Contrastive Multi-View Multiplex Network Embedding with Applications to Robust Network Alignment

Hao Xiong, Junchi Yan, Li Pan

Shanghai Jiao Tong University

Outline

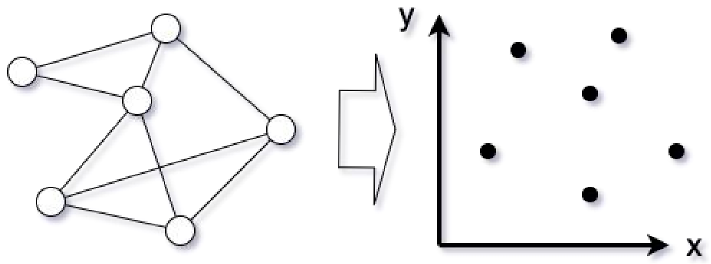
- Background
  - Network embedding and multiplex networks
- Motivations
  - Two challenges in multiplex network embedding
- The Framework: **cM^{2}NE**
- Experiments
- Conclusion

Outline

- Background
  - Network embedding and multiplex networks
- Motivations
  - Two challenges in multiplex network embedding
- The Framework: **cM^2NE**
- Experiments
- Conclusion

Network Embedding (NE)

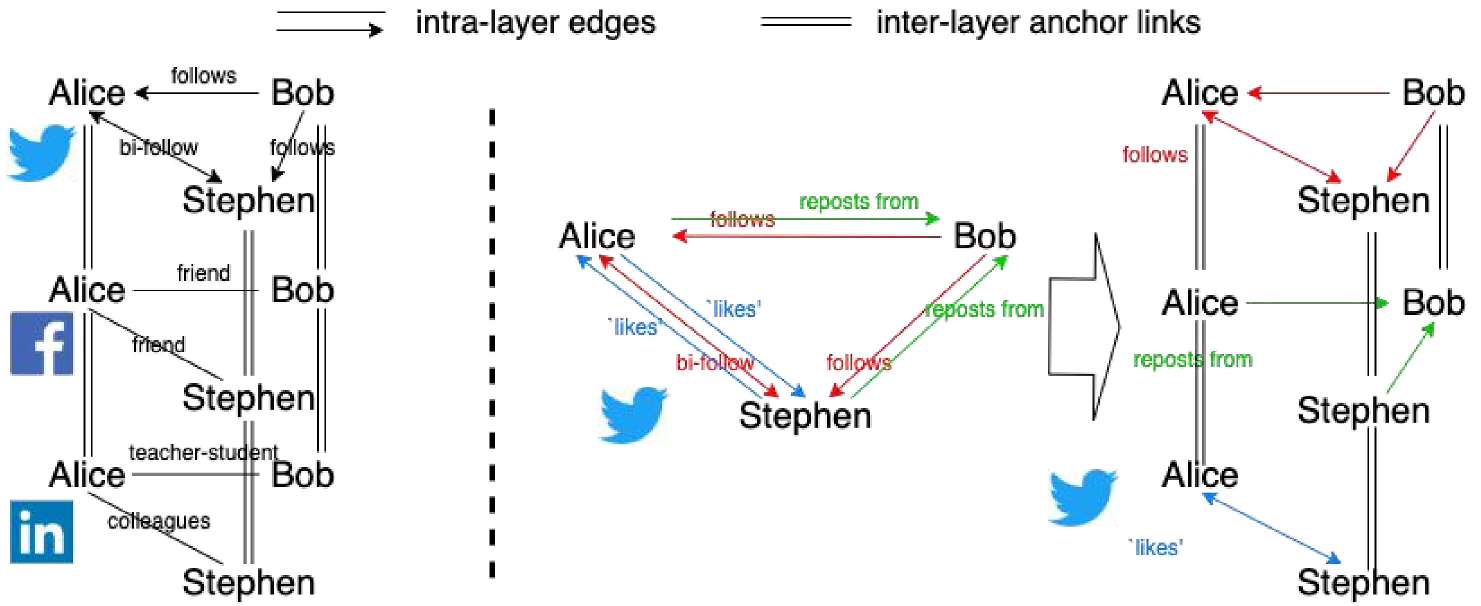
- To represent nodes with low-dimensional vectors
- Why NE?
  - Scalable and easy to parallel
    - E.g. deepwalk [1]
  - Can apply advanced ML algorithms on downstream tasks
    - Classify nodes based on embeddings and labels [1]
    - Align nodes across networks [2]
    - Predict unseen links [3]



[1] Perozzi B, Al-Rfou R, Skiena S. Deepwalk: Online learning of social representations (SIGKDD'2014)
[2] Liu L, Cheung W K, Li X, et al. Aligning Users across Social Networks Using Network Embedding (IJCAI'2016)
[3] Tang J, Qu M, Wang M, et al. Line: Large-scale information network embedding (WWW'2015)

Multiplex Networks (MNs)

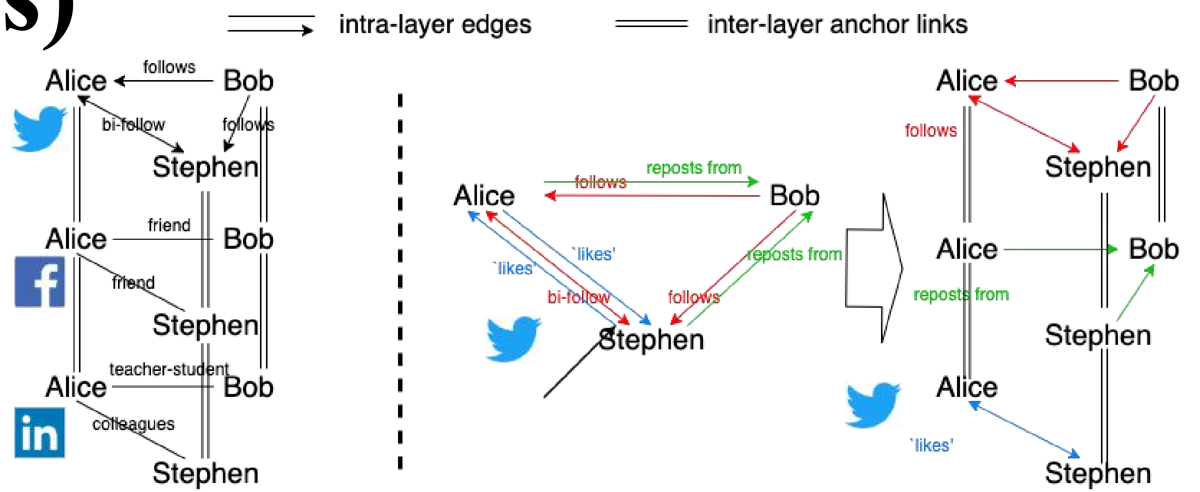
- Multiplex networks
  - Multiple layers, each layer defines one type of interactions
  - Inter-layer *anchor links*



A multiplex network across social platforms

A multiplex network with one social platform

Multiplex Networks (MNs)



A multiplex network across social platforms

A multiplex network with one social platform

- Multiplex networks
  - Multiple layers
  - Inter-layer *Anchor links*
- Definitions
  - A layer: $G = (\mathcal{V}, \mathbf{A})$
  - Anchor links between source layer G^s and target layer G^t: $\mathcal{T}^{s,t} \subseteq \mathcal{V}^s \times \mathcal{V}^t$
  - A multiplex network of N layers: $\mathcal{G} = \{G^g\}_{g=1}^N$
  - Anchor link sets of the MN: $\{\mathcal{T}^{s,t}\}_{s,t=1}^N$

Outline

- Backgrounds
  - Network Embedding and multiplex networks
- Motivations
  - Two challenges in multiplex network embedding
- The Framework: **cM²NE**
- Experiments
- Conclusion

Multiplex Networks Embedding (MNE)

- Challenge 1: intra-layer edges are missing
  - Solution: *multiple structural views* for each layer as data augmentation
    - Random walk
    - Random with restart (Personalize PageRank)
  - How to select views?
    - Low-order information or high-order information?

Multiplex Networks Embedding (MNE)

- Challenge 2: consistency assumption on inter-layer anchor links can be misleading
  - Solution: emphasize the anchor links tending to represent `agreement' across layers and de-emphasize those tending to represent `disagreement'
  - How to determine whether an anchor link represents `agreement' or `disagreement'?

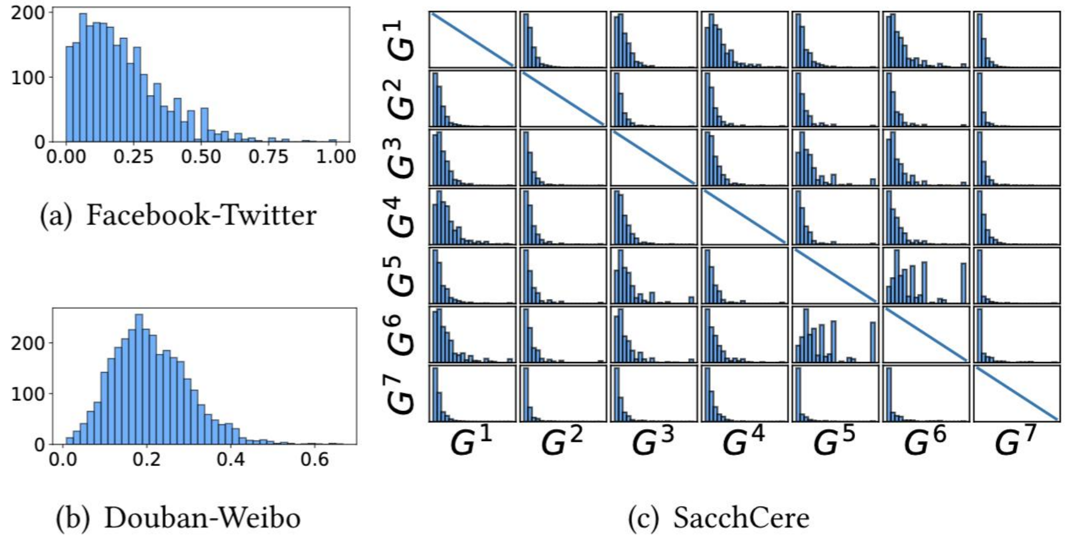
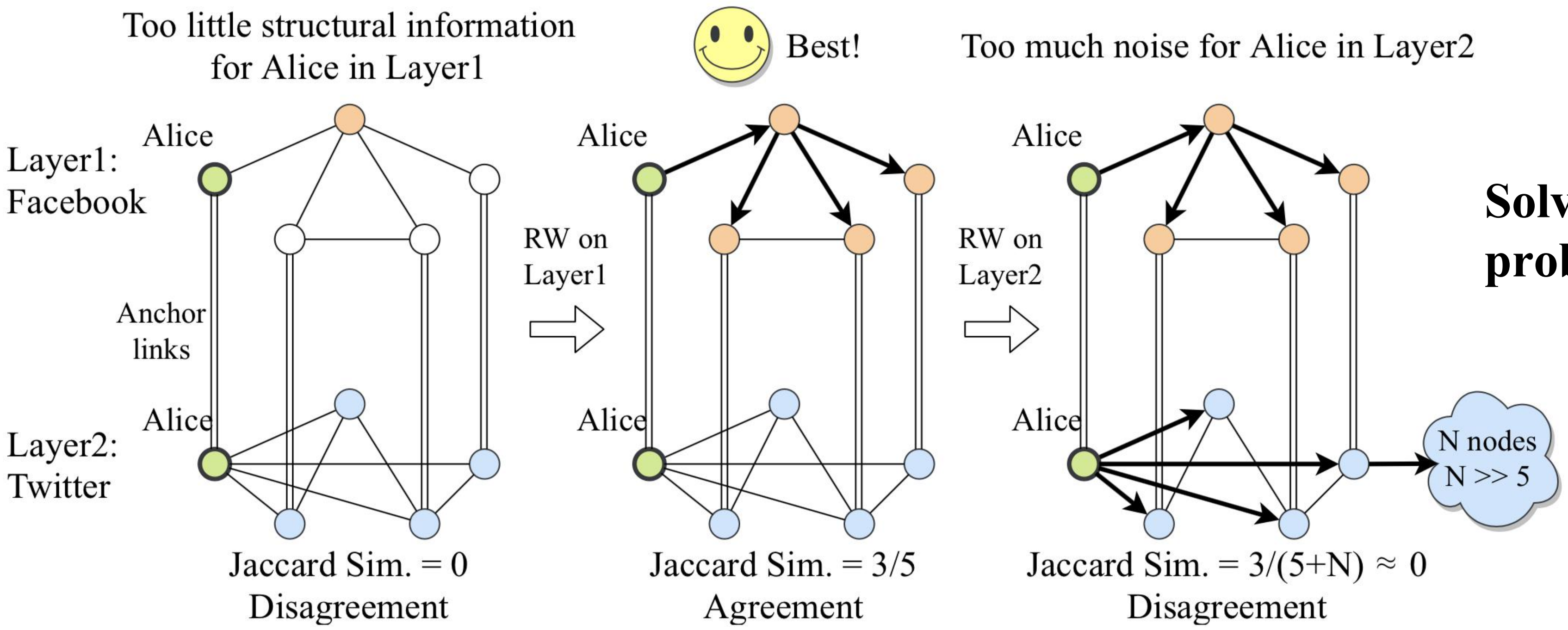


Figure 10: The distribution of Jaccard Similarities between layers in several multiplex networks. In each subplot, x-axis denotes Jaccard Similarity, and y-axis is the number of anchor links whose Jaccard Similarity are in a certain interval.

Multiplex Networks Embedding (MNE)

- How to select views?
- How to determine `agreement/disagreement'?



Solve the two problems jointly

Fig. A toy example of one ideal solution

A Brief View of cM²NE

- Three levels of contrastive learning
  - Intra-view
  - Inter-view
  - Inter-layer

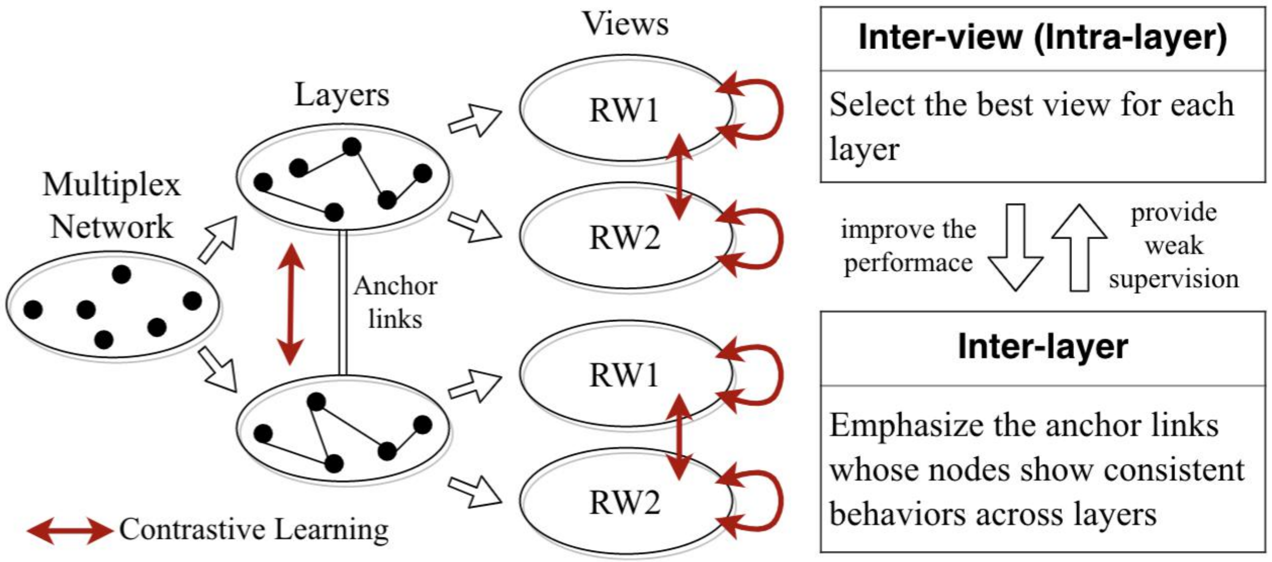


Figure 2: On the left is the sketch of our learning framework, where contrastive learning is performed on intra-view, inter-view, and inter-layer level. On the right is our main motivations of framework design and the inside connections.

Outline

- Backgrounds
  - Network Embedding and multiplex networks
- Motivations
  - Two challenges in multiplex network embedding
- The Framework: **cM²NE**
- Experiments
- Conclusion

Contrastive Learning (CL) for NE

- The (predictive) mutual information [1] between two sets of embeddings:

$$I(\mathbf{X}^{(t)};\mathbf{X}^{(s)}) = \mathbb{E}_{(i,j)\sim p_{pos}} I(\mathbf{x}_j^{(t)};\mathbf{x}_i^{(s)})$$

Node embeddings of the target G^t

Node embeddings of the source G^s

Distribution of positive node pairs

Embedding of node j on G^t

Embedding of node i on G^s

[1] Oord A, Li Y, Vinyals O. Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748, 2018.

Contrastive Learning (CL) for NE

- Mutual information estimators [1]:
  - NT-Xent $$I(\mathbf{x}_j;\mathbf{x}_i)=\log\frac{\exp(\mathbf{x}_i^\top\mathbf{x}_j/\tau)}{\exp(\mathbf{x}_i^\top\mathbf{x}_j/\tau)+\sum_{b=1}^{B}\mathbb{E}_{j'\sim p_{neg}}\exp(\mathbf{x}_i^\top\mathbf{x}_{j'}/\tau)}$$
  - NT-Logistic $$I(\mathbf{x}_j;\mathbf{x}_i)=\log\sigma(\mathbf{x}_i^\top\mathbf{x}_j/\tau)+\sum_{b=1}^{B}\mathbb{E}_{j'\sim p_{neg}}\log\sigma(-\mathbf{x}_i^\top\mathbf{x}_{j'}/\tau)$$
  - Marginal Triplets $$I(\mathbf{x}_j;\mathbf{x}_i)=-\sum_{b=1}^{B}\mathbb{E}_{j'\sim p_{neg}}\max(\mathbf{x}_i^\top\mathbf{x}_{j'}-\mathbf{x}_i^\top\mathbf{x}_j+\gamma,0)$$

Temperature

Distribution of negative node pairs

Number of negative samples

- Mutual information with alignment and uniformity loss [2]:
  - Alignment loss $\mathcal{L}_a(\mathbf{x}_i,\mathbf{x}_j)=\|\mathbf{x}_i-\mathbf{x}_j\|_2^{\rho}$ Explicitly shorten the distance of positive pair of nodes
  - Uniformity loss $\mathcal{L}_u(\mathbf{x}_i)=\log\mathbb{E}_{j'\sim p_{neg}}\exp(-\frac{\|\mathbf{x}_i-\mathbf{x}_{j'}\|_2^2}{\sigma_u})$ Scatter embeddings uniformly on the hypersphere

$$I_{AU}(\mathbf{x}_j;\mathbf{x}_i)=I(\mathbf{x}_j;\mathbf{x}_i)+\beta\Big(\mathcal{L}_a(\mathbf{x}_i,\mathbf{x}_j)+\mathcal{L}_u(\mathbf{x}_i)\Big)$$

(in the paper, we use I_{AU} instead of the vanilla mutual information)

[1] Chen T, Kornblith S, Norouzi M, et al. A simple framework for contrastive learning of visual representations. (ICML'20)
[2] Wang T, Isola P. Understanding contrastive representation learning through alignment and uniformity on the hypersphere. (ICML'2020)

The Framework: cM²NE

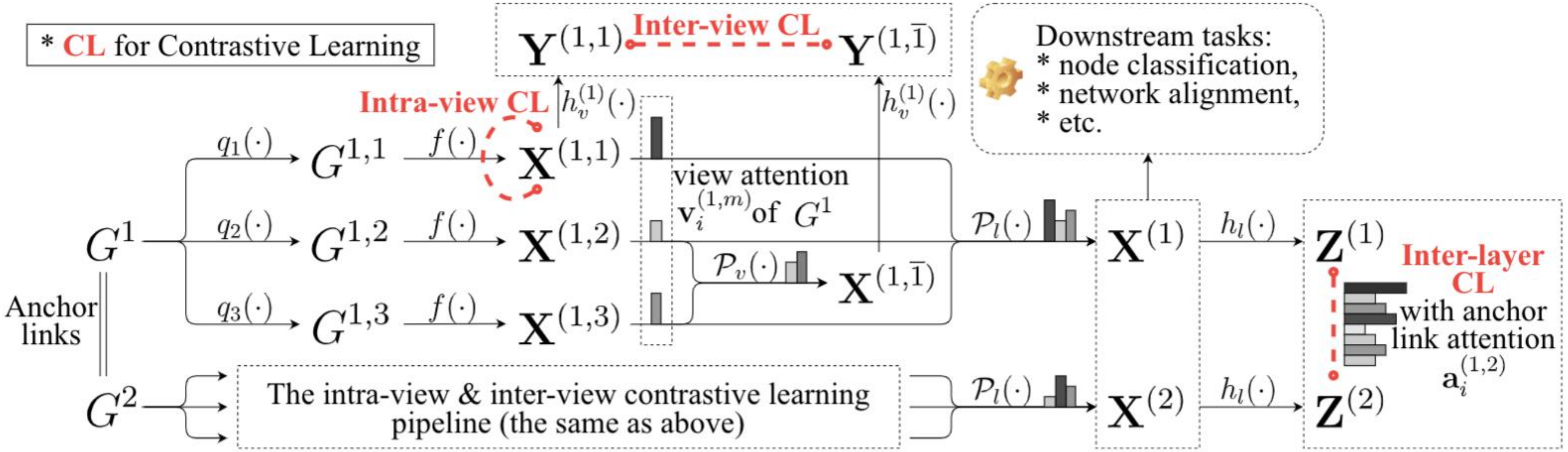
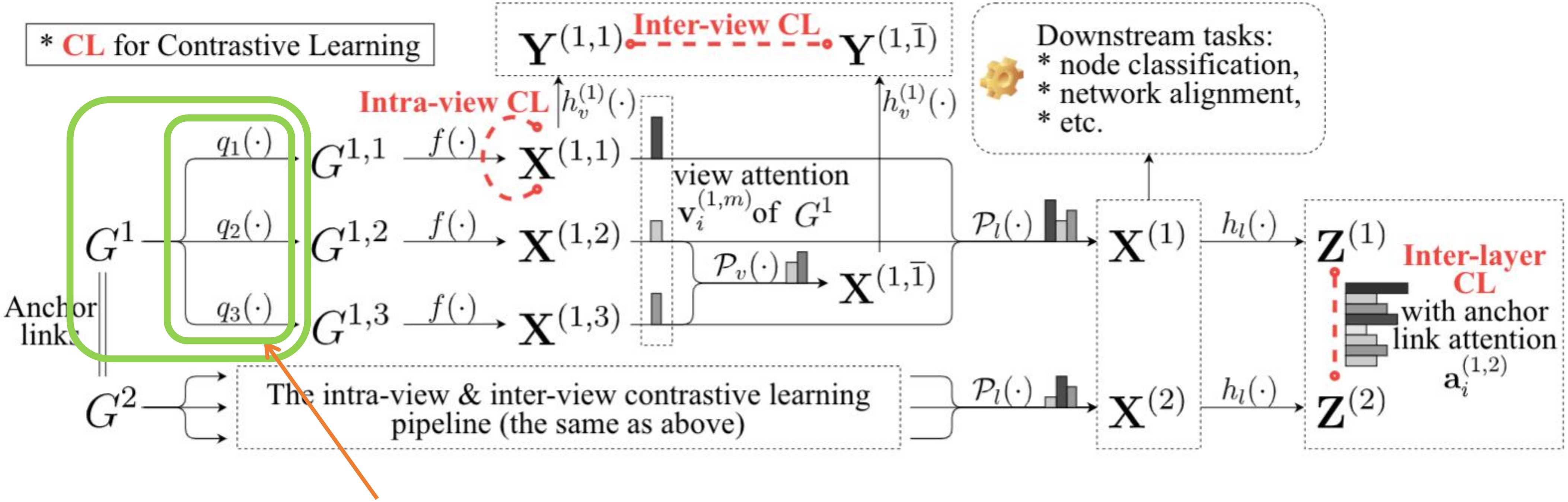


Figure 3: The cM²NE framework with multiplex network $\mathcal{G} = \{G^g\}_{g=1}^N$ as input and embeddings $\{X^{(g)}\}_{g=1}^N$ as output. For layer G^1, M multi-view augmentations $\{G^{1,m}\}_{m=1}^M$ are generated given functions $\{q_m(\cdot)\}_{m=1}^M$. Then by $f(\cdot)$, nodes in view $G^{1,m}$ are embedded into a low-dimensional space, where the embeddings are denoted as $X^{(1,m)}$. Then contrastive learning (CL) is performed on three levels: i) Intra-view CL is conducted directly on $X^{(1,m)}$ to preserve intra-view information. ii) For inter-view CL on layer G^1 between the m-th view and the others, first $\{X^{(1,k)}\}_{k=1,k\neq m}^M$ are aggregated together by inter-view *readout function* $\mathcal{P}_v(\cdot)$ whose results are denoted as $X^{(1,\overline{m})}$, then inter-view CL is performed after mapping $X^{(1,m)}$ and $X^{(1,\overline{m})}$ to $Y^{(1,m)}$ and $Y^{(1,\overline{m})}$ by *projection heads* $h_v^{(1)}(\cdot)$. iii) For inter-layer CL between G^1 and G^2, embeddings of multiple views are aggregated by inter-layer *readout function* $\mathcal{P}_l(\cdot)$ and we get embedding $X^{(1)}$, $X^{(2)}$, then they are mapped to $Z^{(1)}$, $Z^{(2)}$ for inter-layer CL.

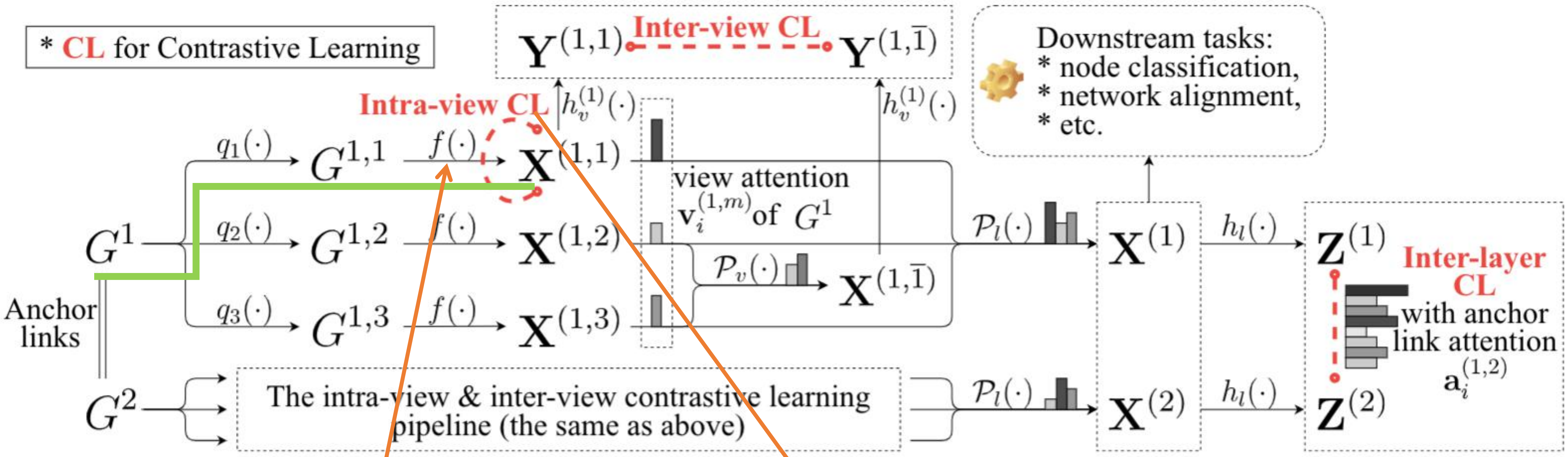
The Framework: cM²NE

- Data augmentation: generate M views for each layer



pre-designed functions to generate multiple structural views

Intra-view Contrastive Learning

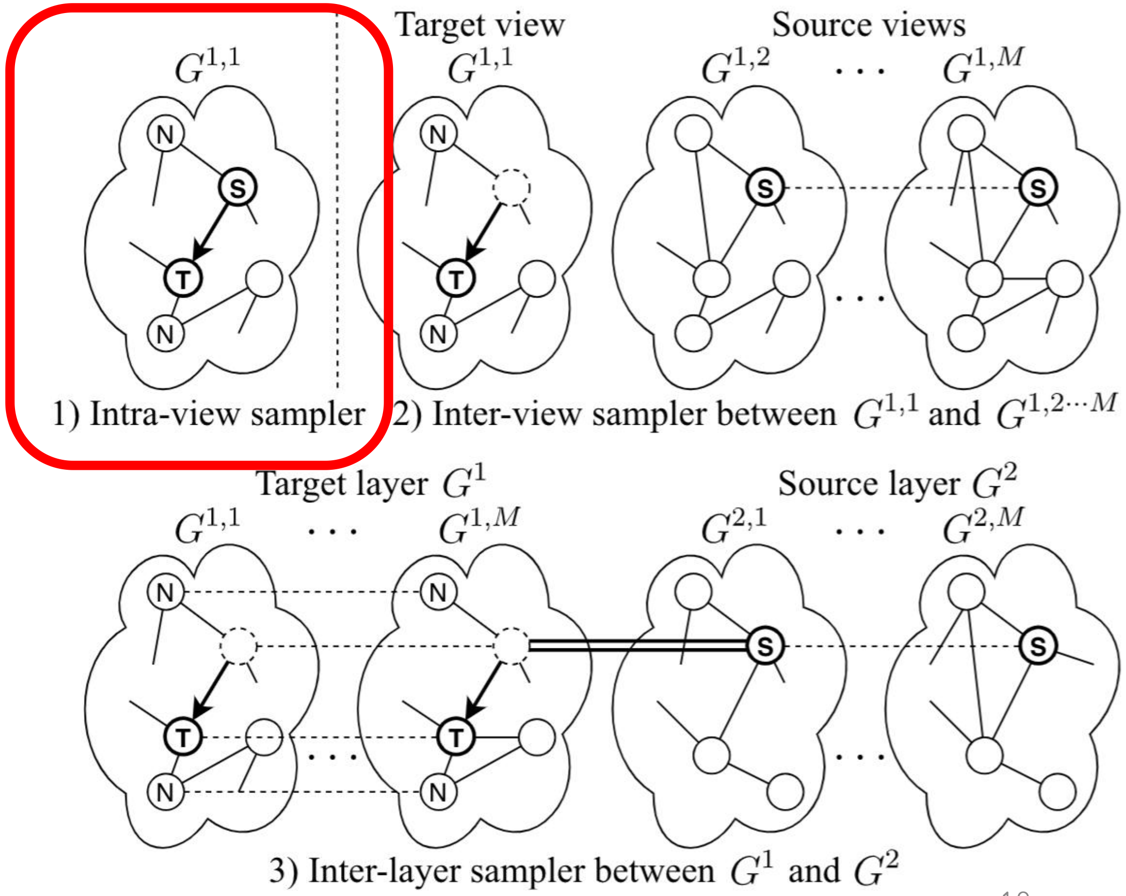


embedding look-up function

$$I(\mathbf{X}^{(1,1)}, \mathbf{X}^{(1,1)})$$

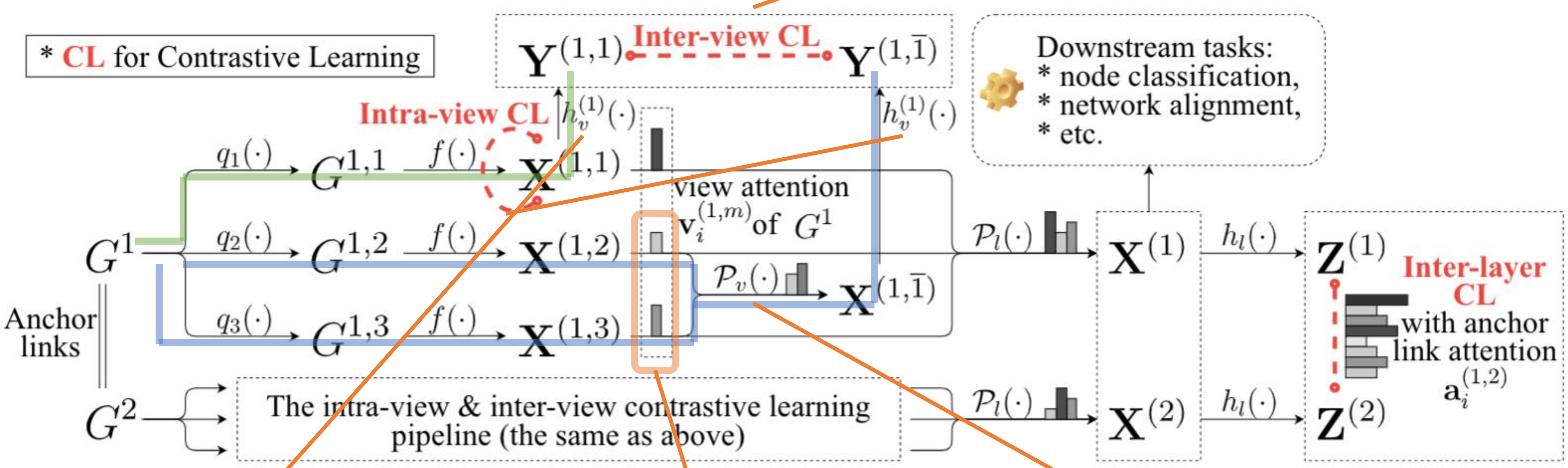
Intra-view Sampling

- Positive samples:
  - Directly sampled from $G^{1,1}$
- Negative samples:
  - On $G^{1,1}$
  - B negative nodes



Inter-view Contrastive Learning

$$I(\mathbf{Y}^{(1,1)}, \mathbf{Y}^{(1,\bar{1})})$$



projection heads to map node embeddings to inter-view space: identical/linear/non-linear mapping

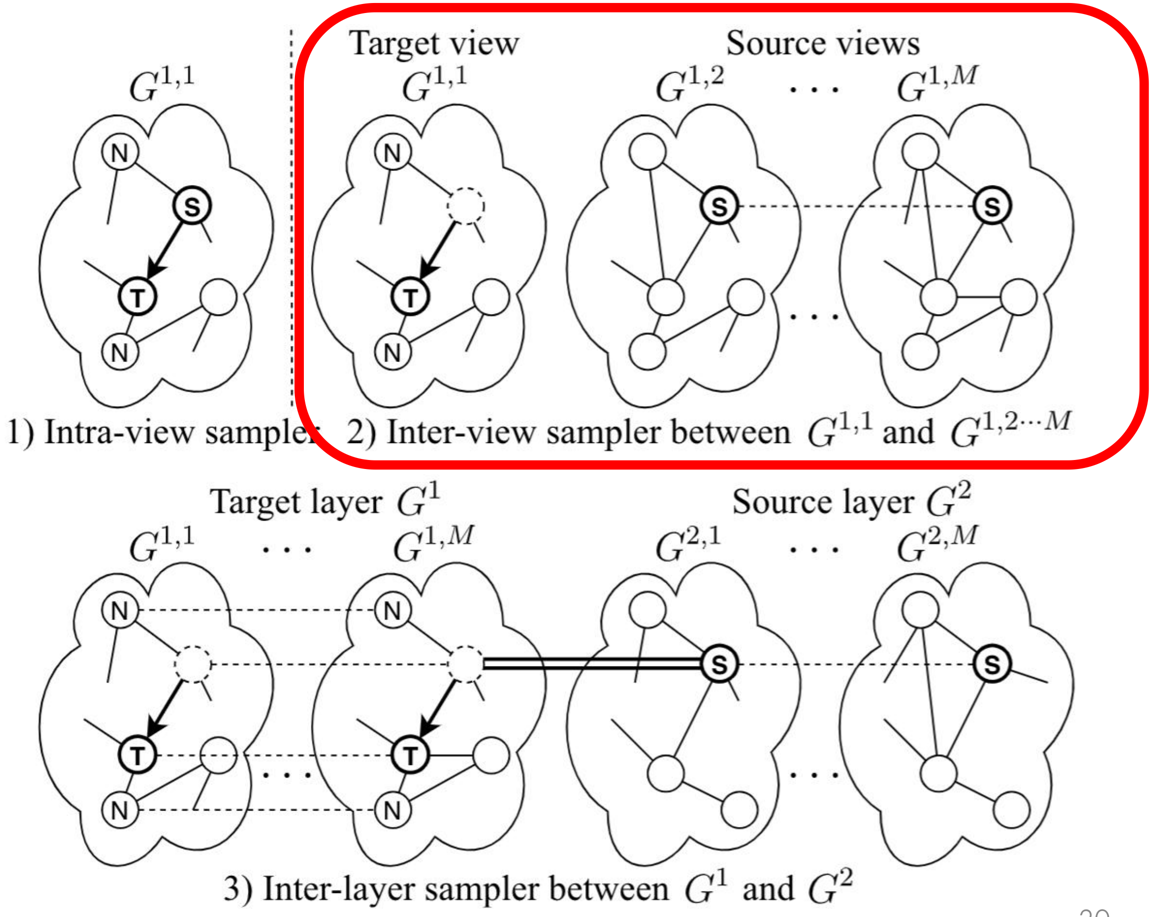
view attention

readout function (pooling):

$$\mathcal{P}_v(\{\mathbf{x}_i^{(g,k)}\}_{k=1,k\neq m}^{M}) = \sum_{k=1}^{M} \frac{\exp \mathbf{v}_i^{(g,k)}}{\sum_{n=1,n\neq m}^{M} \exp \mathbf{v}_i^{(g,n)}} \mathbf{x}_i^{(g,k)}$$

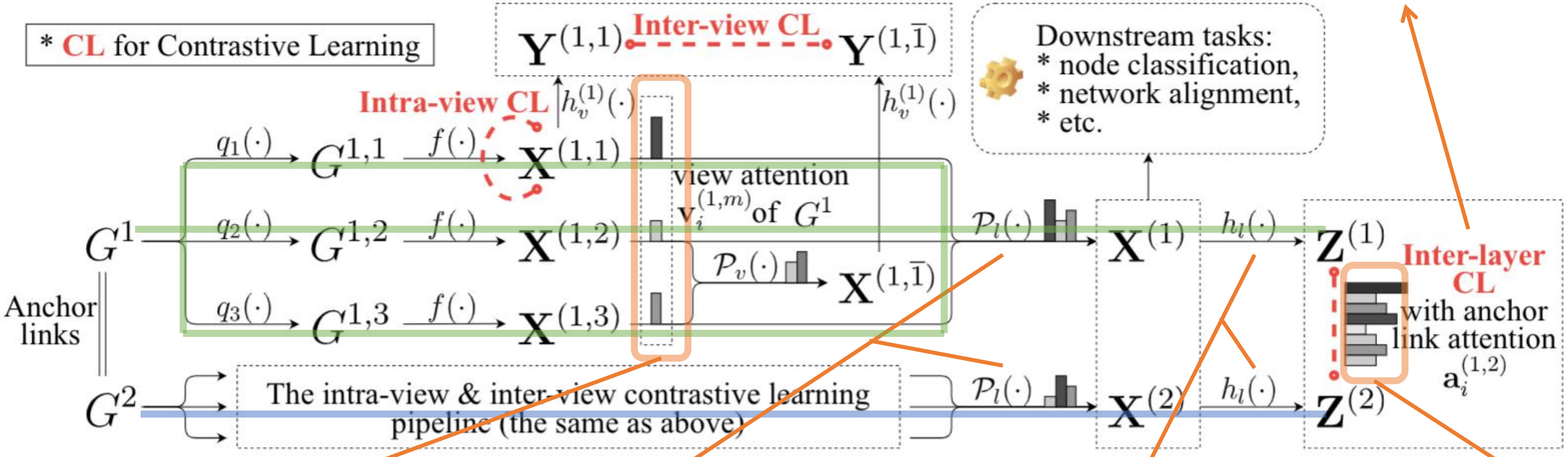
Inter-view Sampling

- Positive samples:
  - Step 1: sample one positive node pair from $G^{1,1}$
  - Step 2: map the source node to other views
  - 1 target node and M-1 source nodes from other views
- Negative samples:
  - On $G^{1,1}$
  - B negative nodes



Inter-layer Contrastive Learning

attention-enhanced mutual information: $\widetilde{I}(\mathrm{Z}^{(t)};\mathrm{Z}^{(s)}) = \underset{\mathcal{B}\sim\{(i,j)\sim p_{pos}\}^{bs}}{\mathbb{E}} \sum_{(i,j)\in\mathcal{B}} \frac{\exp \mathbf{a}_i^{(s,t)}}{\sum_{(i',\cdot)\in\mathcal{B}} \exp \mathbf{a}_{i'}^{(s,t)}} I(\mathbf{z}_j^{(t)};\mathbf{z}_i^{(s)})$



view attention

inter-view readout function (pooling):

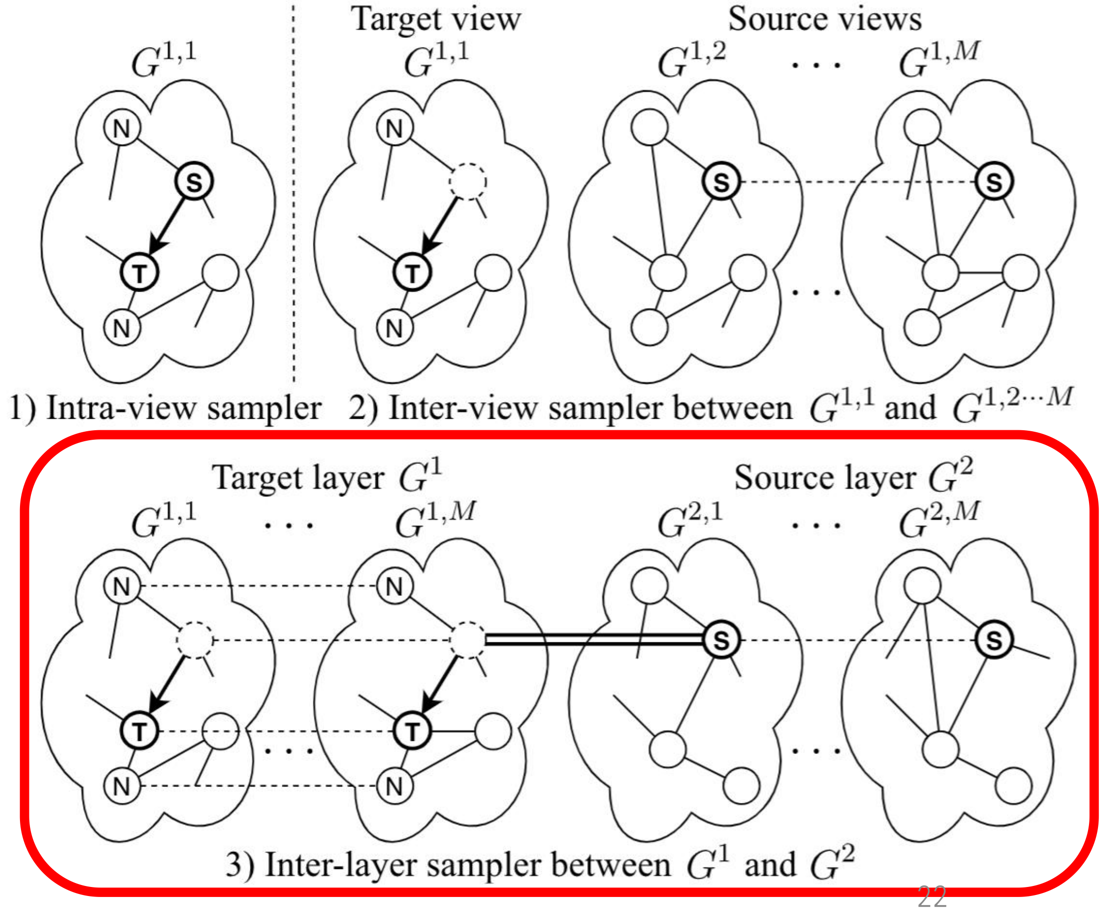
$$\mathcal{P}_l(\{\mathbf{x}_i^{(g,m)}\}_{m=1}^{M}) = \sum_{m=1}^{M} \frac{\exp \mathbf{v}_i^{(g,m)}}{\sum_{k=1}^{M} \exp \mathbf{v}_i^{(g,k)}} \mathbf{x}_i^{(g,m)}$$

projection heads to map embeddings to inter-layer space: identical/linear/non-linear mapping

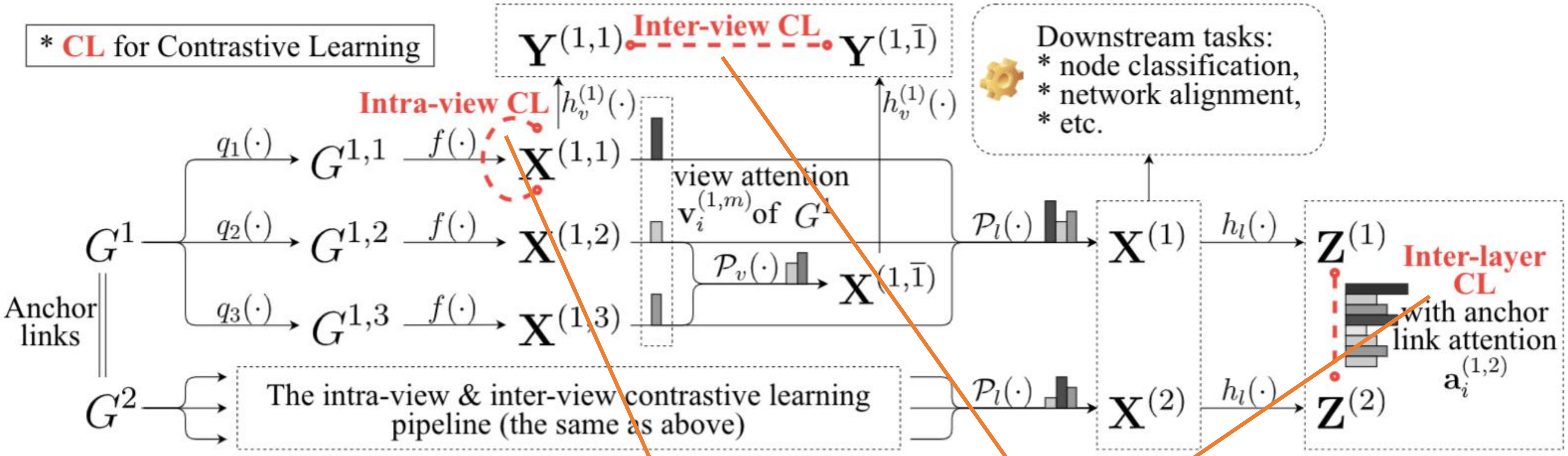
anchor link attention

Inter-layer Sampling

- Positive samples:
  - Step 1: sample one positive node pair from view $G^{1,1}$ (target) with the source node anchored
  - Step 2: map the source node to all the views of the source layer; map the target node to all the views of the target layer.
  - *M* target nodes and *M* source nodes
- Negative samples:
  - Step 1: sample on $G^{1,1}$
  - Step 2: map the negative nodes to all the views of the target layer
  - *B*M* negative nodes



Jointly Learning



$$\mathcal{L} = \sum_{g=1}^{N}\sum_{m=1}^{M}\Big(\underbrace{I_{AU}(\mathbf{X}^{(g,m)};\mathbf{X}^{(g,m)})}_{\text{intra-view MI}} + w_{iv}\,\underbrace{I_{AU}(\mathbf{Y}^{(g,m)};\mathbf{Y}^{(g,\overline{m})})}_{\text{inter-view MI}}\Big)$$

$$+\, w_{il}\sum_{s=1}^{N}\sum_{t=1}^{N}\mathbb{I}_{s\neq t}\,\underbrace{\widetilde{I}_{AU}(\mathbf{Z}^{(t)};\mathbf{Z}^{(s)})}_{\text{inter-layer MI}},$$

Outline

- Backgrounds
    - Network Embedding and multiplex networks
- Motivations
    - Two challenges in multiplex network embedding
- The Framework: **cM²NE**
- Experiments
- Conclusion

Experiments

- Network alignment
  - Predict unseen anchor links
  - Datasets:
    - Facebook-Twitter
    - Douban-Weibo
    - SacchCere

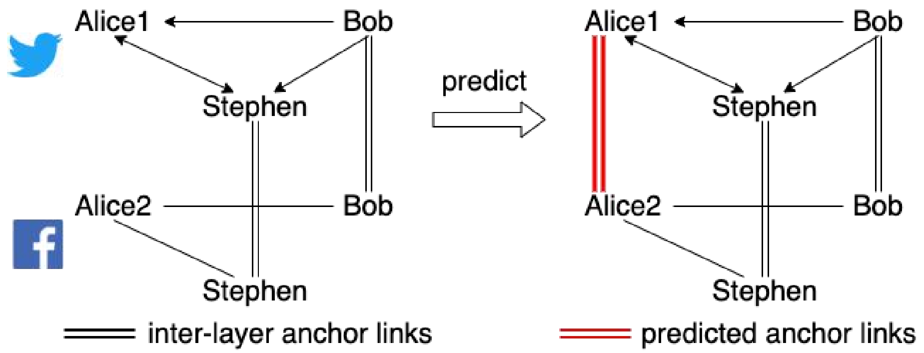


Fig. A toy example of network alignment

Table 4: Network statistics.

Dataset	PPI	BlogCatalog (simulated)	Facebook-Twitter	Douban-Weibo	SacchCere
Domain	Biological	Social	Social	Social	Biological
Task	Node classification		Network alignment		
$\lvert\mathcal{G}\rvert$ (N)	1	3	2	2	7
$\lvert\mathcal{V}^g\rvert$	3890	[10312, 10312, 10312]	[2458, 2458]	[3154, 3154]	[5928, 4850, 5042, 4694, 1401, 1130, 4949]
$\lvert\mathcal{E}^g\rvert$	76,584	[380078, 380304, 380120]	[40298, 95034]	[301074, 241736]	[66150, 37241, 29599, 37106, 2188, 1426, 109045]
Edge	Undirected	Directed	Directed	Directed	Directed
#Labels	50	39	/	/	/
Avg. $\lvert\mathcal{T}^{s,t}\rvert$	/	10312	2458	3154	[3499.8, 3247.2, 3223.8, 3194.8, 1179.3, 986.7, 3278.7]

Experiments

- Network alignment
  - Predict unseen anchor links
  - Datasets:
    - Facebook-Twitter
    - Douban-Weibo
    - SacchCere

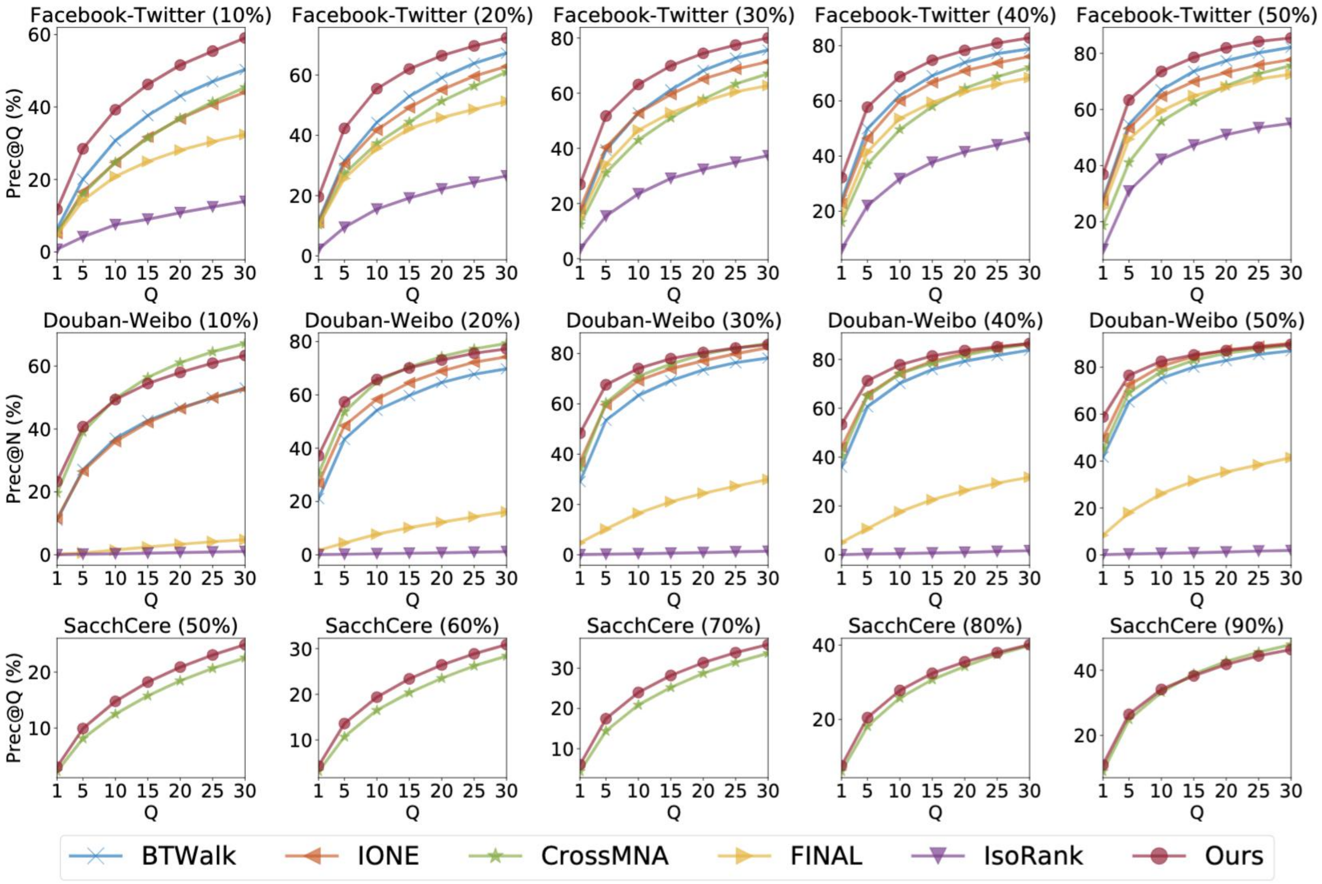


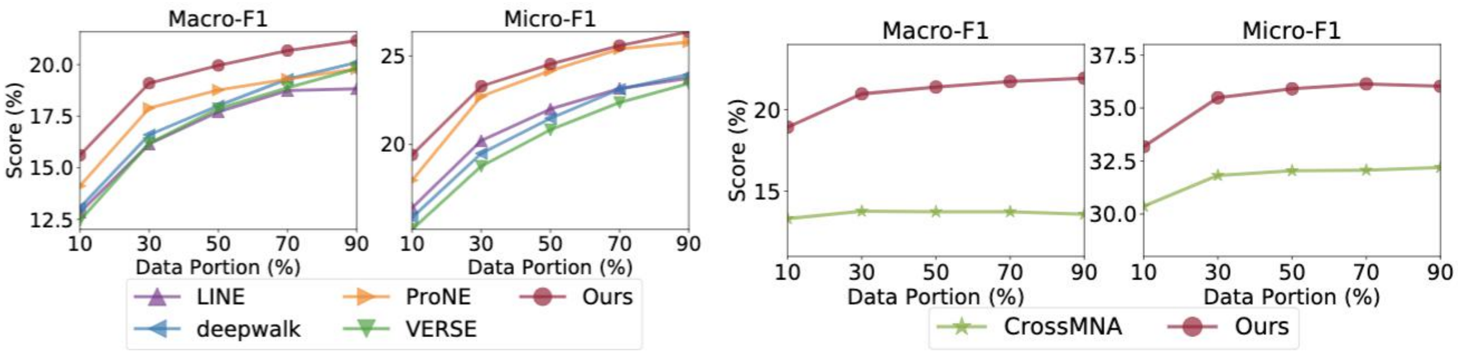
Figure 5: Network alignment results.

Experiments

- Node classification
  - Predict node labels
  - Datasets:
    - BlogCatalog
    - PPI

Table 4: Network statistics.

Dataset	PPI	BlogCatalog (simulated)	Facebook-Twitter	Douban-Weibo	SacchCere
Domain	Biological	Social	Social	Social	Biological
Task	Node classification		Network alignment		
$\|\mathcal{G}\|$ (N)	1	3	2	2	7
$\|\mathcal{V}^g\|$	3890	[10312, 10312, 10312]	[2458, 2458]	[3154, 3154]	[5928, 4850, 5042, 4694, 1401, 1130, 4949]
$\|\mathcal{E}^g\|$	76,584	[380078, 380304, 380120]	[40298, 95034]	[301074, 241736]	[66150, 37241, 29599, 37106, 2188, 1426, 109045]
Edge	Undirected	Directed	Directed	Directed	Directed
#Labels	50	39	/	/	/
Avg. $\|\mathcal{T}^{s,t}\|$	/	10312	2458	3154	[3499.8, 3247.2, 3223.8, 3194.8, 1179.3, 986.7, 3278.7]



(a) (single-layer) PPI (b) (multi-layer) BlogCatalog

Figure 4: Node classification results.

Experiments

- Abaltion study

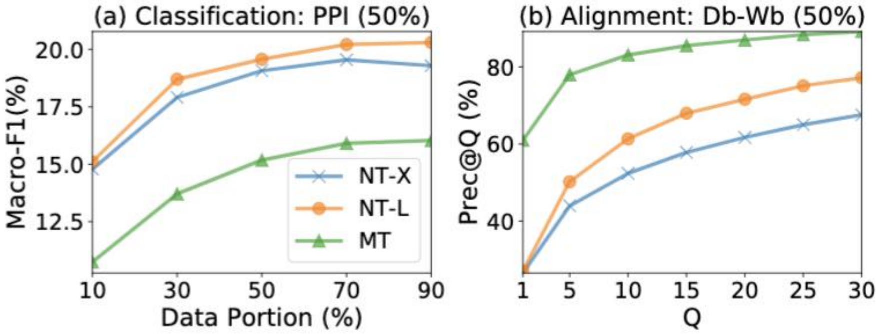


Figure 6: Ablation study of MI estimators. 'NT-X' is short for 'NT-Xent', 'NT-L': 'NT-Logistic', 'MT': 'Marginal Triplets'.

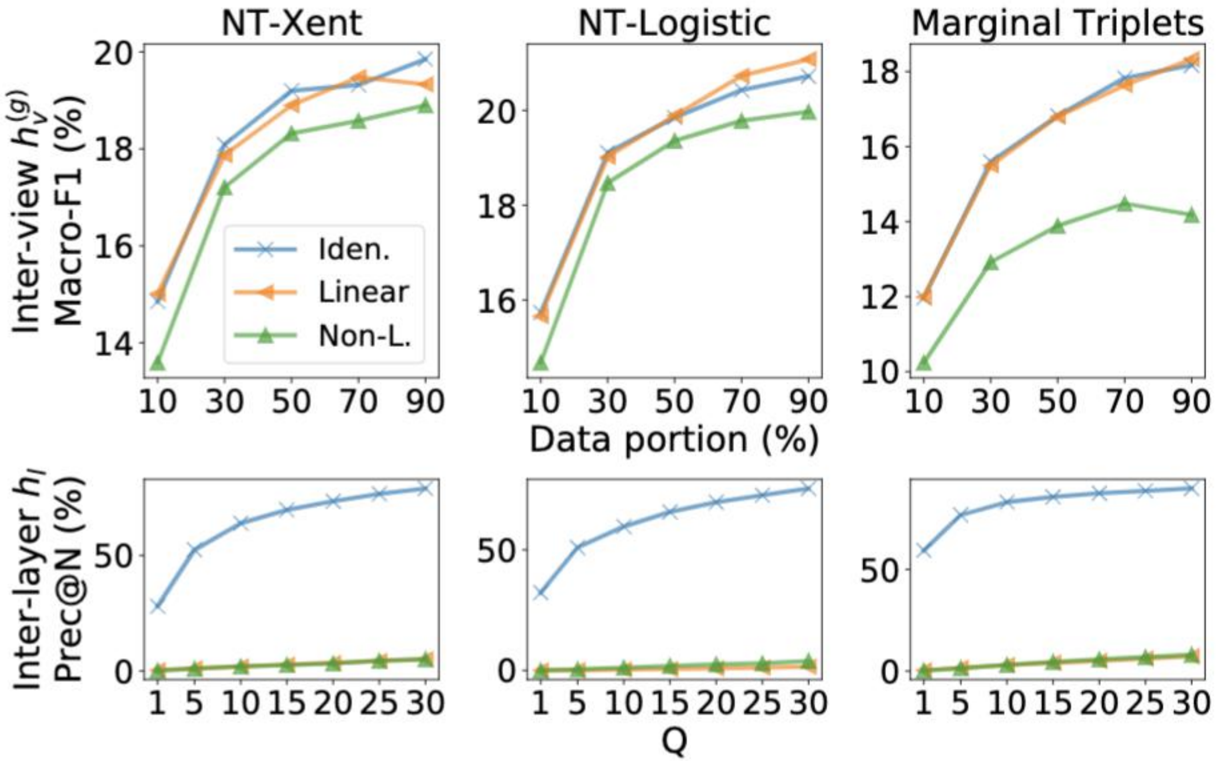


Figure 7: Ablation study for readout functions.

Experiments

- Case study
  - Different layers show different preferences on structural views, while low-order information are consistently prefered.
  - Anchor link attention is usually positively related with the Jaccard Similarity of the neighborhoods of the two anchored nodes.

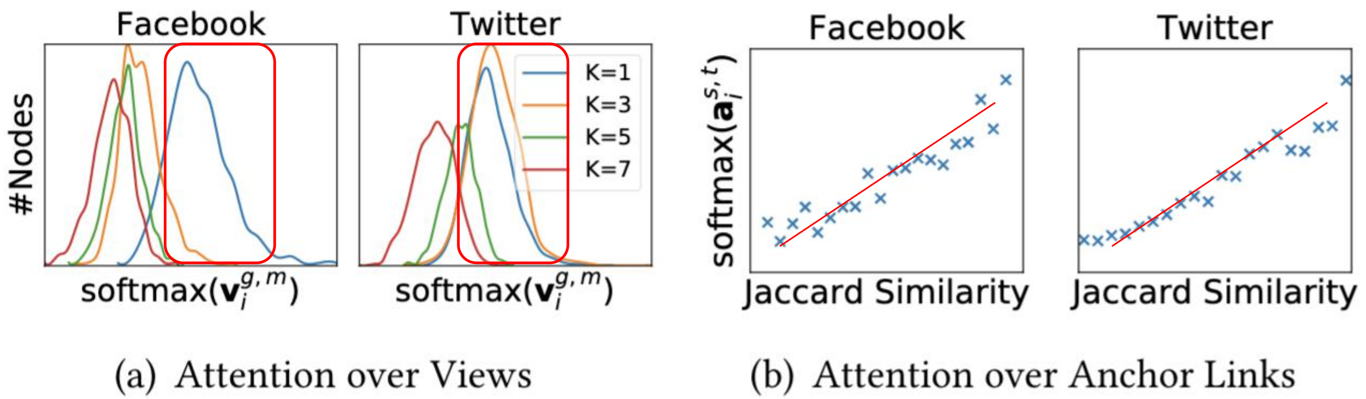


Figure 8: Case study for Facebook-Twitter. (a) distribution of attention over views. (b) positive correlation between the learned attention over anchor links and Jaccard Similarities.

Conclusion

- Summary of contributions:
- i) It is the first work to explore multiple structural views for multiplex network embedding.
- ii) Our multi-view contrastive learning framework is modulated by a tensorized attention mechanism, which adaptively learns the importance of each view and the agreement level of each anchor link.
- iii) The framework is equipped with several plug-in components that are new to the literature on MNE, including projection heads, embedding readout functions, and mutual information estimators.