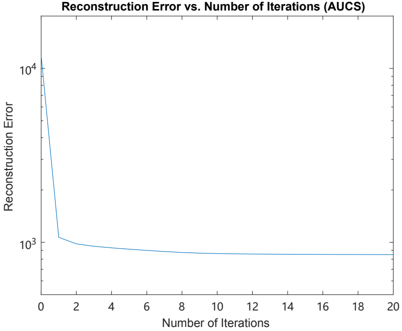

Figure 2: Reconstruction Error vs. Number of Iterations (AUCS).

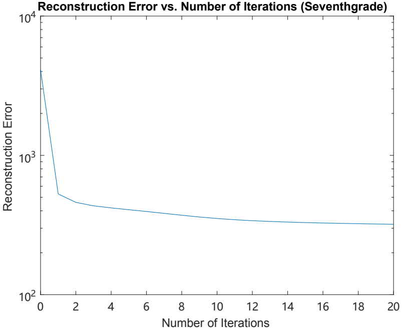

Figure 3: Reconstruction Error vs. Number of Iterations (Seventhgrade).

Convergence Analysis

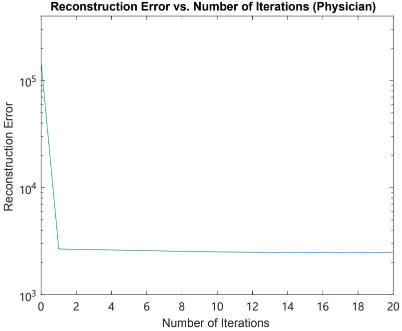

Figure 4: Reconstruction Error vs. Number of Iterations (Physician).

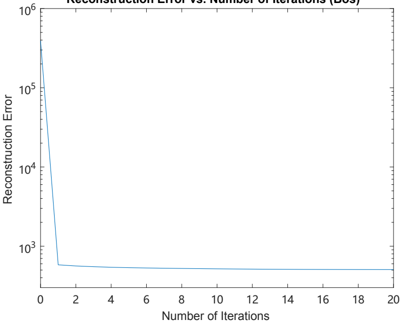

Figure 5: Reconstruction Error vs. Number of Iterations (Bos).

Convergence Analysis

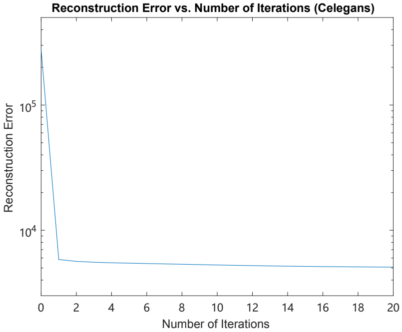

Figure 6: Reconstruction Error vs. Number of Iterations (Celegans).

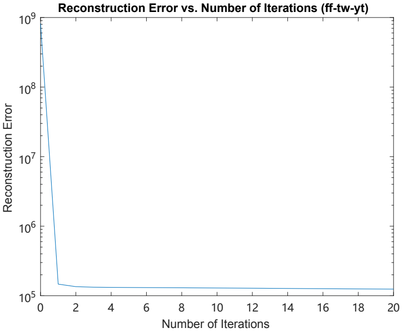

Figure 7: Reconstruction Error vs. Number of Iterations (ff-tw-yt).

Comparative Analysis

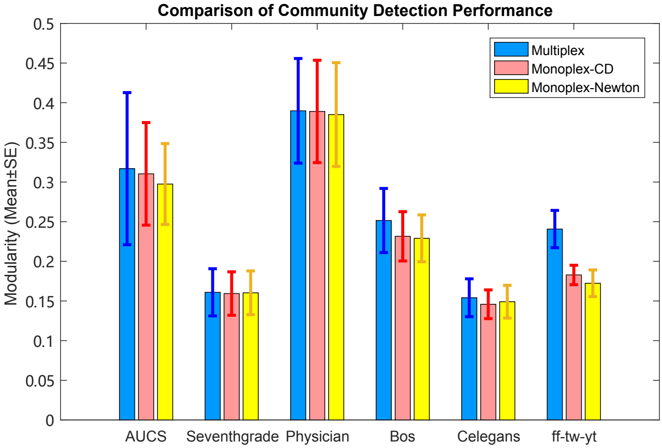

Figure 8: Comparison of community detection performance in terms of modualrity. Error bars depict standard errors.

Noisy Condition Analysis

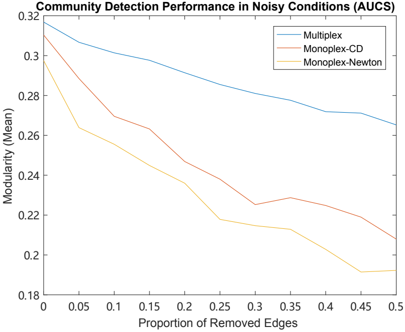

Figure 9: Community detection performance in noisy conditions (AUCS).

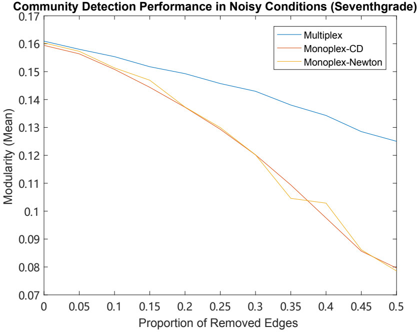

Figure 10: Community detection performance in noisy conditions (Seventhgrade).

Noisy Condition Analysis

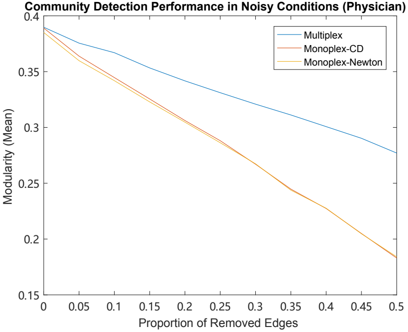

Figure 11: Community detection performance in noisy conditions (Physician).

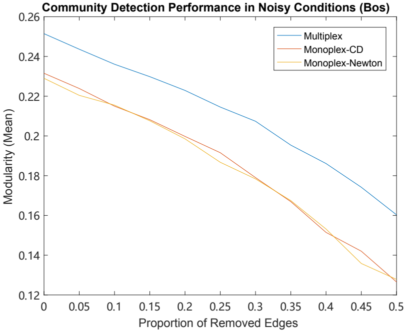

Figure 12: Community detection performance in noisy conditions (Bos).

Conclusion

- Summary:
  - A novel objective for multiplex community detection.
  - A representation-based multiplex community detection framework.
  - Increased performance compared to other monoplex methods in several multiplex network datasets, especially in noisy conditions.
- Future work:
  - Other implementations of the learning and clustering algorithms for more efficiency and accuracy.
  - Handling other types of networks (e.g. directed and weighted network, temporal network, heterogeneous network, etc).
  - Dealing with 'negative transfer' in the community detection context.