IWML 2018

Semi-supervised cluster invariant constraint for network representation learning

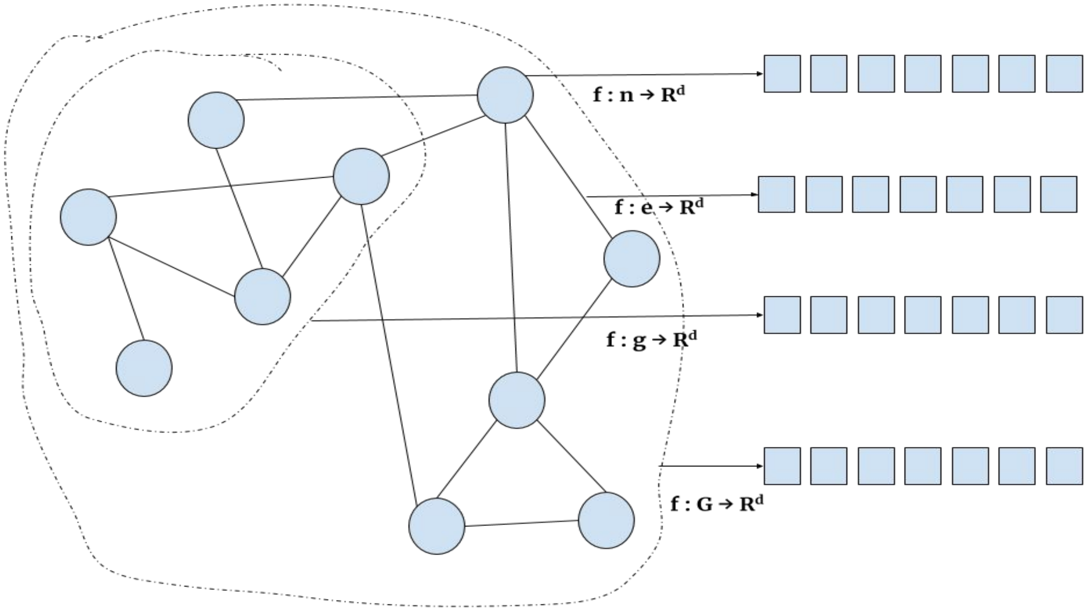
Anasua Mitra

Collaborators: Priyesh Vijayan & Dr. Balaraman Ravindran
(Indian Institute of Technology Madras)

Indian Institute of Technology Guwahati
July 2018

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?

$f : n \rightarrow U_v \in R^m$

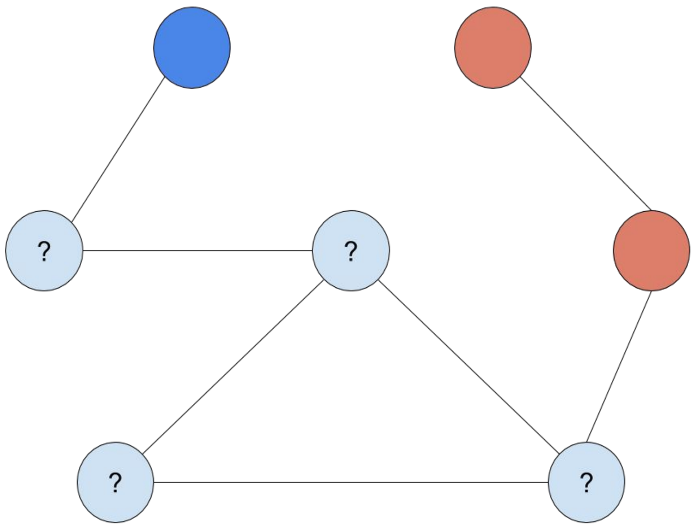
Paradigms for NRL :

- Factorization
- Random walk
- Deep learning
- Graph kernels
- Generative models
- Hybrid models

Experiment Setting

Relational Classification setup

- $G = (V, E, Y)$
- $N = L + UL$
- $V = \{v_1, v_2, \ldots, v_N\}$
- $E \in R^{NxN}$
- $Y \in R^{qxN}$
- $C = \{c_1, c_2, \ldots, c_q\}$
- Relational Learning is challenging when given networked data has link sparsity and/ or label sparsity.



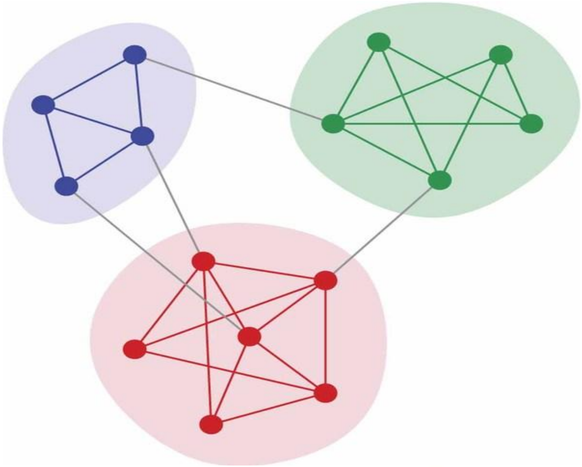
Assumption: Network exhibits Homophily.

Applicable to non-attributed, (un)-directed, (un)-weighted graphs.

Community vs Clusterability

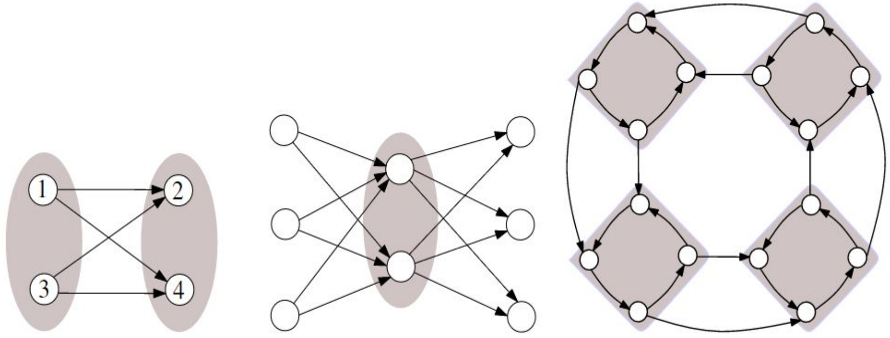
Sources of information for non-attributed graphs :

- Local structure
- Global structures, i.e., communities, clusters
- Labels



Community detection is a form of clustering to discover modular structures in graph data.

Clustering is a more general concept which groups entities together based on certain criteria. No notion of prior connectedness among entities is assumed.



(a) Citation-based cluster (b) Citation-based cluster (c) Flow-based cluster

Some useful local invariance assumptions

Embedding invariance: $\sum_{i,j} e_{ij} \| u_i - u_j \|^2 = U^T \Delta U$

Label invariance: $\sum_{i,j} e_{ij} \| f(u_i) - f(u_j) \|^2 = f^T \Delta f$

$\Delta = D - E$, Graph Laplacian

Smoothness assumption in local neighborhood.

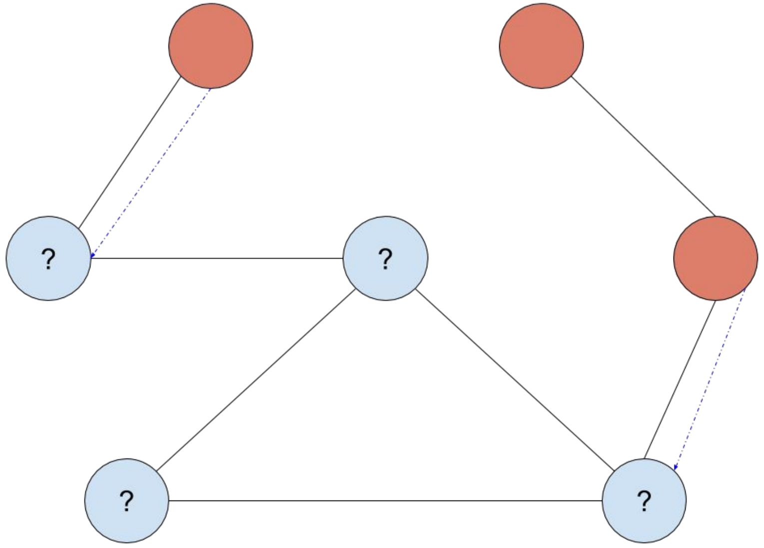
Semi-supervised learning leverages both labeled and unlabeled data for prediction.

$A \in R^{N\mathrm{x}N}$, $N = L + UL$.

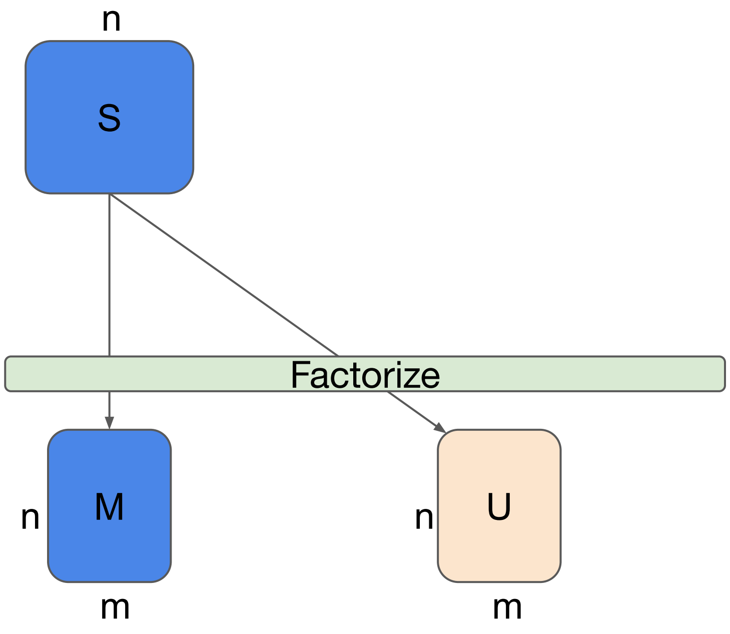
Graph based SSL assumes Label Invariance.

$$\sum_{i = 1 \text{ to } L} \text{loss}\,(y_i, f(u_i)) + \lambda \left(\sum_{i,j} e_{ij} \| f(u_i) - f(u_j) \|^2 \right)$$

$$= \sum_{i = 1 \text{ to } L} \text{loss}\,(y_i, f(u_i)) + \lambda\,(f^T \Delta f)$$



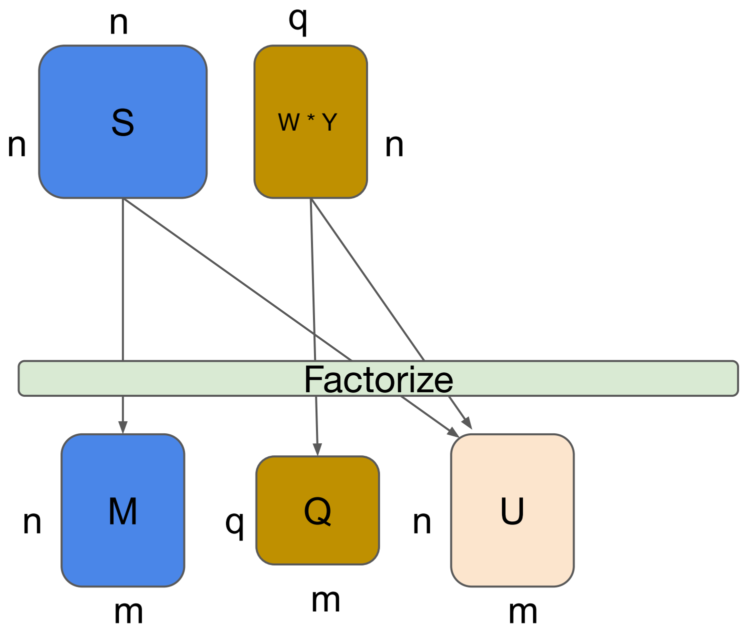
Important factorization based baselines



Deep-walk as matrix factorization (MFDW)

- $\min_{M,U} \alpha(S - U^T.M)^2 + \lambda(U^2 + M^2)$
- Prior information in use : Local neighborhood.

$$S_{i,j} = \log \frac{e_i(A + A^2 + A^3 + ... + A^t)_j}{t}$$



MFDW + Label Matrix Factorization/ Max-Margin Loss

- $\min_{M,U} \alpha(S - U^T.M)^2 + \theta(W * |Y - Q.U|)^2 + \lambda(U^2 + M^2 + Q^2)$
- Prior information in use : Local neighborhood, label info.

Our assumption of invariance for node clusterability

<u>Label guided cluster invariance to capture global structure :</u>

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

$$\sum_{i,j} e'_{ij} \, || \, h(u_i)-h(u_j) \, || \,^2 = H\Delta H^T$$

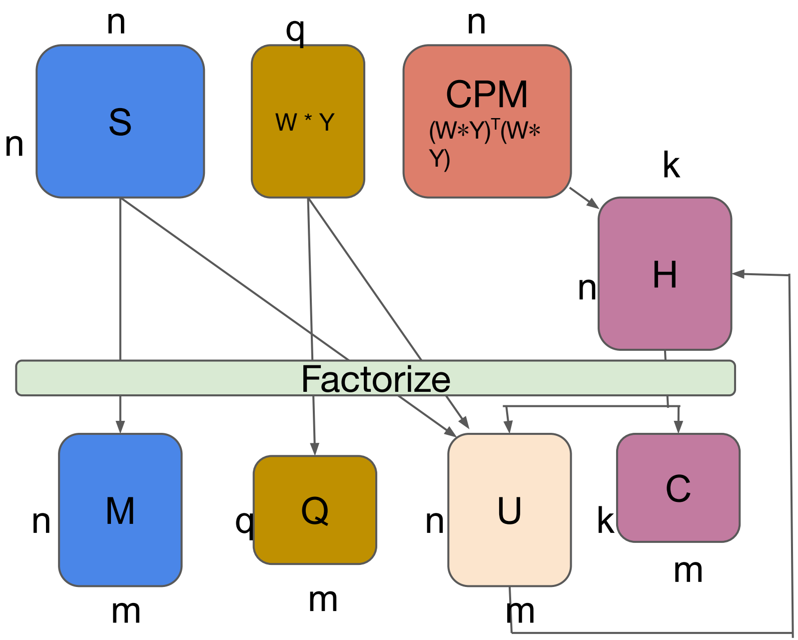
$\Delta = D - E'$, Graph Laplacian

$E' = (W \ast Y)^T(W \ast Y)$, train-label similarity graph/ matrix

Flow of supervision knowledge from labeled nodes to unlabeled nodes irrespective of their positions in graph !

Incorporating semi-supervised clusterability constraint in NRL setup.

SS-NMF Framework



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

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

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

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

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’ = $\text{argmax}_j(Y'^T_{ij})$

Components of competing algorithms

Factorization baselines	Network info	Label info	Well-separated clusters	Label smoothness	Semi-supervised/ clustering	Misc.
MFDW	$\alpha \|\|S - U^T M\|\|^2$					
MMDW	$\alpha \|\|S - U^T M\|\|^2$	Max-margin loss				
MFDWL	$\alpha \|\|S - U^T M\|\|^2$	$\theta \|\|W \odot (Y - QU)\|\|^2$				
MF-Planetoid	$\alpha \|\|S - U^T M\|\|^2$	$\theta \|\|W \odot (Y - QU)\|\|^2$		$\phi Tr\{U \mathcal{L}(E) U^T\}$		
MNMF	$\alpha \|\|S - M' U'^T\|\|^2$		$\zeta \|\|H'^T H' - I\|\|^2$		$\beta \|\|H' - U' C^T\|\|^2$	$\gamma Tr\{H'^T B H'\}$
MNMFL	$\alpha \|\|S - M' U'^T\|\|^2$	$\theta \|\|W \odot (Y - QU'^T)\|\|^2$	$\zeta \|\|H'^T H' - I\|\|^2$		$\beta \|\|H' - U' C^T\|\|^2$	$\gamma Tr\{H'^T B H'\}$
SS-NMF	$\alpha \|\|S - U^T M\|\|^2$	$\theta \|\|W \odot (Y - QU)\|\|^2$	$\zeta \|\|H H^T - I\|\|^2$	$\phi Tr\{H \mathcal{L}(E) H^T\}$	$\beta \|\|H - CU\|\|^2$	

<u>*References:*</u>

Perozzi et al., "Deepwalk: Online learning of social representations." *KDD 2014.*
Tu, C., Zhang, W., Liu, Z., Sun, M.: Max-margin deepwalk: Discriminative learning of network representation. In: IJCAI, pp. 3889–3895 (2016).
Yang et al., "Revisiting Semi-Supervised Learning with Graph Embeddings." ICML 2016.
Wang, Xiao, et al. "Community Preserving Network Embedding." *AAAI*. 2017.

Experiment Results > Node Classification

	Semi-supervised (N/ LR Method)					LR Method			
N vs LR	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

Experiment Results > Node Clustering

		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

Experiment Results > Ablation Study

Q. How each component of our proposed equation influences node representations?

- We have incrementally built our model to understand this.

MFDW : $\min_{M,U} \alpha(S - U^T.M)^2$

MFDWL : $\min_{M,U,Q} \alpha(S - U^T.M)^2 + \theta(W * |Y - Q.U|)^2$

MF-Planetoid : $\min_{M,U,Q} \alpha(S - U^T.M)^2 + \theta(W * |Y - Q.U|)^2 + \phi Tr\{ U.L(E').U^T\}$

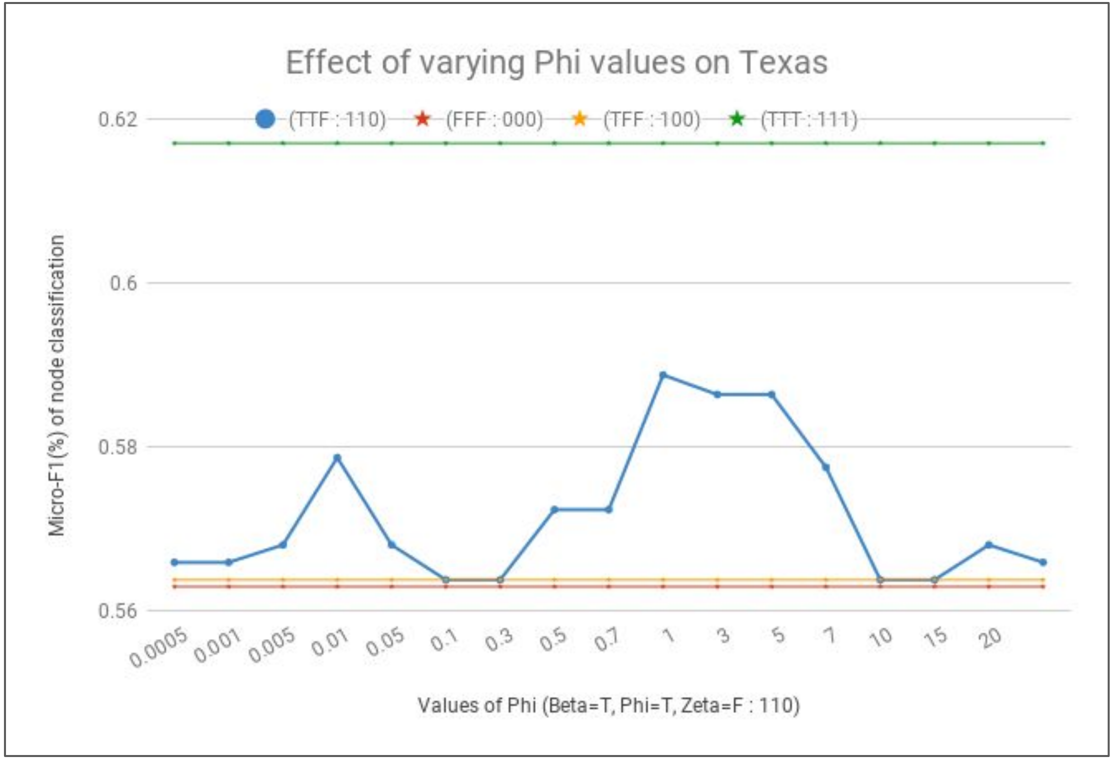
SS-NMF : $\min_{M,U,C,Q,H} \alpha(S - U^T.M)^2 + \theta(W * |Y - Q.U|)^2 + \phi Tr\{ H.L(E').H^T\} + \beta || H - C.U||^2 + \xi|| HH^T - I ||^2$

Experiment Results > Prior Analysis

Q. How each cluster related term contributes in learning cluster membership matrix ?

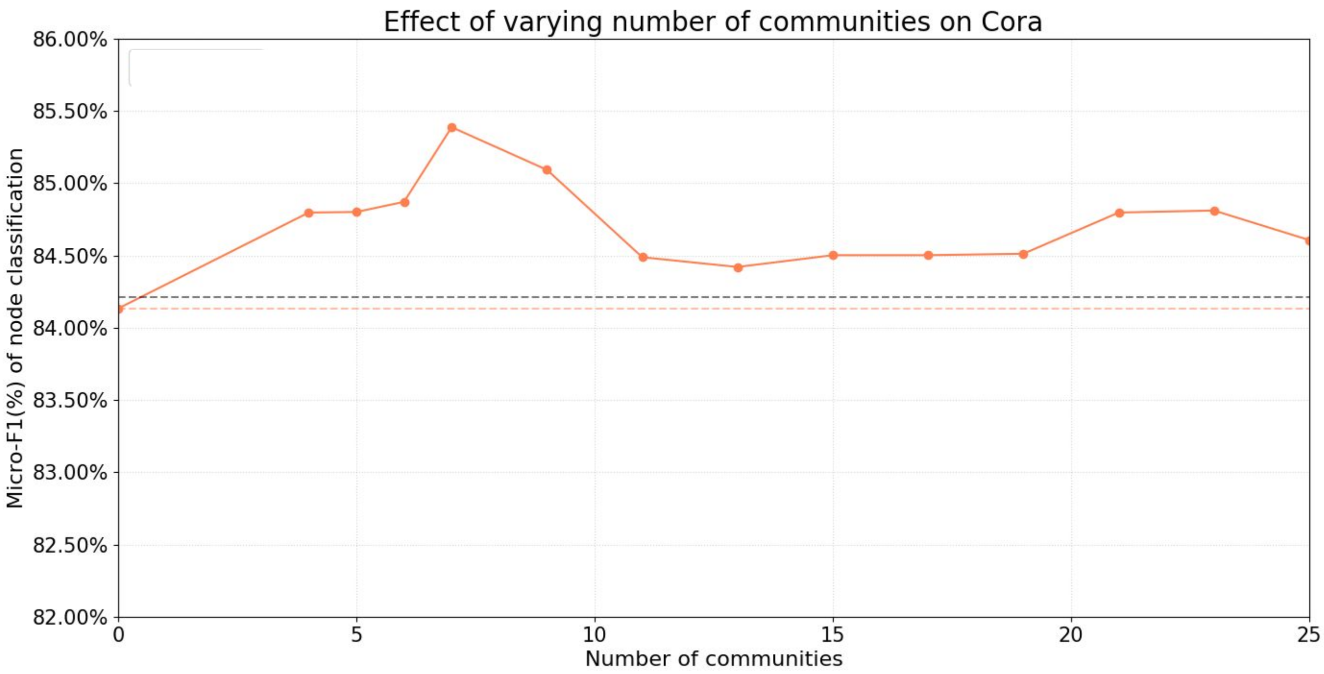
$$\min_{M,U,C,Q,H} \alpha(S - U^T.M)^2 + \theta(W * |Y - Q.U|)^2 + \beta \,||\, H - C.U \,||^2 + \phi Tr\{ H.L(E').H^T\} + \xi ||\, HH^T - I \,||^2 + \lambda(L2_{reg})$$

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



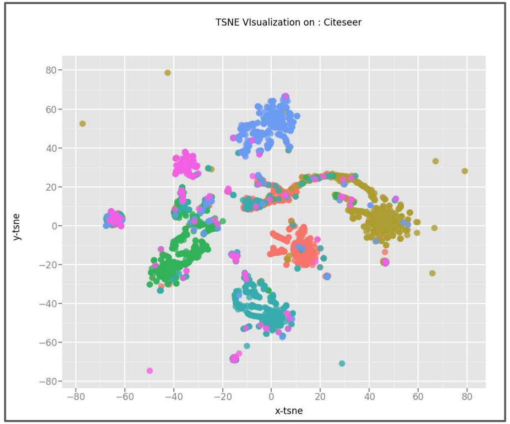
Experiment Results > Parameter Sensitivity

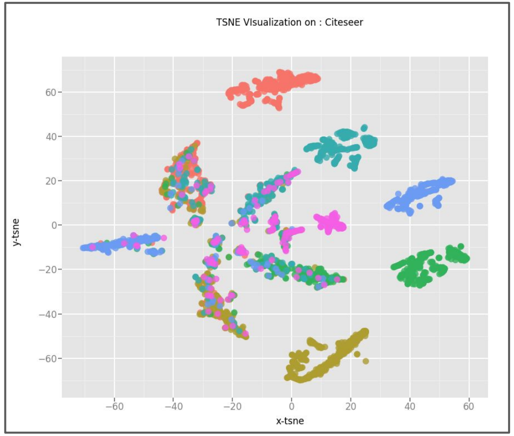
Varying number of clusters



Experiment Results > t-SNE Plots

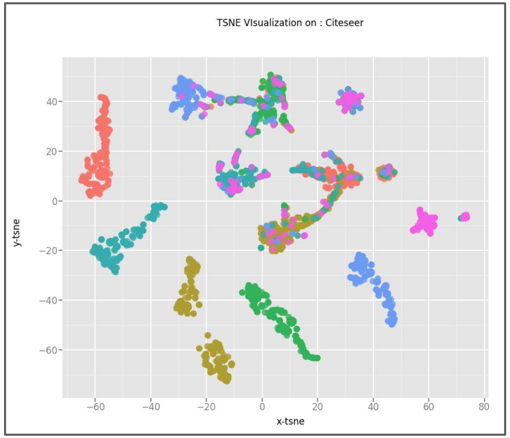
MFDWL

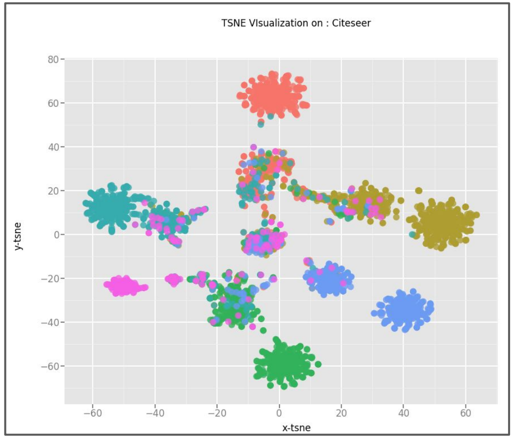




MNMFL

MF-P





SS-NMF

Thank you !

References

Perozzi, B., Al-Rfou, R., Skiena, S.: Deepwalk: Online learning of social representations. In: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 701–710. ACM(2014).
URL: https://dl.acm.org/citation.cfm?id=2623732 .

Grover, A., Leskovec, J.: node2vec: Scalable feature learning for networks. In: Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 855–864. ACM (2016).
URL: https://dl.acm.org/citation.cfm?id=2939754 .

Tu, C., Zhang, W., Liu, Z., Sun, M.: Max-margin deepwalk: Discriminative learning of network representation. In: IJCAI, pp. 3889–3895 (2016).
URL: http://www.thunlp.org/~tcc/publications/ijcai2016_mmdw.pdf .

Yang, Z., Cohen, W.W., Salakhutdinov, R.: Revisiting semi-supervised learning with graph embeddings. arXiv preprint arXiv:1603.08861 (2016).
URL: http://proceedings.mlr.press/v48/yanga16.pdf .

Wang, X., Cui, P., Wang, J., Pei, J., Zhu, W., Yang, S.: Community preserving network embedding. In: AAAI, pp. 203–209 (2017).
URL: http://www.aaai.org/ocs/index.php/AAAI/AAAI17/paper/download/14589/13763 .