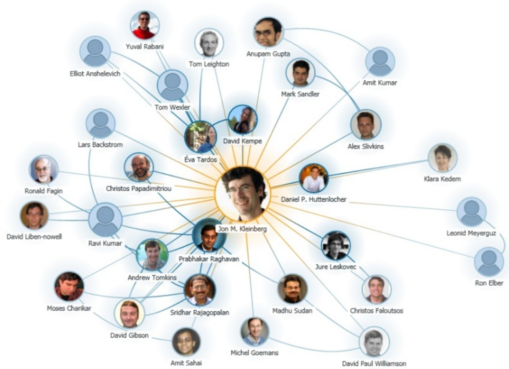
Co-Regularized Deep Multi-Network Embedding

Jingchao Ni[1], Shiyu Chang[2], Xiao Liu[3], Wei Cheng[4], Haifeng Chen[4],
Dongkuan Xu[1], and Xiang Zhang[1]

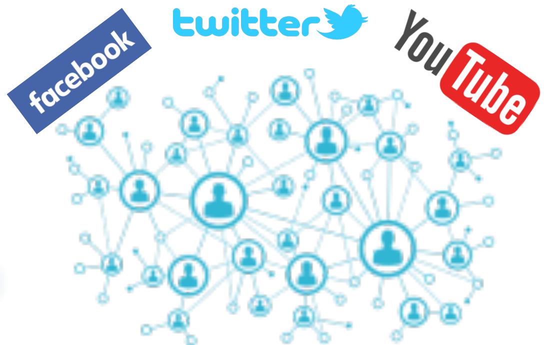
[1]College of Information Sciences and Technology, Pennsylvania State University
[2]IBM Thomas J. Watson Research Center
[3]Department of Biomedical Engineering, Pennsylvania State University
[4]NEC Laboratories America

The Web Conference 2018

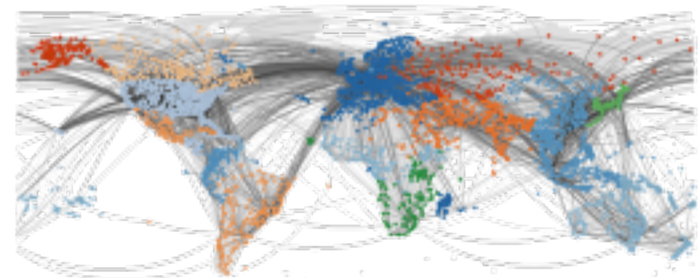
Information Networks Are Prevalent



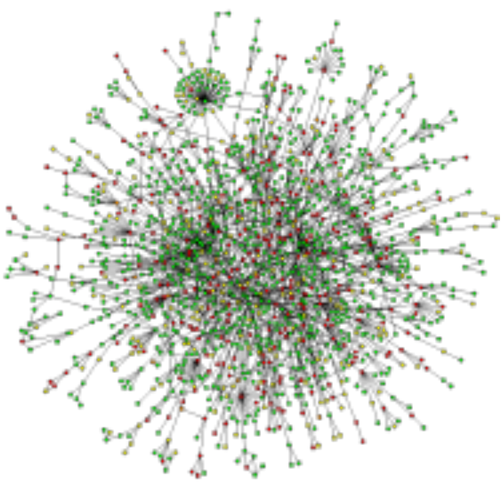
Collaboration network



Social network

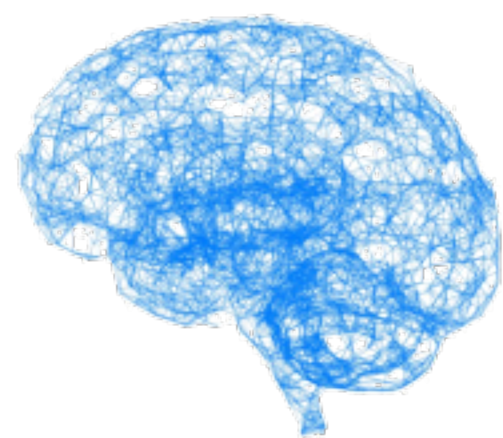


Traffic network



Protein-protein-interaction network

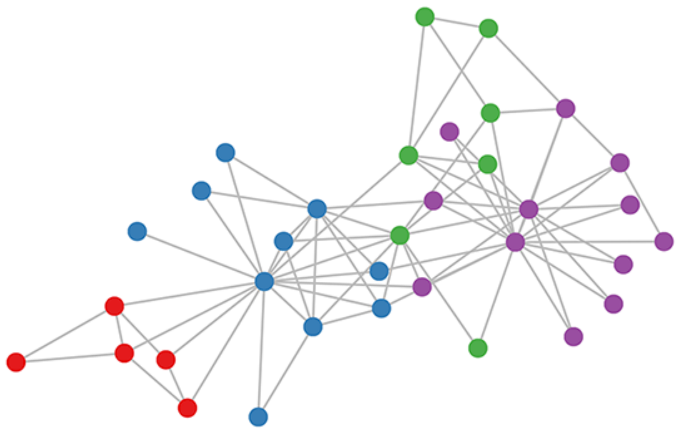
Disease similarity network



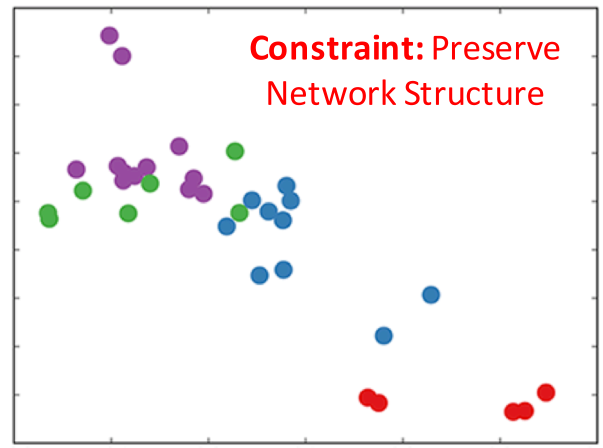
Brain network

Network Embedding

Low-dimensional space



Embedding



G = (V, E)

G = $[\vec{v}_1, \ldots, \vec{v}_n]$

Existing Methods

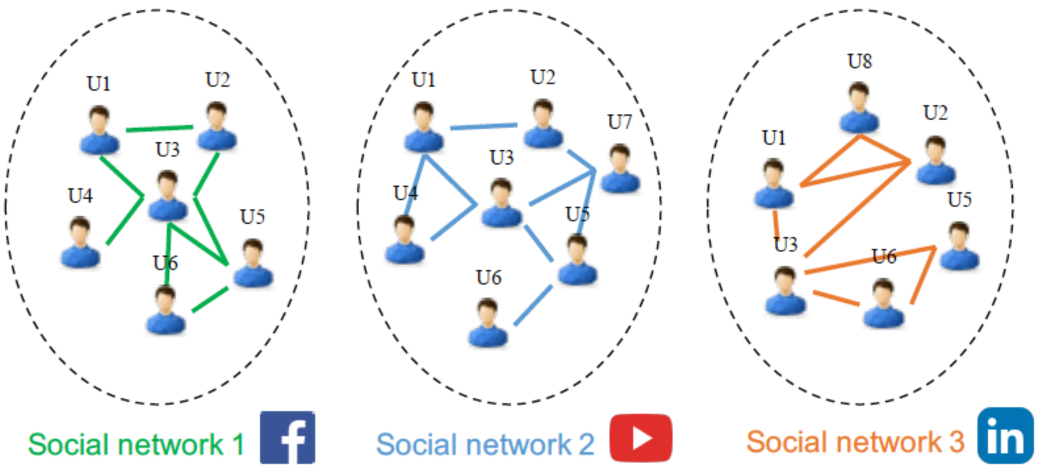
- DeepWalk [Perozzi et al., KDD'14]
- LINE [Tang, et al., WWW'15]
- GraRep [Cao et al., CIKM'15]
- Node2vec [Grover and Leskovec, KDD'16]
- DNGR [Cao et al., AAAI'16]
- …

- Node Classification
- Node Clustering
- Anomaly Detection
- Link Prediction
- …

Multi-Network Data

Social Domain

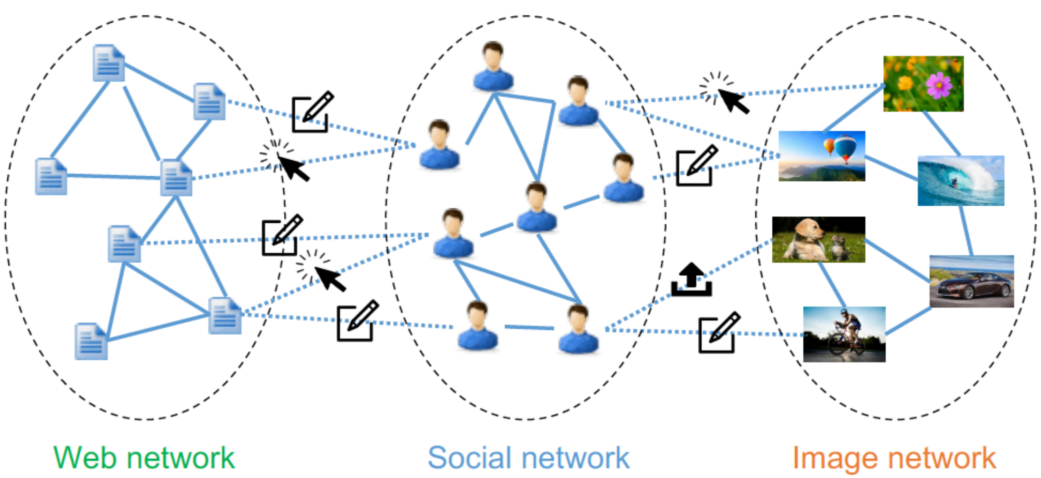
Case 1: multiple social networks



Common users: U1, U2, U3, …

Unique users: U7, U8, …

Case 2: inter-connected domain networks



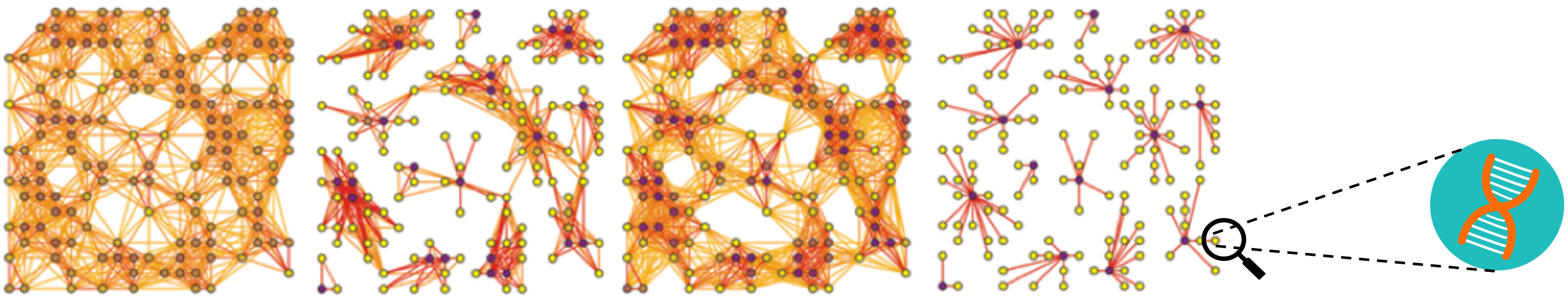
: Within-network relationship

: Cross-network relationship

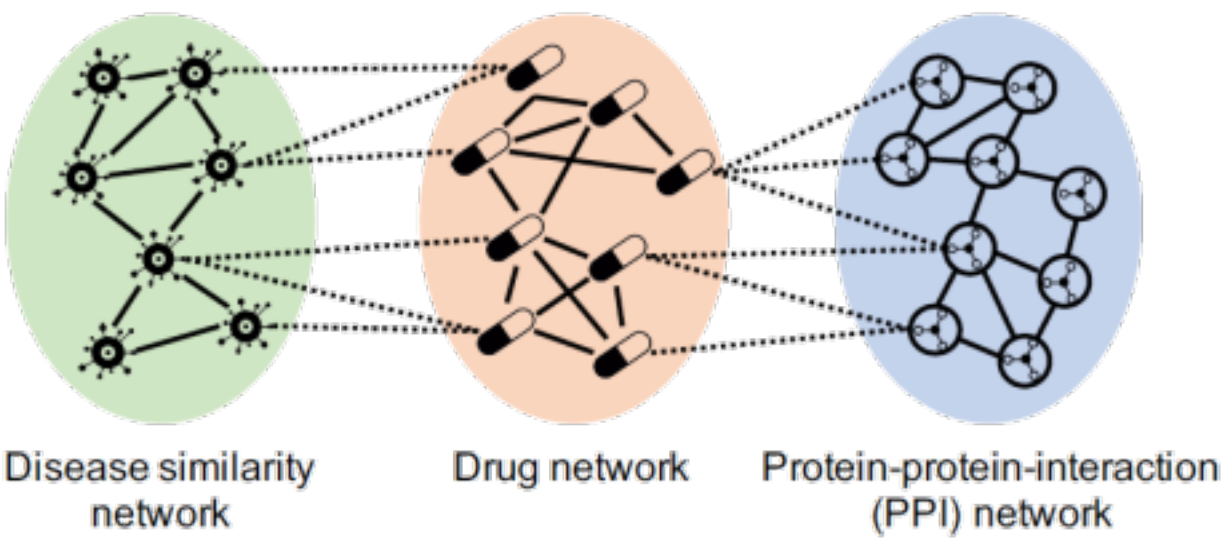
Multi-Network Data

Scientific Domain

Case 1: gene co-expression networks from multiple tissues



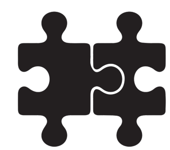
Case 2: inter-connected medical networks



Multi-Network Embedding

Motivation

- ✓ Wide applications 

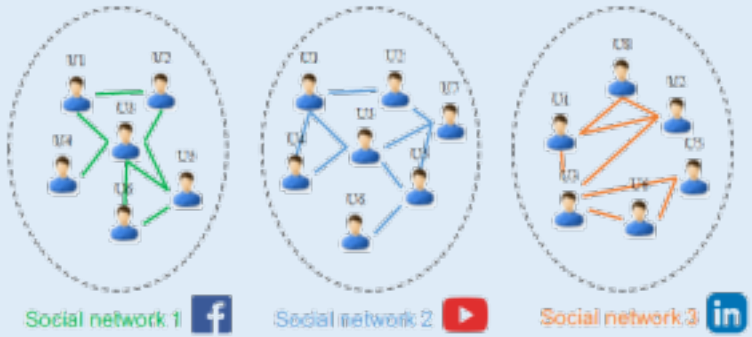
- ✓ Complementary information 

- ✓ Robustness 

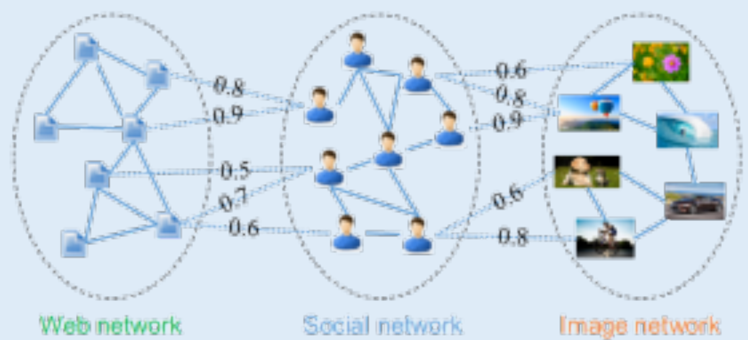
Multi-Network Embedding

Challenges

Case 1: multiple social networks



Case 2: inter-connected domain networks



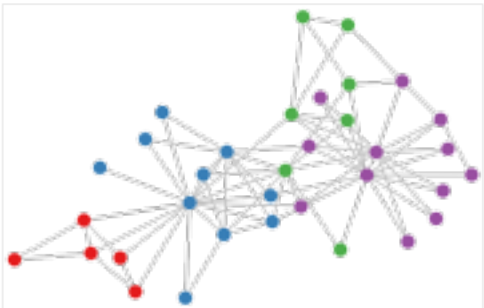
A general example

- Networks:
  - ➢ different sizes
- Cross-network relationships:
  - ➢ many-to-many
  - ➢ weighted
  - ➢ Incomplete

Both Case 1 & 2 can be represented by the general example.

Deep Multi-Network Embedding (DMNE)

Preliminary: Structural Context Extraction[1]

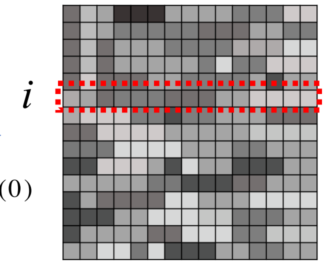


Random Walk from Each Node

$$\mathbf{p}^{(k)} = c\mathbf{p}^{(k-1)}[(\mathbf{D})^{-1}\mathbf{G}] + (1-c)\mathbf{p}^{(0)}$$

Degree matrix

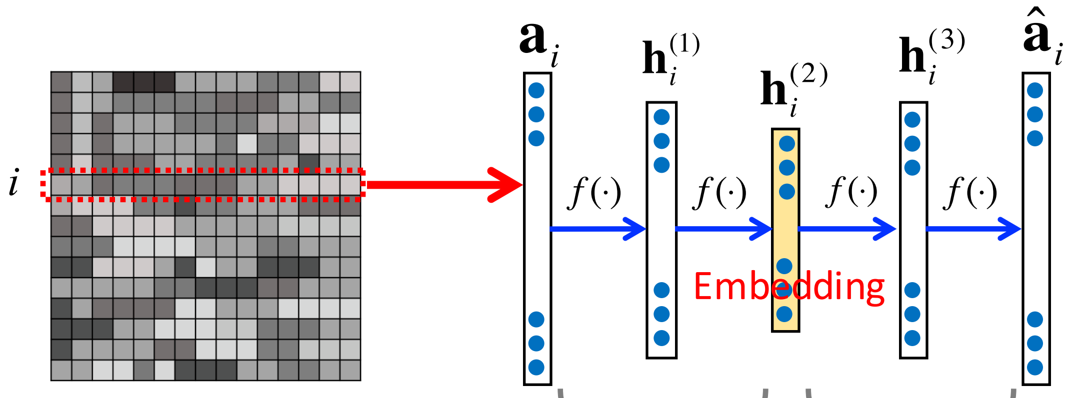
$$\mathbf{D}_{xx} = \sum_{y=1}^{n} \mathbf{G}_{xy}$$



$\mathbf{A}$

Local community of node i

Network Embedding: Deep Model[1,2]



Input matrix $\mathbf{A}$

Encoder

Decoder

Reconstruction error

$$\min_{\{\mathbf{W}_l, \mathbf{b}_l\}_{l=1}^{L}} \left\|\mathbf{A} - \hat{\mathbf{A}}\right\|_F^2 + \lambda \sum_{l=1}^{L} \left\|\mathbf{W}_l\right\|_F^2$$

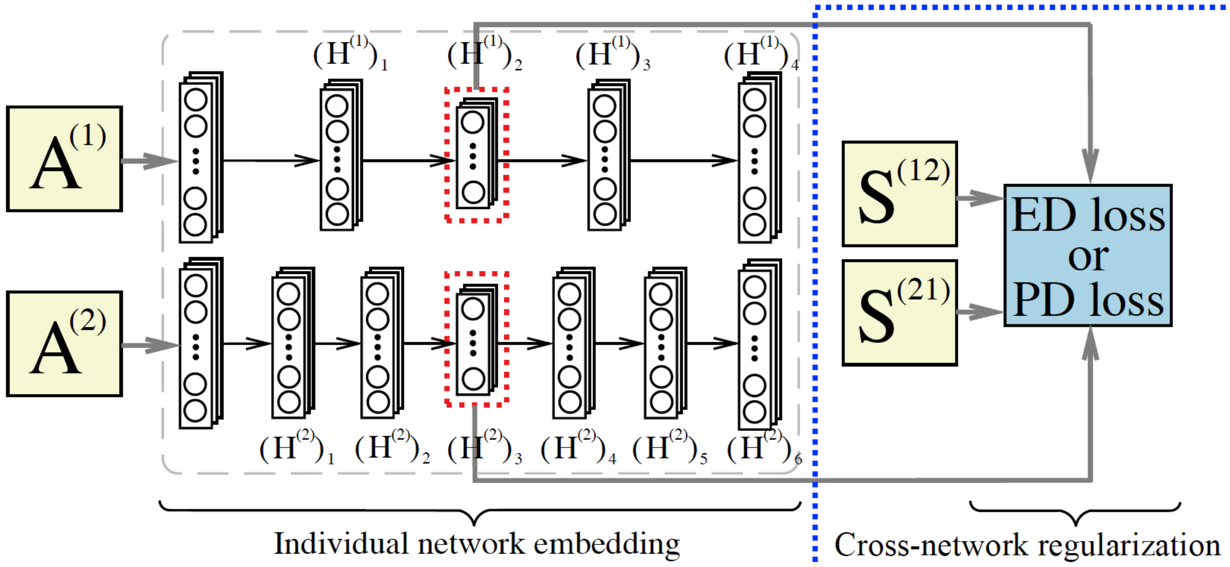
Reconstruction of $\mathbf{A}$

Activation function

$$f(\mathbf{x}) = \sigma(\mathbf{W}_l \mathbf{x} + \mathbf{b}_l)$$

1. S. Cao, W. Lu, and Q. Xu. *Deep Neural Networks for Learning Graph Representations*. In AAAI, 2016.
2. D. Wang, P. Cui, and W. Zhu. *Structural deep network embedding*. In KDD, 2016.

DMNE: Architecture



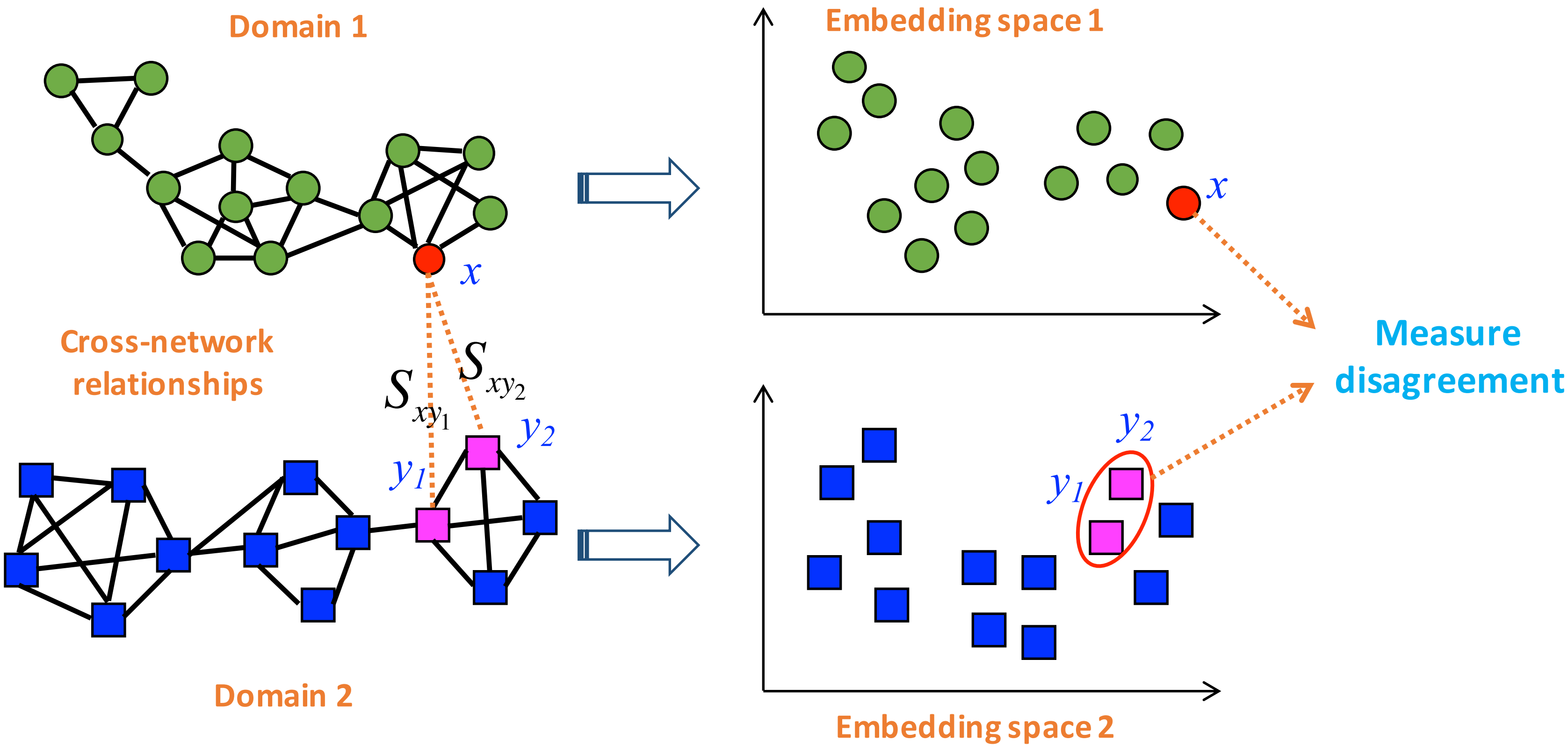
Coordinate Multiple Embedding Spaces

Objective function

$$\min_{\{\theta^{(i)}\}_{i=1}^{g}} L = \underbrace{\sum_{i=1}^{g} L_A^{(i)}}_{\text{Individual network embedding}} + \alpha \underbrace{\sum_{(i,j)\in I} L_R^{(ij)}}_{\text{Cross-network regularization}}$$

DMNE: Cross-Network Regularization

Method #1: Embedding Disagreement (ED)



Loss function

$$\min \left\| \mathbf{h}_x^{(1)} - \mathbf{h}_x^{(1\to 2)} \right\|_F^2 \qquad \mathbf{h}_x^{(1\to 2)} = \frac{\sum_y \mathbf{S}_{xy}^{(12)} \mathbf{h}_y^{(2)}}{\sum_y \mathbf{S}_{xy}^{(12)}}$$

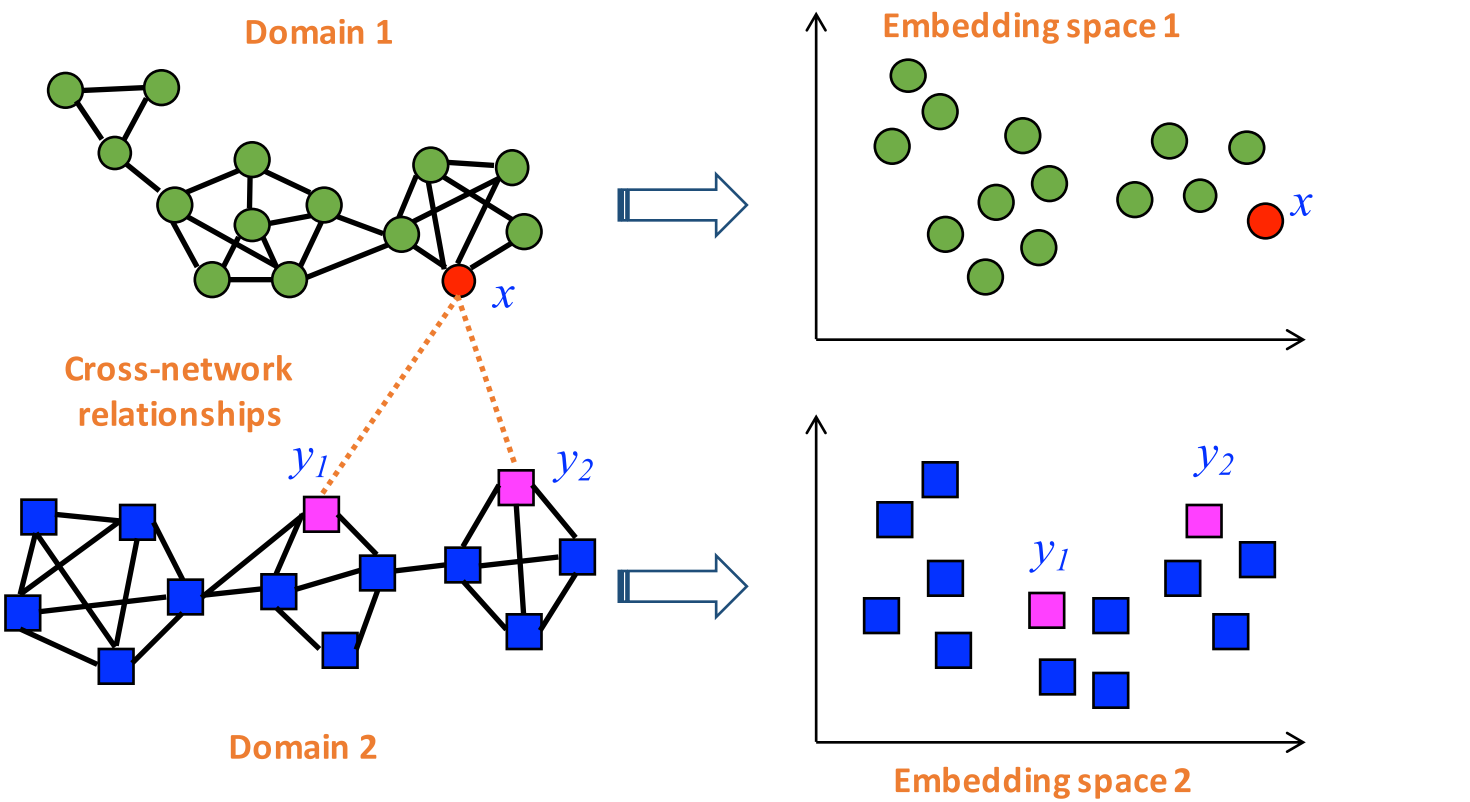
Matrix form

$$L_{ED}^{(12)} = \left\| \mathbf{O}^{(12)} \mathbf{H}^{(1)} - \tilde{\mathbf{S}}^{(12)} \mathbf{H}^{(2)} \right\|_F^2$$

$\mathbf{S}_{xy}^{(12)}$: link weight between node x and y

DMNE: Cross-Network Regularization

Method #1: Embedding Disagreement (ED)

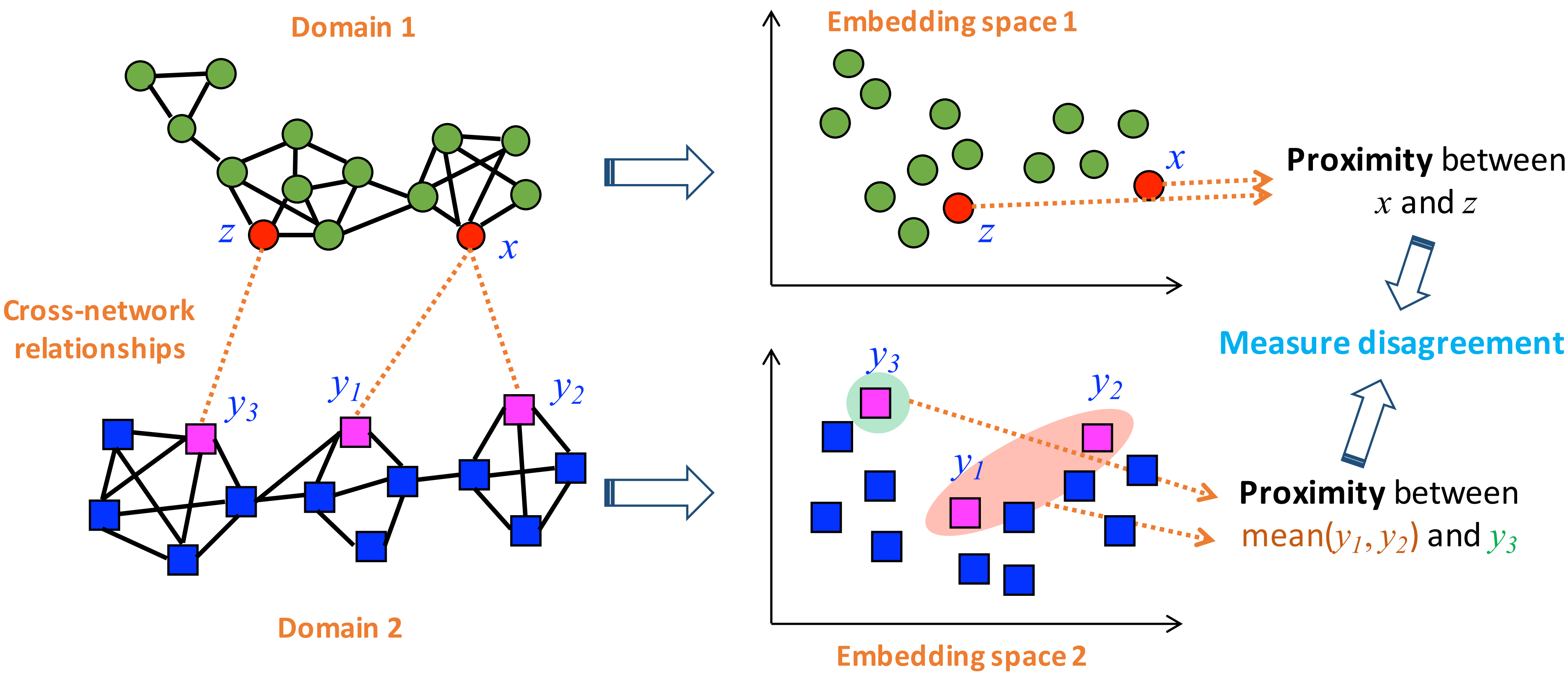


If y_1 and y_2 are far from each other in Domain 2

mean(y_1, y_2) consistent embedding(x)

DMNE: Cross-Network Regularization

Method #2: Proximity Disagreement (PD)



Proximity between x and z

$$\min \quad [(\mathbf{h}_x^{(1)})^T \mathbf{h}_z^{(1)} - (\mathbf{h}_x^{(1\rightarrow 2)})^T \mathbf{h}_z^{(1\rightarrow 2)}]^2$$

Matrix form

$$L_{PD}^{(12)} = \left\| (\mathbf{O}^{(12)}\mathbf{H}^{(1)})^T (\mathbf{O}^{(12)}\mathbf{H}^{(1)}) - (\tilde{\mathbf{S}}^{(12)}\mathbf{H}^{(2)})^T (\tilde{\mathbf{S}}^{(12)}\mathbf{H}^{(2)}) \right\|_F^2$$

Experiments: Datasets

Dataset	#Networks	#Nodes	#Links	#CrossLinks	LabeledNet.	#Classes
6-NG	*5*	*4,500*	*16,447*	*66,756*	*All*	*6*
9-NG	*5*	*6,750*	*24,778*	*100,585*	*All*	*9*
DP-NET	*2*	*13,583*	*51,918*	*2,107*	*Disease*	*18*
DBIS	*2*	*24,535*	*85,184*	*38,035*	*Collaboration*	*4*
CiteSeer-M10	*3*	*15,533*	*56,548*	*11,828*	*Collaboration*	*10*

- **6-NG**, **9-NG** (Document Network)
  - From 20-Newsgroup
  - Edge weight: cosine similarity between *tf-idf* vectors
  - Each network is a K-NN graph (k = 5)

Class Names[3,4]

- **6-NG**: *alt.atheism, comp.sys.mac.hardware, rec.motorcycles, rec.sport.hockey, soc.religion.christian, talk.religion.misc.*
- **9-NG**: *talk.politics.mideast, talk.politics.misc, comp.os.ms-windows.misc, comp.sys.ibm.pc.hardware, sci.electronics, sci.crypt, sci.med, sci.space, misc.forsale.*

3. F. Tian, B. Gao, Q. Cui, E. Chen, and T. Liu. *Learning deep representations for graph clustering*. In AAAI, 2014.
4. S. Cao, W. Lu, and Q. Xu. *Deep Neural Networks for Learning Graph Representations*. In AAAI, 2016.

Experiments: Datasets

Dataset	#Networks	#Nodes	#Links	#CrossLinks	LabeledNet.	#Classes
6-NG	*5*	*4,500*	*16,447*	*66,756*	*All*	*6*
9-NG	*5*	*6,750*	*24,778*	*100,585*	*All*	*9*
DP-NET	*2*	*13,583*	*51,918*	*2,107*	*Disease*	*18*
DBIS	*2*	*24,535*	*85,184*	*38,035*	*Collaboration*	*4*
CiteSeer-M10	*3*	*15,533*	*56,548*	*11,828*	*Collaboration*	*10*

- **DP-NET** (Disease-Protein Network)
  - A disease network: 5,080 nodes, 19,729 links
  - A protein interaction network: 8,503 nodes, 32,189 links
  - # Disease-Protein Links: 2,107 (*many-to-many*)

Experiments: Datasets

Dataset	#Networks	#Nodes	#Links	#CrossLinks	LabeledNet.	#Classes
6-NG	*5*	*4,500*	*16,447*	*66,756*	*All*	*6*
9-NG	*5*	*6,750*	*24,778*	*100,585*	*All*	*9*
DP-NET	*2*	*13,583*	*51,918*	*2,107*	*Disease*	*18*
DBIS	*2*	*24,535*	*85,184*	*38,035*	*Collaboration*	*4*
CiteSeer-M10	*3*	*15,533*	*56,548*	*11,828*	*Collaboration*	*10*

- **DBIS** (Author-Paper Network)
  - A collaborator network: 12,002 nodes, 37,587 links
  - A paper similarity network: 12,533 nodes, 47,597 links
  - # Author-Paper Links: 38,035 (*many-to-many*)

Experiments: Datasets

Dataset	#Networks	#Nodes	#Links	#CrossLinks	LabeledNet.	#Classes
6-NG	*5*	*4,500*	*16,447*	*66,756*	*All*	*6*
9-NG	*5*	*6,750*	*24,778*	*100,585*	*All*	*9*
DP-NET	*2*	*13,583*	*51,918*	*2,107*	*Disease*	*18*
DBIS	*2*	*24,535*	*85,184*	*38,035*	*Collaboration*	*4*
CiteSeer-M10	*3*	*15,533*	*56,548*	*11,828*	*Collaboration*	*10*

- **CiteSeer-M10** (Author-Paper-Paper Network)
  - A collaborator network: 3,284 nodes, 13,781 links
  - A paper similarity network: 10,214 nodes, 39,411 links
  - A paper citation network: 2,035 nodes, 3,356 links
  - # Author-Paper Links: 7,173 & 2,634 (*many-to-many*)
  - # Paper-Paper Cross-Links: 2,021 (*one-to-one*)

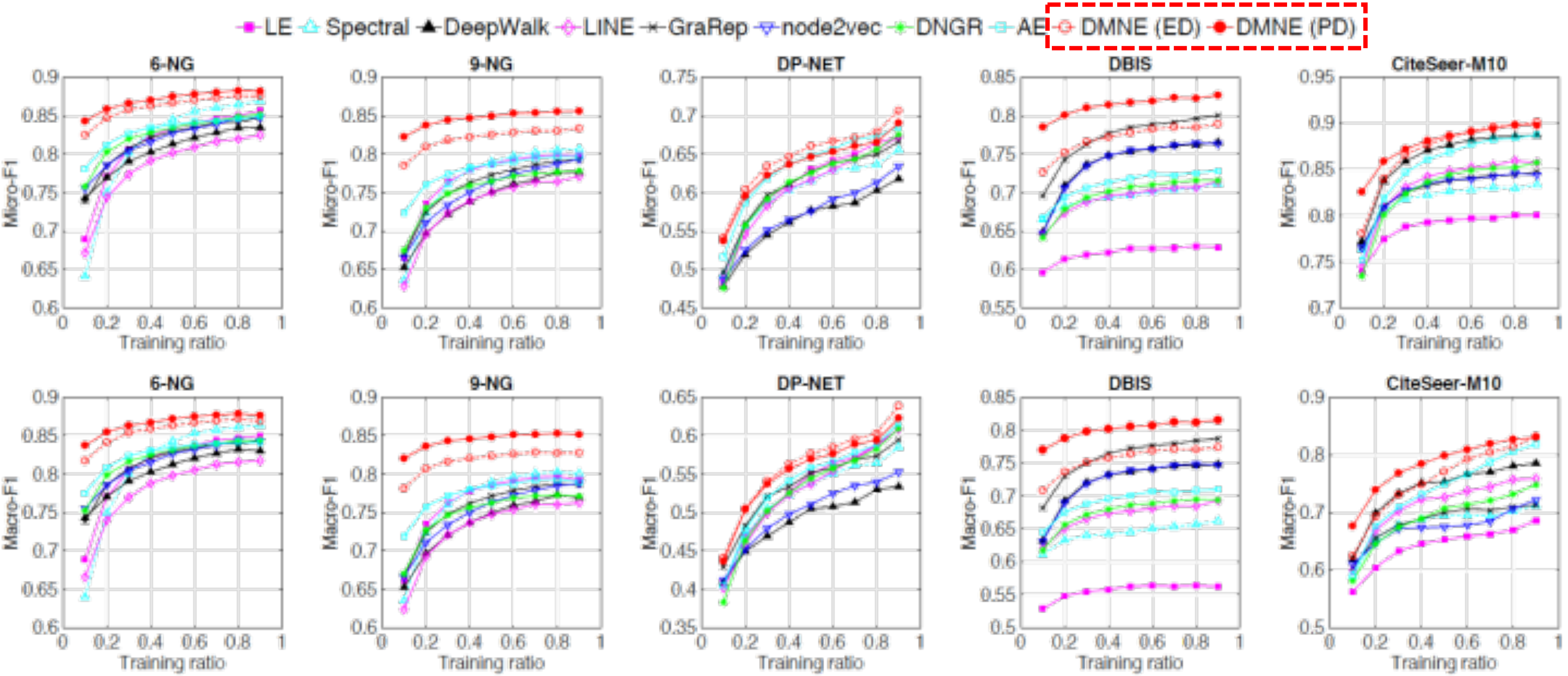
Experiments: Multi-Label Classification

Dataset	#Networks	#Nodes	#Links	#CrossLinks	LabeledNet.	#Classes
6-NG	*5*	*4,500*	*16,447*	*66,756*	*All*	*6*
9-NG	*5*	*6,750*	*24,778*	*100,585*	*All*	*9*
DP-NET	*2*	*13,583*	*51,918*	*2,107*	*Disease*	*18*
DBIS	*2*	*24,535*	*85,184*	*38,035*	*Collaboration*	*4*
CiteSeer-M10	*3*	*15,533*	*56,548*	*11,828*	*Collaboration*	*10*

- **Compared Methods**
  - **LE**: Laplacian Eigenmaps [Belkin and Niyogi, *Neural Comput.'03*]
  - **Spectral**: Spectral clustering [Shi and Malik, *TPAMI'00*]
  - **DeepWalk** [Perozzi et al., *KDD'14*]
  - **LINE** [Tang, et al., *WWW'15*]
  - **GraRep** [Cao et al., *CIKM'15*]
  - **node2vec** [Grover and Leskovec, *KDD'16*]
  - **DNGR** [Cao et al., *AAAI'16*]
  - **AE**: AutoEncoder [Hinton and Salakhutdinov, *Science'06*]

Experiments: Multi-Label Classification

Proposed methods



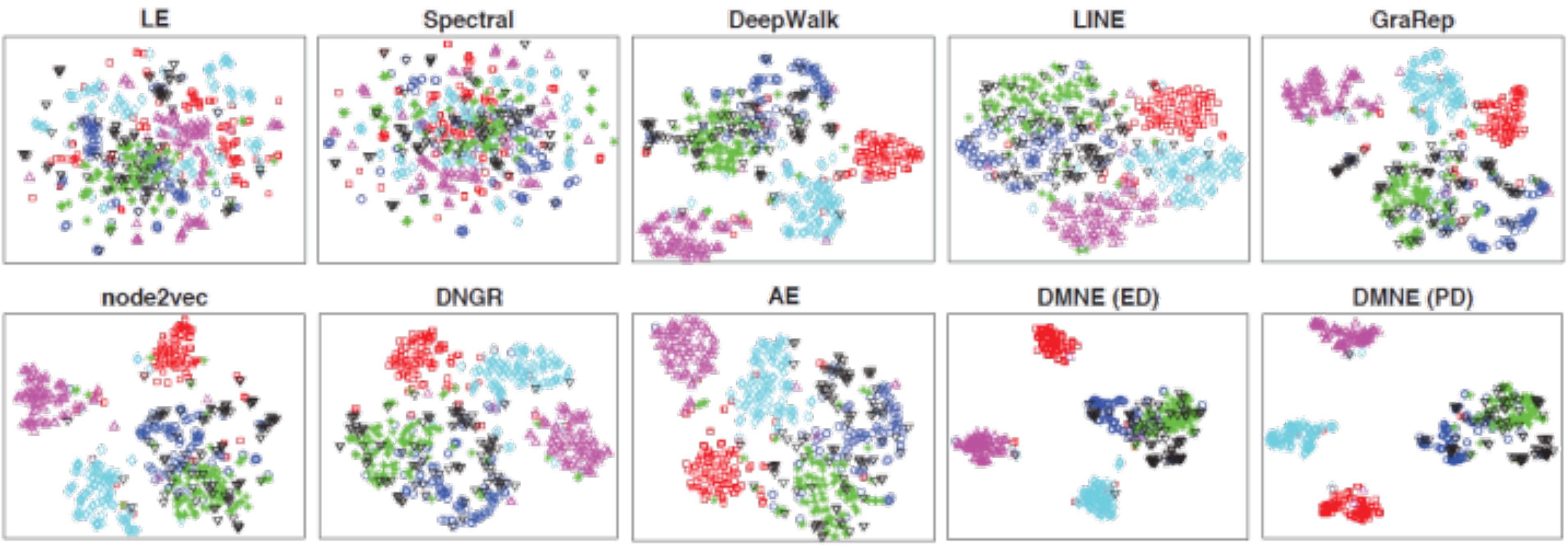
Feed learned embeddings (100-D) to SVM[5]

- **Evaluation Method**: Micro-F1 Score (1st Row) & Macro-F1 Score (2nd Row)
- **Setting**: varying training ratio from 0.1 to 0.9

5. R., Fan, K. Chang, C. Hsieh, X. Wang, and C. Lin. *LIBLINEAR: A library for large linear classification*. J. Mach. Learn. Res., 2008.

Experiments: Visualization

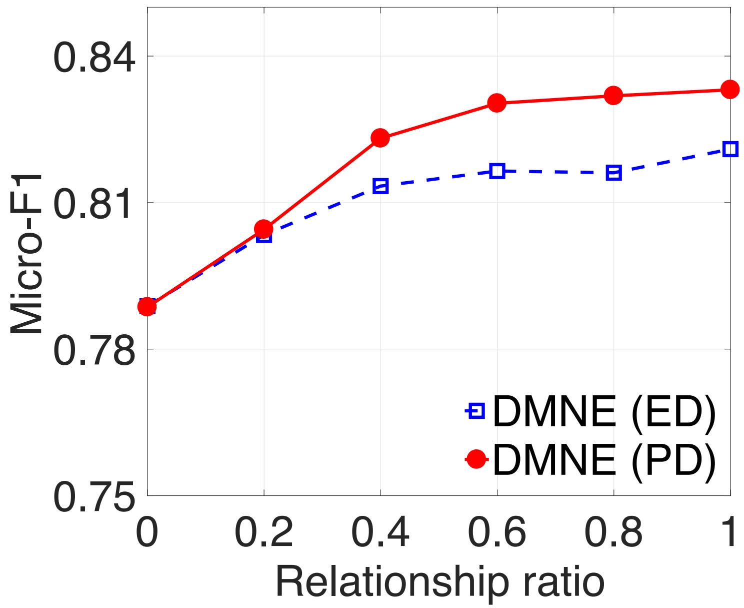
- **Dataset**: the first network of 6-NG (6 classes)
- **Method**: use t-SNE[6] to project all embeddings to a 2-D space



6. L. Maaten, and G. Hinton. *Visualizing data using t-SNE*. J. Mach. Learn. Res., 2008.

Experiments: Insights of Effectiveness

- Dataset: 6-NG
- Cross-network relationships ➔ 5 equal parts (20% each)
- Each time ➔ add one part



Meaningful relationships ➔ better performance

Conclusion

- ✓ Investigate a general multi-network embedding problem
- ✓ Propose effective algorithms: DMNE (ED) and DMNE (PD)
- ✓ Experimental evaluations

Thanks!