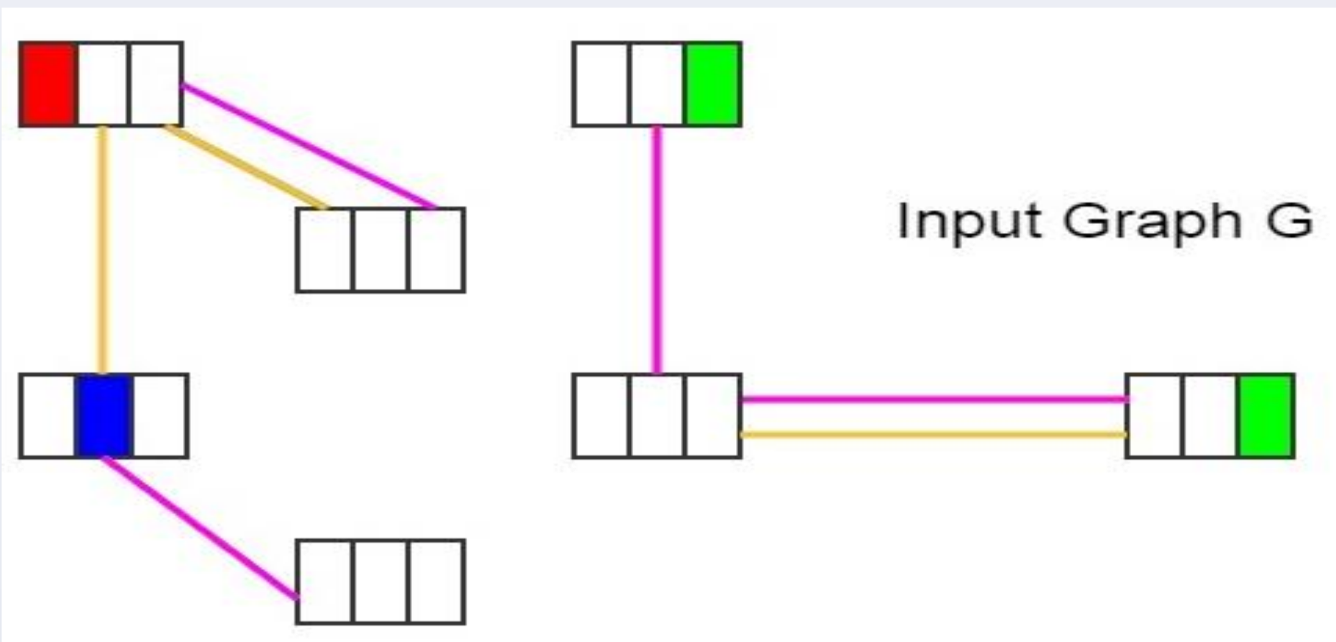
Joint Link Prediction and Multi-label Learning in Heterogeneous Networks

Srichandra Chilappagari and Sumit Negi

Xerox Research Centre India

PROBLEM STATEMENT

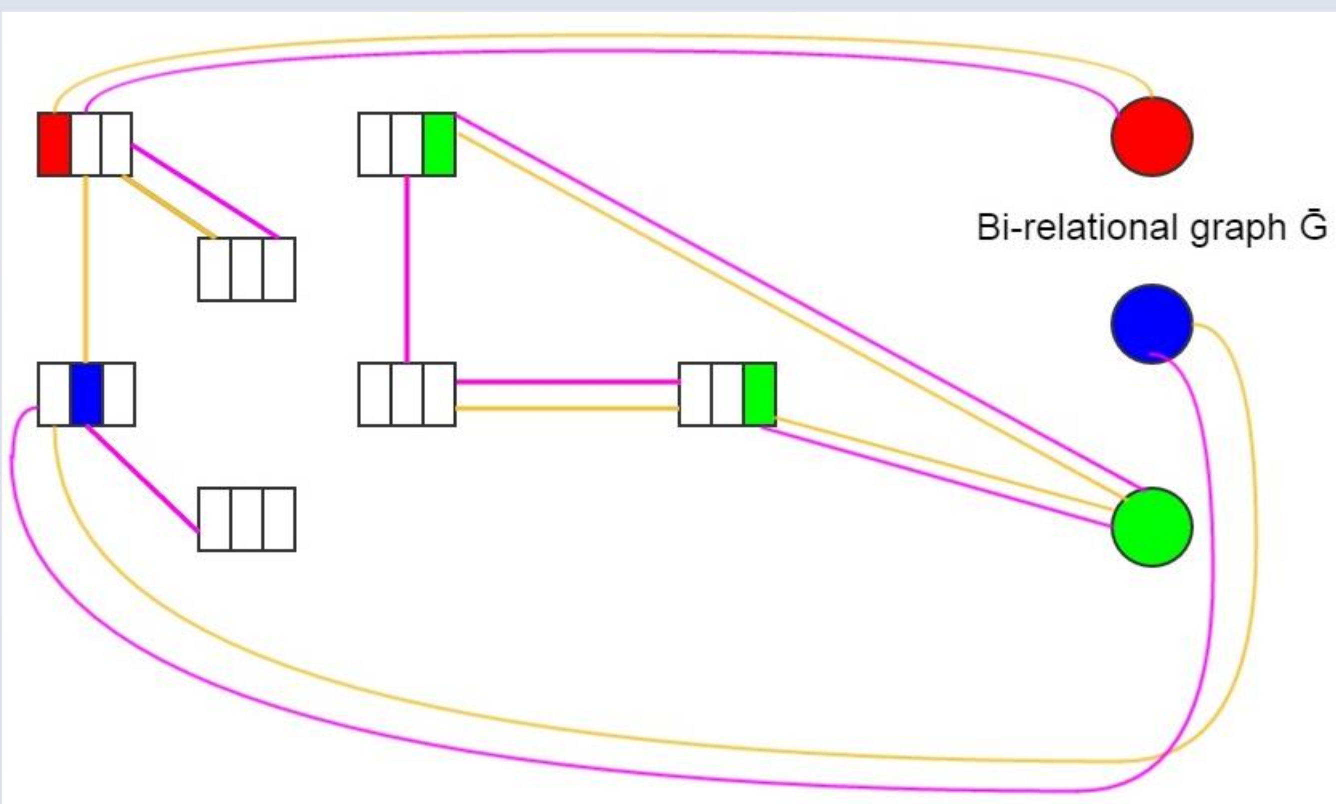
We consider the problems of LP and MLL on a graph $G = (\mathbf{G}^1, \mathbf{G}^2, \ldots, \mathbf{G}^m, \mathbf{Y})$ of n nodes and m link types, where $\mathbf{G}^l \in \{0,1\}^{n \times n}$ denote the partially observed relational graphs between nodes corresponding to link type l. $\mathbf{G}^l_{ij} = 1$ iff we observe a link of type l between nodes i and j and $\mathbf{G}^l_{ij} = 0$ otherwise. $Y \in \{0,1\}^{K \times n}$ represents partially observed labels of nodes, where K is the total number of nodes and $\mathbf{Y}_{kj} = 1$ iff node i is associated with the k^{th} label and $\mathbf{Y}_{kj} = 0$ otherwise.



Without loss of generality, we assume two link types (m = 2) and three labels(K = 3). We assume that labels of first u nodes are given(albeit incomplete) and the other labels are unknown.

APPROACH

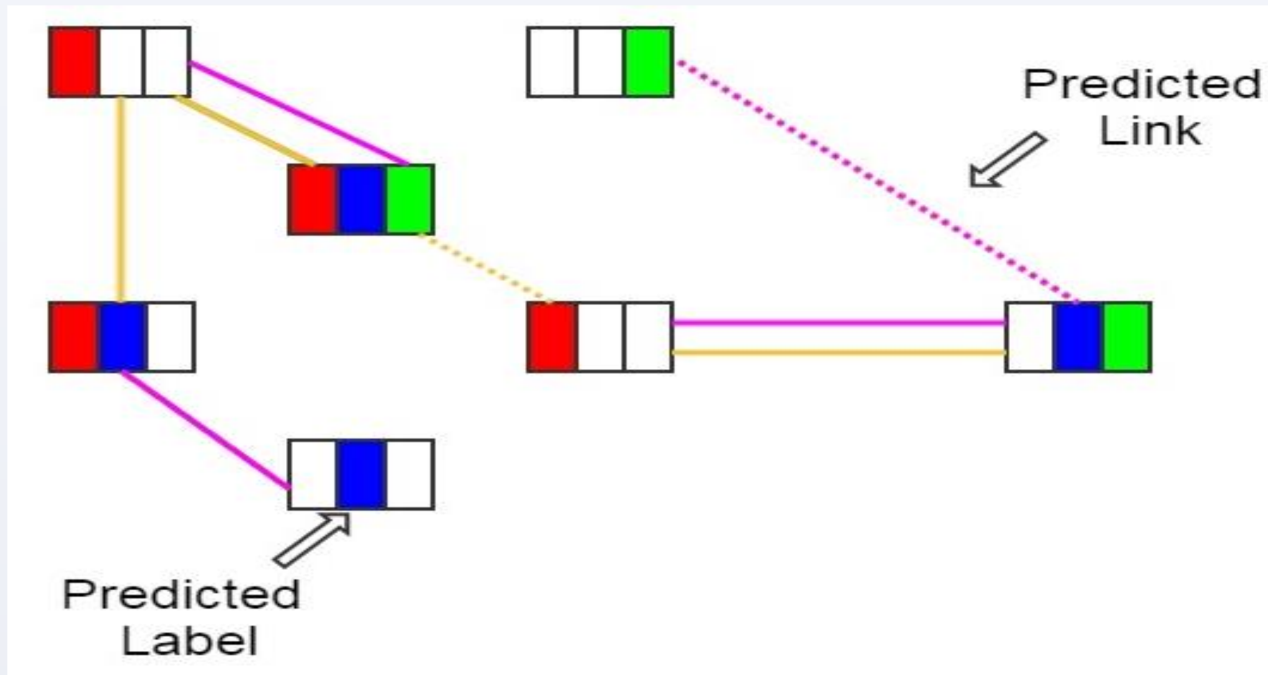
We introduce K new nodes, one for each label and refer to these as *label nodes* and original nodes as *data nodes*. We then create links of all m types between the i^{th} data node and k^{th} label node if node i carries a label k i.e. $\mathbf{Y}_{ki} = 1$. See the figure below for a schematic illustration.



We use $\bar{\mathbf{G}}$ to denote the link matrix of the new graph. $\bar{\mathbf{G}} \in \{0,1\}^{(2n+k) \times (2n+k)}$ can be decomposed as

$$\bar{G} = \begin{bmatrix} G^1 & S & Y^T \\ S^T & G^2 & Y^T \\ Y & Y & H \end{bmatrix}$$

G^1, G^2 are adjacency matrices of original graphs corresponding to link types 1 and 2 respectively. Y is the label matrix and H is the cosine similarity between observed labels [1]. S captures the similarity between G^1 and G^2, which models the engagement between two nodes via different link types and nodes. We claim to perform a better job of LP and MLL by allowing information flow in both directions based on a marginalized denoising framework [2] and combining the problems into a single joint objective function[3].



ALGORITHM

Inputs: $\mathbf{Y}$, Graphs $\mathbf{G}^1$, $\mathbf{G}^2$, Parameters α, $\beta > 0$ and level of corruption: $0 < p < 1$

Output: Link prediction scores $\hat{\mathbf{G}}^1$, $\hat{\mathbf{G}}^2$ and label prediction scores $\hat{\mathbf{Y}}$.

- Set $H_{ij} = \frac{y_i^T y_j}{\|y_i\|\,\|y_j\|}$, where $y_i \in \{0,1\}^{n \times 1}$ is the i^{th} column of matrix
- Set $S_{ij} = \frac{g_i^1 (g_j^2)^T}{\|g_i^1\|\,\|g_j^2\|}$, where $g_i^1 \in \{0,1\}^{1 \times n}$ is i^{th} row of matrix G^1 and $g_j^2 \in \{0,1\}^{1 \times n}$ is i^{th} row of matrix G^2.
- $\bar{G} = [G^1, S, Y^T; S^T, G^2, Y^T; Y^1, Y^2, H]$, $L^1/L^2 \leftarrow$ graph Laplacian of G^1/G^2
- $Q = (1-p)^2 \bar{G}\bar{G}^T + p(1-p)\,\mathrm{diag}(\bar{G}\bar{G}^T)$
- $P = (1-p) * W\bar{G}\bar{G}^T$. Initialize $\hat{Y}^1$, $\hat{Y}^2$ and $W \in \{0,1\}^{(2n+k) \times (2n+k)}$ randomly.
- Let $\bar{G}_1$, $\bar{G}_2$, $\bar{G}_3$ represent the first n, second n and last K columns of matrix $\bar{G}$. Let W_1, W_2, W_3 represent the first n, second n and last K rows of matrix W.
- $Z_1 = (\frac{\alpha}{4}\bar{G}_1 \bar{G}_1^T + \beta Q)^{-1}$
- $Z_2 = (\frac{\alpha}{4}\bar{G}_2 \bar{G}_2^T + \beta Q)^{-1}$
- $Z_3 = (\frac{\alpha}{4}\bar{G}_3 \bar{G}_3^T + \beta Q)^{-1}$
- $Z_4 = (L + \alpha I)-1$
- $Z_{objective} = \mathrm{Tr}(\hat{Y}^1 L^1 \hat{Y}^1) + \mathrm{Tr}(\hat{Y}^2 L^2 \hat{Y}^2) + \alpha \|\hat{Y}^1 - 0.5(W_3\bar{G}_1 + \bar{G}_3^T W_1^T)\|_F^2 + \alpha\|\hat{Y}^2 - 0.5(W_3\bar{G}_2 + \bar{G}_3^T W_2^T)\|_F^2 + \beta \mathrm{Tr}(WQW^T - 2(1-p)W\bar{G}\bar{G}^T)$
- Repeat
- $W_3^1 \leftarrow (\frac{\alpha}{2}(\hat{Y}^1 - \bar{G}_3^T W_1^T)\bar{G}_1^T + \beta P_3)Z_1$
- $W_3^2 \leftarrow (\frac{\alpha}{2}(\hat{Y}^2 - \bar{G}_3^T W_2^T)\bar{G}_2^T + \beta P_3)Z_2$
- $W_3 \leftarrow \frac{1}{2}(W_3^1 + W_3^2)$
- $W_1 \leftarrow (\frac{\alpha}{2}((\hat{Y}^1)^T - \bar{G}_1^T W_3^T)\bar{G}_3^T + \beta P_1)Z_3$
- $W_2 \leftarrow (\frac{\alpha}{2}((\hat{Y}^2)^T - \bar{G}_2^T W_3^T)\bar{G}_3^T + \beta P_2)Z_3$
- $\hat{Y}^1 \leftarrow (\frac{\alpha}{2}(W_3\bar{G}_1 + \bar{G}_3^T W_1^T)Z_4$
- $\hat{Y}^2 \leftarrow (\frac{\alpha}{2}(W_3\bar{G}_2 + \bar{G}_3^T W_2^T)Z_4$
- Until convergence
- $\hat{Y} = \hat{Y}^1 \,\&\, \hat{Y}^2$, $\hat{G}^1 = \frac{1}{2}(W_1\bar{G}_1 + \bar{G}_1^T W_1^T)$, $\hat{G}^2 = \frac{1}{2}(W_2\bar{G}_2 + \bar{G}_2^T W_2^T)$
- Return $\hat{\mathbf{Y}}$, $\hat{\mathbf{G}}^1$, $\hat{\mathbf{G}}^2$

EXPERIMENTS (WIP)

We intend to evaluate our algorithm on the real-world dataset extracted from the DBLP Bibliography data where:

- Nodes represent Authors
- Link type can be co-author relationship or shared community(journal, conference)
- Labels are the Research areas

REFERENCES

- [1] Wang, H.; Huang, H.; and Ding, C. 2011. Image Annotation Using Bi-Relational Graph of Images and Semantic Labels. In *CVPR*, 793-800. IEEE.
- [2] Chen, M.; Xu, Z.; Weinberger, K.; and Sha, F. 2012. Marginalized Denoising Autoencoders for Domain Adaptation. In *ICML*. ACM. 767-774.
- [3] Chen, Z.; Chen, M.; Weinberger, K.; and Zhang, W. 2015. Marginalized Denoising for Link Prediction and Multi-label Learning. In *AAAI*, 1707-1713.