Semi-supervised cluster invariant constraint for network representation learning

Anasua Mitra
Department of CSE
IIT Guwahati

Priyesh Vijayan, Balaraman Ravindran
Department of CSE
IIT Madras

Abstract

We propose a Semi-Supervised Learning (SSL) approach that explicitly encodes different necessary priors to learn efficient representations for nodes in a network. We introduce a novel semi-supervised cluster invariance constraint that explicitly groups nodes of similar labels together. We show that explicitly encoding this largely ignored clustering constraint allows for learning better clusterable representation with qualitative visualizations and quantifiable evaluations. Further, we show that such semi-supervised clusterable representations provide improved performance for semi-supervised node classification across a variety of datasets.

Network Representation Learning

Network representation learning aims to learn a low dimensional representation of graph structure.

Q. How to learn more informative node representation for graphs?

Symbols used

G	(V, E, Y) Networked data	N = L+U L : #Labelled instances U : #Unlabelled instances	C	Community embedding $\in \mathbb{R}^{\mathbf{k}\times\mathbf{m}}$
E	(Un)Weighted connection	$\in \mathbb{R}^{(\mathbf{L+U})\times(\mathbf{L+U})}$	B	Modularity (un)directed $\in \mathbb{R}^{\mathbf{N}\times\mathbf{N}}$
V	Vertices	$\{\mathbf{v}_1, ..., \mathbf{v}_N\}$	α	Proximity factorization $\in \mathbb{R}$
Y	Label matrix	$\in \mathbb{R}^{\mathbf{q}\times\mathbf{N}}$	β	Community factorization $\in \mathbb{R}$
W	Penalty matrix	$\in \mathbb{R}^{\mathbf{q}\times\mathbf{N}}$	θ	Label factorization $\in \mathbb{R}$
S	Proximity matrix (PMI, 1st + 2nd order)	$\in \mathbb{R}^{\mathbf{N}\times\mathbf{N}}$	γ	Modularity maximization $\in \mathbb{R}$
U, U'	Node embedding	$\in \mathbb{R}^{\mathbf{m}\times\mathbf{N}}, \in \mathbb{R}^{\mathbf{N}\times\mathbf{m}}$	ϕ	Label similarity based clustering $\in \mathbb{R}$
M, M'	Context/ neighborhood embedding	$\in \mathbb{R}^{\mathbf{m}\times\mathbf{N}}, \in \mathbb{R}^{\mathbf{N}\times\mathbf{m}}$	ζ	Non-overlapping communities $\in \mathbb{R}$
Q	Label embedding	$\in \mathbb{R}^{\mathbf{q}\times\mathbf{m}}$	λ	L2 regularisation $\in \mathbb{R}$
H, H'	Community membership	$\in \mathbb{R}^{\mathbf{k}\times\mathbf{N}}, \in \mathbb{R}^{\mathbf{N}\times\mathbf{k}}$		N : #Nodes, q : #Labels, k : #Communities m : Dimension of latent space

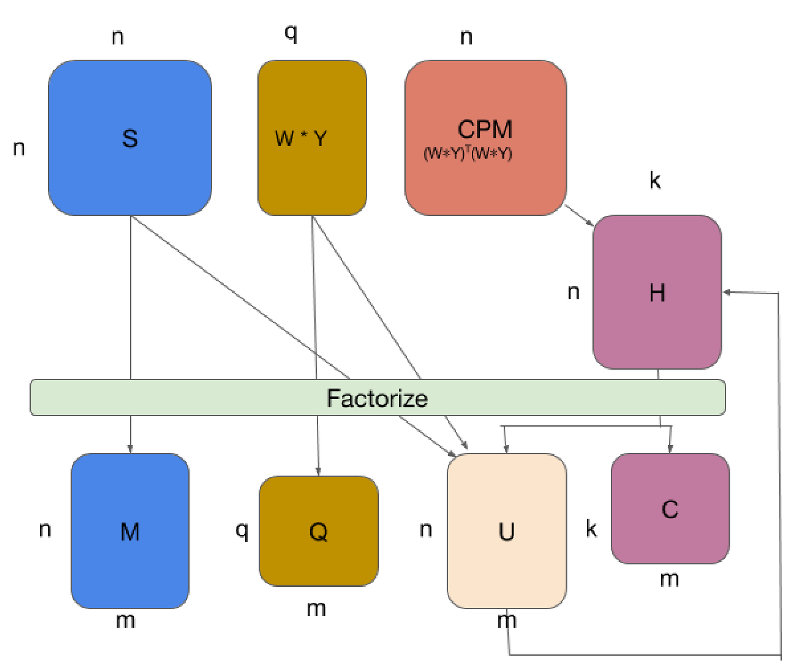
Semi-Supervised Clusterability Constraint

Similar points (nodes belonging to same clusters) tend to have the same labels.

$$\sum_{i,j} \hat{e}_{i,\,j} \| h(u_i) - h(u_j) \|^2 = H\mathcal{L}(\hat{\mathcal{E}})H^T$$

$\hat{E} = (W \odot Y)^T (W \odot Y) \in \mathbb{R}^{N\times N}$ train-label similarity graph/ matrix. $\mathcal{L}(\hat{\mathcal{E}}) = D - \hat{E}$ is the unnormalized Laplacian operator on graph $\hat{E}$. Flow of supervision knowledge from labeled nodes to unlabeled nodes irrespective of their positions in graph !

SS-NMF Framework



Algorithm SS-NMF

Input S, Y, W, H

Output M, U, C, Q, H

Algorithm
1: Initialize M, U, C, Q, H randomly
2: repeat
3: Update M, C, Q, H, U respectively using derived update equations
4: until convergence

Inference
5: compute approximated Y' = Q.U
6: for each unlabeled node i do
7: q' = $\text{argmax}_j(Y'^T_{ij})$

$$L_1 = \min_{M,U} \|S - U^T M\|^2 : M \geq 0, U \geq 0$$

$$L_2 = \min_{Q,U} \|W \odot (Y - QU)\|^2 : Q \geq 0, U \geq 0$$

$$L_3 = \min_{H,C,U} \beta\|H - CU\|^2 + \phi Tr\{H\mathcal{L}(\hat{E})H^T\} + \zeta\|HH^T - I\|^2$$

$$\boxed{L = \alpha L_1 + \theta L_2 + L_3 + \lambda(L2_{reg}) : M, U, Q, C, H \geq 0}$$

Experiment setup

Factorization baselines	Network info	Label info	Well-separated classes	Label smoothing based clusterability	Cluster/ Community info	Misc.
MFDW	$\alpha\|\mathbf{S} - \mathbf{U^T M}\|^2$					
MFDWL	$\alpha\|\mathbf{S} - \mathbf{U^T M}\|^2$	$\theta\|\mathbf{W} \odot (\mathbf{Y} - \mathbf{QU})\|^2$				
MF-Planetoid	$\alpha\|\mathbf{S} - \mathbf{U^T M}\|^2$	$\theta\|\mathbf{W} \odot (\mathbf{Y} - \mathbf{QU})\|^2$		$\phi\mathbf{Tr}\{\mathbf{UL}((\mathbf{W} \odot \mathbf{Y})^\mathbf{T}(\mathbf{W} \odot \mathbf{Y}))\mathbf{U^T}\}$		
MNMF	$\alpha\|\mathbf{S} - \mathbf{M'U'^T}\|^2$		$\zeta\|\mathbf{H'^T H'} - \mathbf{I}\|^2$		$\beta\|\mathbf{H'} - \mathbf{U'C^T}\|^2$	$\gamma\mathbf{Tr}\{\mathbf{H'^T B H'}\}$
MNMFL	$\alpha\|\mathbf{S} - \mathbf{M'U'^T}\|^2$	$\theta\|\mathbf{W} \odot (\mathbf{Y} - \mathbf{QU'^T})\|^2$	$\zeta\|\mathbf{H'^T H'} - \mathbf{I}\|^2$		$\beta\|\mathbf{H'} - \mathbf{U'C^T}\|^2$	$\gamma\mathbf{Tr}\{\mathbf{H'^T B H'}\}$
SS-NMF	$\alpha\|\mathbf{S} - \mathbf{U^T M}\|^2$	$\theta\|\mathbf{W} \odot (\mathbf{Y} - \mathbf{QU})\|^2$	$\zeta\|\mathbf{HH^T} - \mathbf{I}\|^2$	$\phi\mathbf{Tr}\{\mathbf{HL}((\mathbf{W} \odot \mathbf{Y})^\mathbf{T}(\mathbf{W} \odot \mathbf{Y}))\mathbf{H^T}\}$	$\beta\|\mathbf{H} - \mathbf{CU}\|^2$	

Datasets

	$\|V\|$	$\|E\|$	$\|Y\|$	Multi Label	Min Deg.	Max Deg.	Avg Deg.	Singleton Nodes
Washington	230	596	5	False	2	123	4.88	0
Wisconsin	265	724	5	False	2	123	5.0	0
Texas	186	464	5	False	2	105	4.51	0
Cornell	195	478	5	False	2	9	4.12	0
Cora	2,708	5,278	7	False	0	166	2.00	1143
Citeseer	3,312	4,732	6	False	0	99	1.42	1361
Wiki	2,405	17,981	19	False	0	258	6.87	360
PPI	3,890	76,584	50	True	1	594	19.69	0
Blogcatalog	10,312	3,33,983	39	True	1	3992	64.78	0

Experimental Results

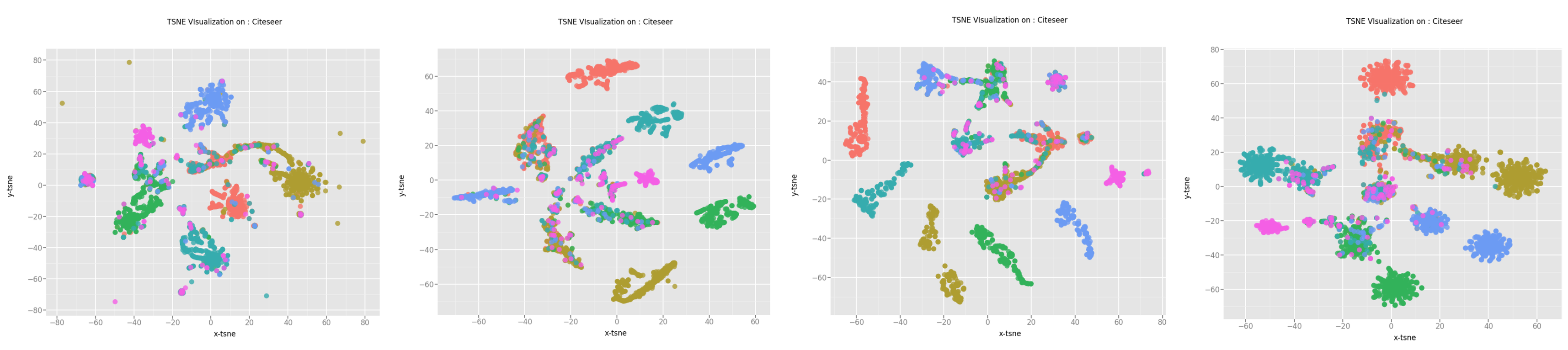
Node Classification Results (10% train data)

N vs LR	Semi-supervised (N/ LR Method)						LR Method		
	Matrix Factorization Approaches							RW Based Approaches	
Train : 10%	Proposed	SoA	Proposed Baseline Variants			SoA		SoA	SoA
Datasets	SS-NMF	MMDW	MFDWL	MF-Planetoid	MNMFL	MNMF	MFDW	DW	N2V
Cora	**77.56**	68.27	70.30	76.47	73.43	74.50	74.44	74.79	74.99
Citeseer	**55.59**	47.38	49.27	55.54	50.09	54.37	53.43	52.72	54.19
Wiki	**57.74**	50.81	56.10	56.19	54.14	56.36	55.50	55.75	56.03
Washington	**56.73**	40.79	50.19	52.40	52.60	57.60	53.70	49.52	50.00
Wisconsin	**53.14**	30.78	45.21	52.72	48.03	48.03	40.71	38.91	43.10
Texas	**55.86**	47.10	53.13	55.38	55.27	55.15	55.03	55.15	55.15
Cornell	**44.63**	38.73	36.89	40.95	40.45	43.84	37.20	22.60	25.42
PPI	14.94	13.85	13.26	14.25	13.96	**17.22**	16.05	16.20	16.88
Blogcatalog	28.17	18.59	27.16	27.31	27.57	28.32	27.88	34.92	**35.16**
Rank	**1.88**	8.44	7.	3.55	5.55	2.88	5.55	5.55	4.55
Score	**1.1262**	10.9094	5.8864	2.5861	4.3281	2.1236	4.5065	5.9937	4.842
LR vs LR	SS-NMF	MMDW	MFDWL	MF-Planetoid	MNMFL	MNMF	MFDW	DW	N2V
Rank	**1.44**	4.88	6.	3.44	3.66	4.77	8.11	7.	5.66
Score	**0.7739**	2.9083	3.4598	1.6377	1.982	2.3559	4.7388	6.226	5.0743

Clustering results (50% train data)

		Semi-supervised (N/ LR Method)					LR Method			
	N vs LR	Matrix Factorization Approaches							RW Based Approaches	
	Train : 50%	Proposed	SoA	Proposed Baseline Variants			SoA		SoA	SoA
		SS-NMF	MMDW	MFDWL	MF-Planetoid	MNMFL	MNMF	MFDW	DW	N2V
Purity	Rank	**1.1429**	6.2857	5.2857	2.	2.8571	4.8571	8.5714	7.5714	6.4286
	Score	**0.2728**	11.9328	9.629	2.9642	7.0497	9.9101	16.6532	14.1462	12.8244
NMI/ ONMI	Rank	**1.2222**	6.3333	4.7778	2.6667	3.1111	4.8889	8.	7.8889	6.1111
	Score	**0.0258**	17.0103	12.4661	4.1336	7.0906	13.3816	20.077	20.2128	19.2078
Omega Index	PPI	6.49	4.43	4.12	6.14	5.81	5.20	3.37	6.25	**6.90**
	Blogcatalog	**4.64**	3.67	3.99	4.30	4.07	3.82	2.71	2.06	3.19

Visualization of the node embeddings in the projected space



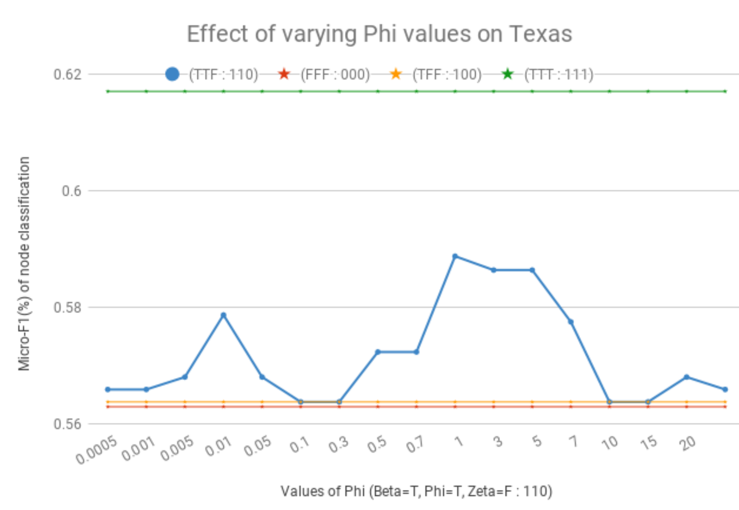
a. MFDWL b. MNMFL c. MF-Planetoid d. SS-NMF

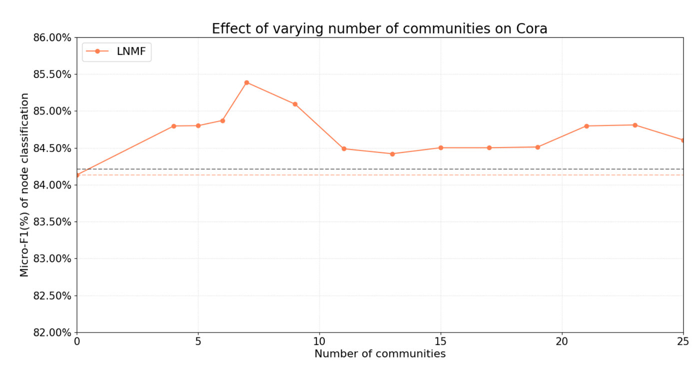
Prior Analysis & Parameter Sensitivity Analysis

How each cluster related prior information, a) Cluster matrix factorization term: $\beta\|H - CU\|^2$, b) Regularization based cluster learning from Cluster Proximity Matrix(CPM): $\phi Tr\{H\mathcal{L}(\hat{E})H^T\}$, c) Minimum overlapping cluster constraint: $\zeta\|HH^T - I\|^2$: contributes in learning cluster membership matrix H.

Table: Prior experiment settings

β	ϕ	ζ	Meaning
0	0	0	(FFF) : no cluster H learning
1	0	0	(TFF) : cluster H learns from U
1	1	0	(TTF) : cluster H directly learns from CPM and U
1	0	1	(TFT) : cluster H learns from U and orthogonality constraint
1	1	1	(TTT) : original setup





References

- Tu, Cunchao, et al. "Max-Margin DeepWalk: Discriminative Learning of Network Representation." IJCAI. 2016.
- Yang, Zhilin, William W. Cohen, and Ruslan Salakhutdinov. "Revisiting Semi-Supervised Learning with Graph Embeddings" Journal of Machine Learning Research 9 (2016): 175-180.
- Wang, Xiao, et al. "Community Preserving Network Embedding." AAAI. 2017.