Motif Aware Graph Embedding

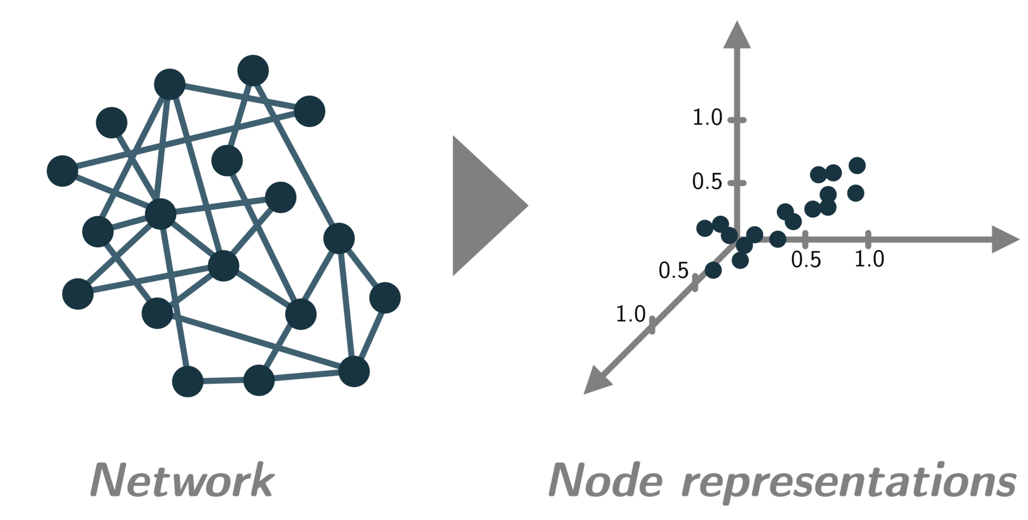
Hoang Nguyen, Tsuyoshi Murata

Department of Computer Science, Tokyo Institute of Technology

TOKYO TECH

Objective

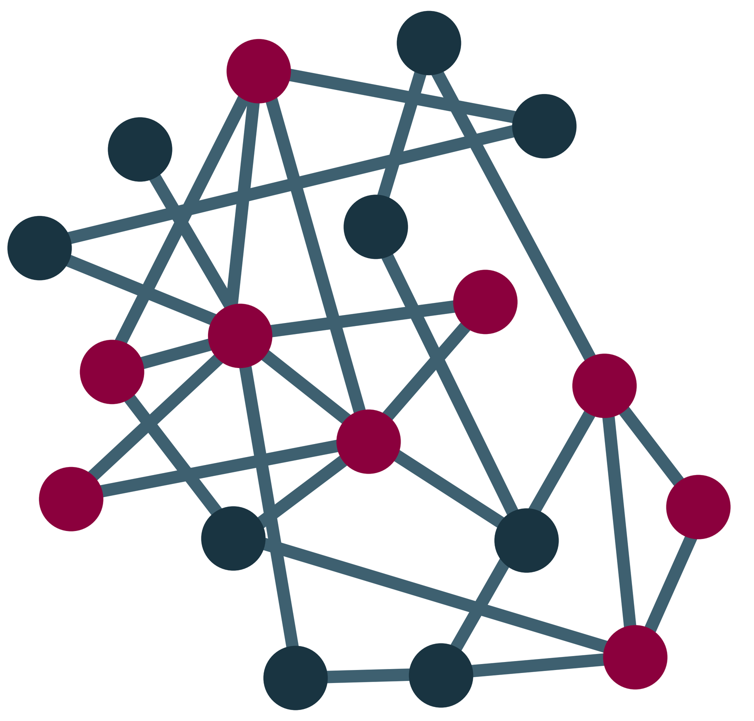
To learn a higher-order latent representation of a complex network:



The learned representation will be used in various machine learning tasks.

Key Ideas

Inject a targeted motif structure into the representation learning process.



Motif structure injection can be realized by:

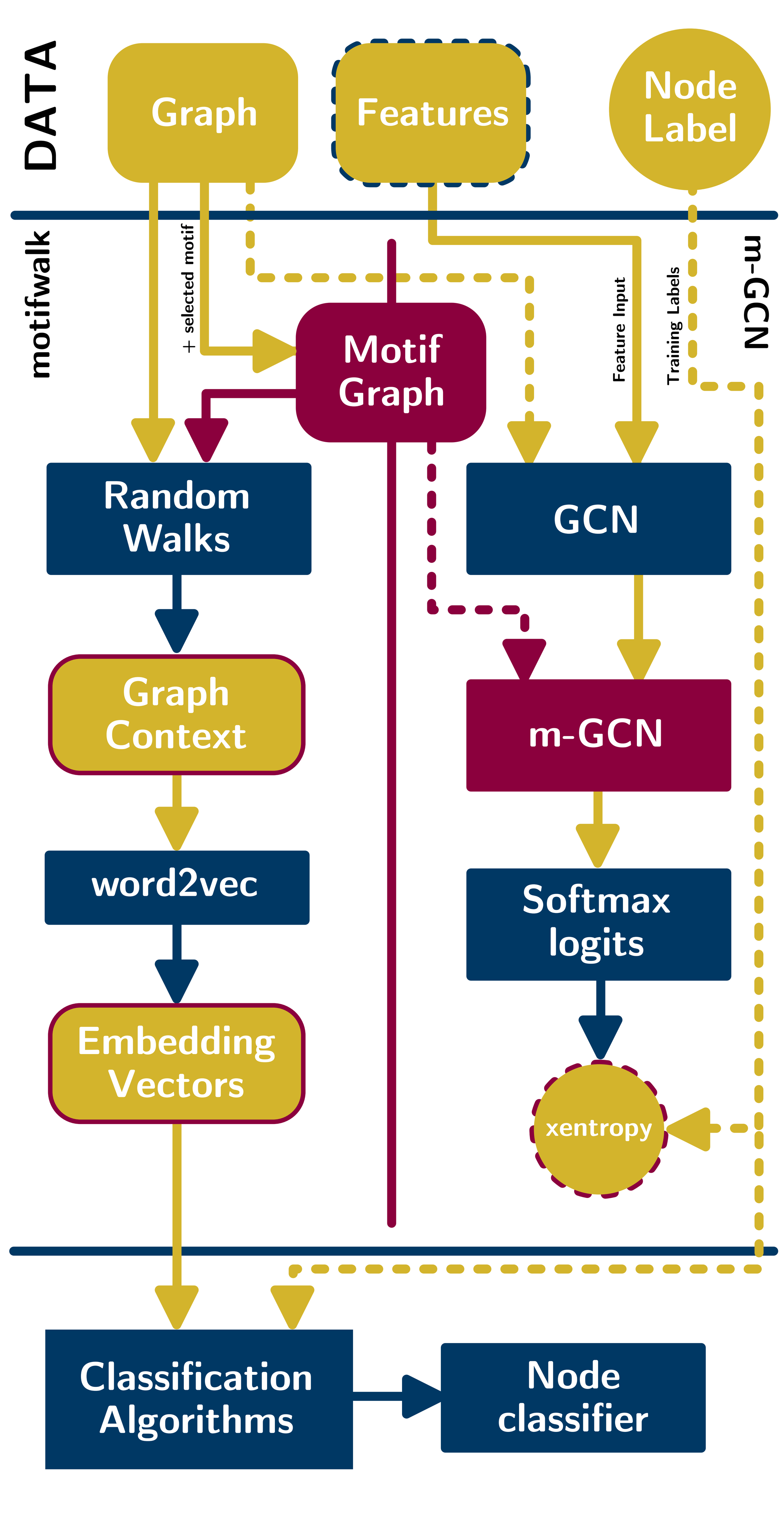
- Biased random walk.
- Wavelet basis defined by a motif matrix.

Which motif?

Measure the motif's significant by z-score:

$$\text{z-score} = \frac{N_m(G) - N_m(G_{\text{random}})}{\sigma_m(G_{\text{random}})}$$

Model Overview



Datasets

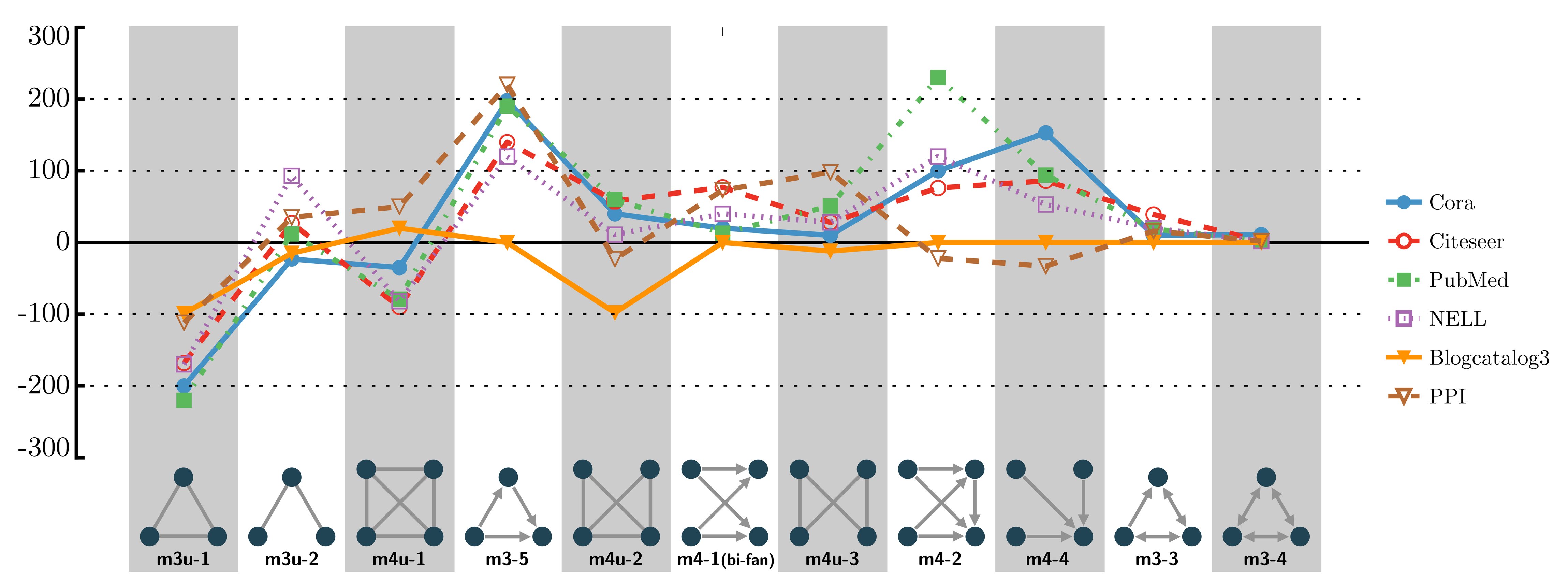
Dataset	#Classes	#Nodes	#Edges	Training ratio
BlogCatalog	39	10,312	333,983	0.5
Citeseer	6	3,327	4,732	0.5

Table 1: Datasets for unsupervised embeddings

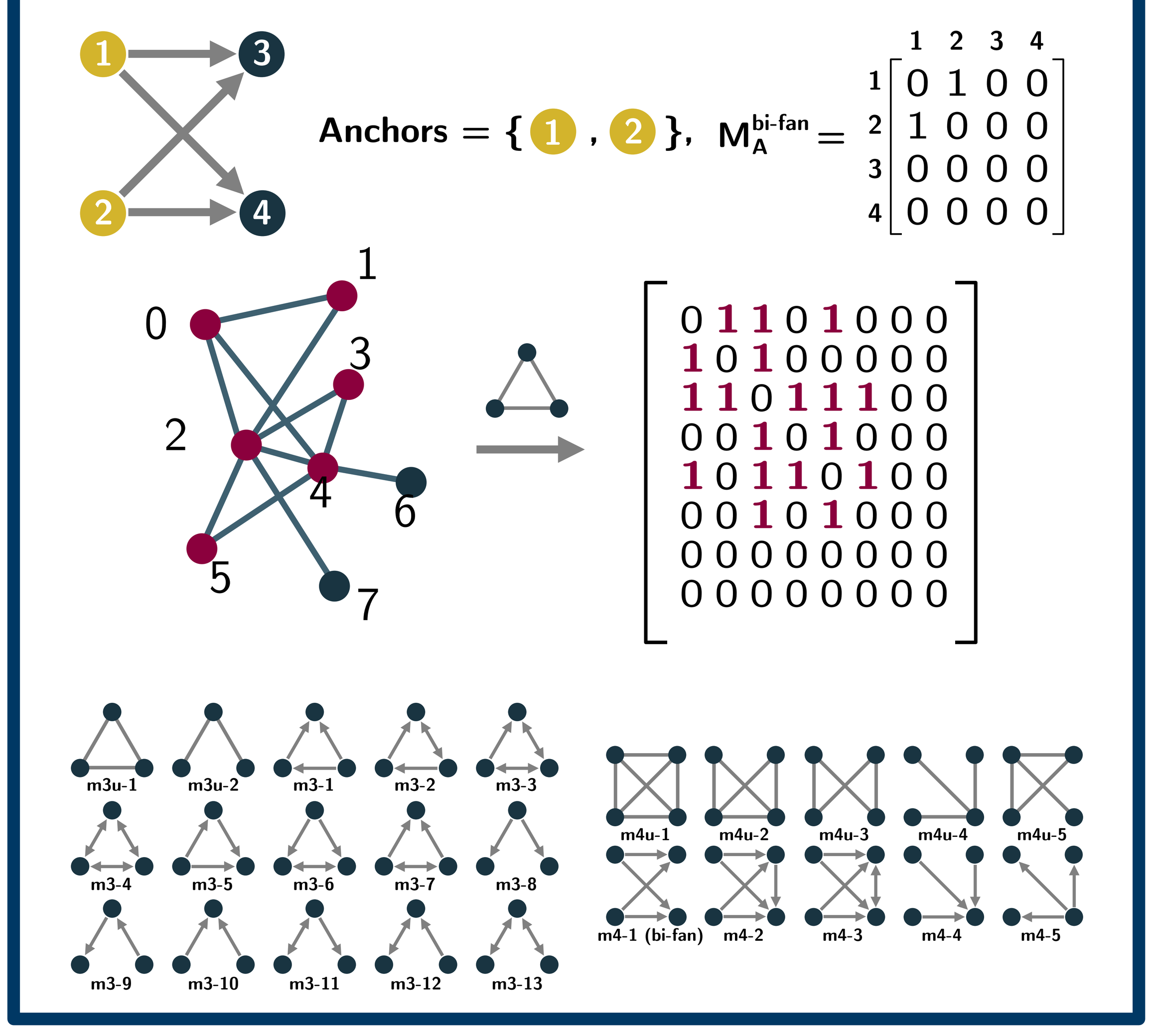
Dataset	#Classes	#Nodes	#Edges	#Features
Citeseer	39	10,312	333,983	3,703
Cora	6	2,708	4,732	1,433
Pubmed	3	19,717	44,338	76,584
NELL	210	65,755	266,144	5,414

Table 2: Datasets for semi-supervised embeddings

Motif z-score (configuration model)



Motif Adj. Matrix



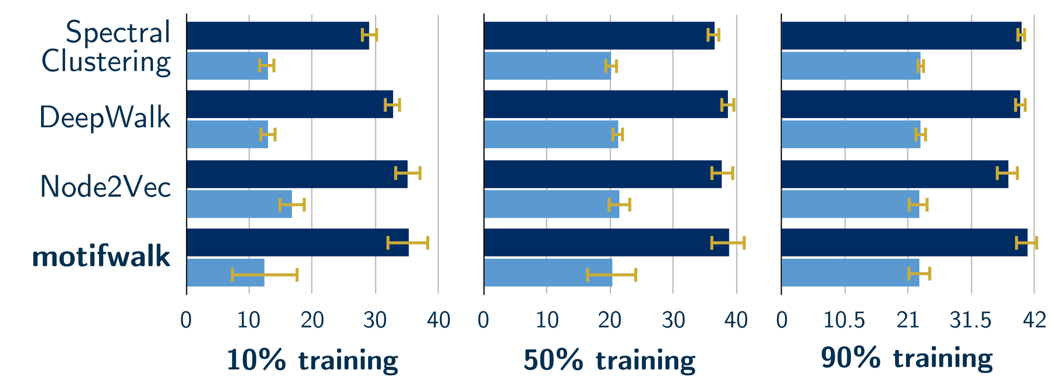
GCN & m-GCN

$$g_\theta \star x = U g_\theta U^\top x$$

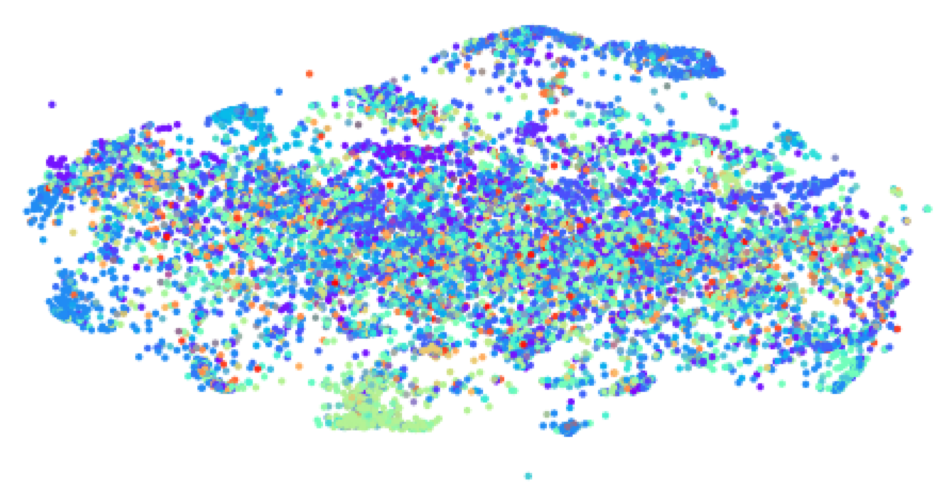
$$Z = \widetilde{D}^{-\frac{1}{2}} \widetilde{A} \widetilde{D}^{-\frac{1}{2}} X W$$

Results

- Motifwalk + OVR - Linear Regression.



BlogCatalog 3



- m-GCN

Method	Citeseer	Cora	Pubmed	NELL
Deepwalk	43.2	67.2	65.3	58.1
motifwalk	45.7	68.0	64.9	58.8
Planetoid	64.7	75.7	77.2	61.9
GCN	70.3	81.5	79.0	66.0
m-GCN	**71.2**	**82.1**	**79.5**	**66.1**
m-GCN (rand. splits)	70.2 ± 0.5	81.1 ± 0.5	79.3 ± 0.7	62.0 ± 1.4

References

(Perezzi, 2014) Deepwalk: Online learning of social representations.

(Benson, 2016) Higher-order organization of complex networks

(Kipf, 2017) Semi-supervised classification with graph convolutional networks.