HDMI:
High-order Deep Multiplex Infomax

Presenter: Baoyu Jing
Contact: baoyuj2@illinois.edu

Baoyu Jing
(UIUC)

Chanyoung Park
(KAIST)

Hanghang Tong
(UIUC)

Outline

- Introduction
- Preliminary
- Methodology
- Experiments
- Conclusion

Outline

- **Introduction**
- Preliminary
- Methodology
- Experiments
- Conclusion

Introduction: Ubiquity of Network

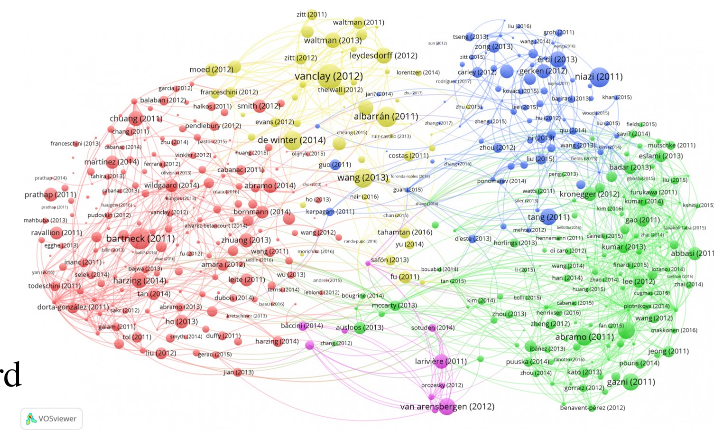
- Network in various applications.
  - Nodes are connected by various relations.

- Friendship
- Colleague
- Classmates

Social Network

- Also View
- Also Buy
- Bought Together

Product Network



- Citation
- Same Author
- Same Keyword

Paper Network

https://www.lebow.drexel.edu/news/relational-networking-not-just-collecting-contacts
https://atechnologyjobisnoexcuse.com/2006/09/11/network-of-amazon-products/
https://www.cwts.nl/blog?article=n-r2r294

Introduction: Self-supervised Learning

- Self-supervised learning
  - Train models without external training signals.
    - Do not need human labeling.
  - Pre-trained models perform well for down-stream tasks.
    - E.g., classification and clustering etc.
  - Key challenge: how to build the training signal?
- Deep Graph Infomax (DGI) for graphs
  - Mutual Information (MI) based training signal
  - Key idea:
    - Maximize the MI between node embedding h_n and summary vector s.

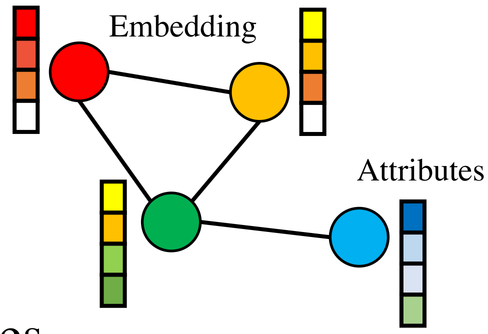
Velickovic, Petar, et al. "Deep Graph Infomax." *ICLR* . 2019.

Introduction: DGI Limitation #1

1. It only considers the **extrinsic** (global) information.
   - **Intrinsic** (node attribute) information is also important.

   × Existing methods:
   - Use reconstruction error
   - Reconstruction error doesn't imply high quality!

   ✓ In our work:
   - Maximize MI between node embedding and attributes
   - We propose to use **High-order Mutual Information** to jointly capture both extrinsic and intrinsic signals.
   - We propose a novel **High-order Deep Infomax (HDI)** as the training signal

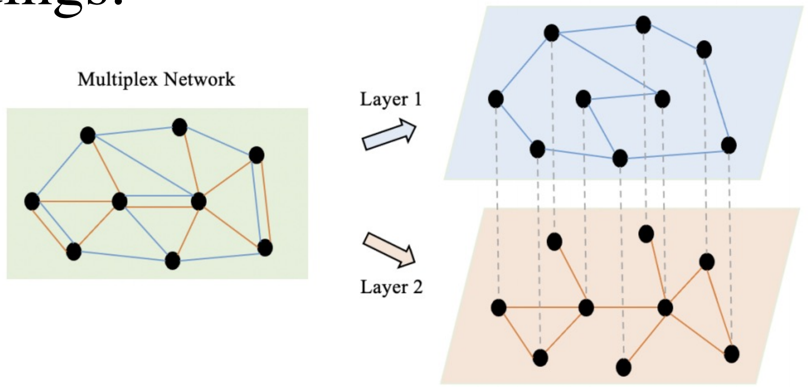


Introduction: DGI Limitation #2

2. DGI assumes a single type of relations among nodes.
   - Nodes are connected by **multiple** relations (**Multiplex Graph**).
     - Each relation is a layer of the graph.
     - Common strategy:
       1) Separately consider each layer.
       2) Combine embedding from different layers.

× Simplest way to combine embeddings:
- Average pooling

✓ In our work:
- Attention based fusion module.

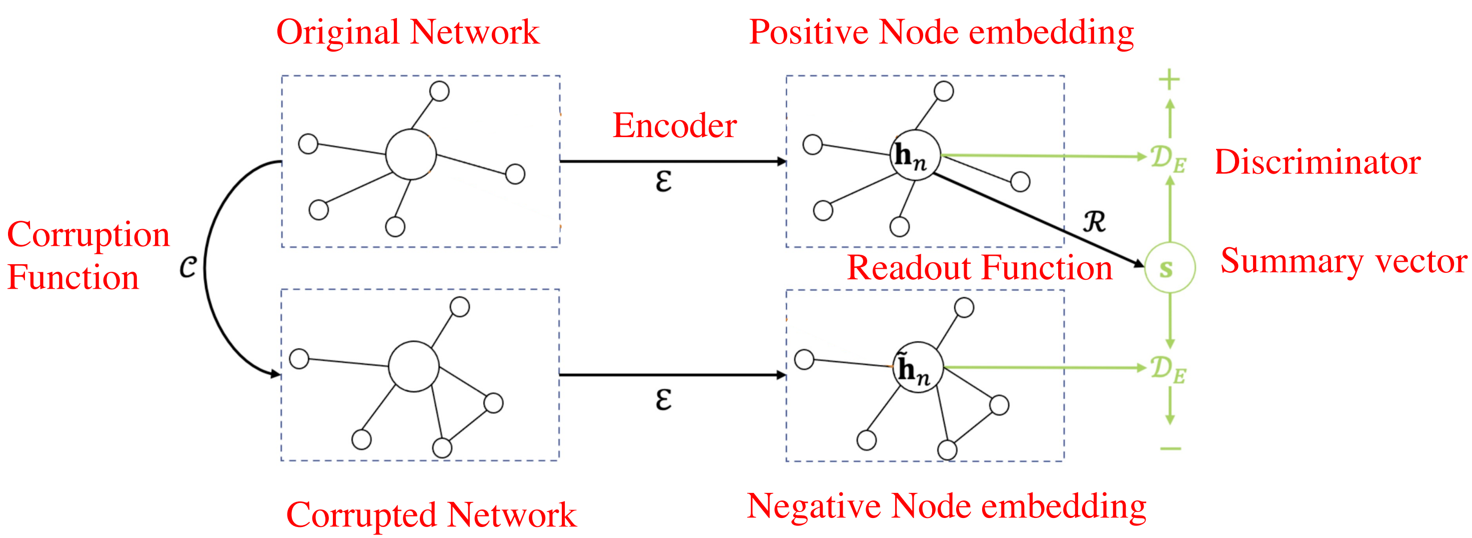


Outline

- Introduction
- ➢ **Preliminary**
- Methodology
- Experiments
- Conclusion

Preliminary: DGI

- Key steps:
  1) Generate a corrupted network via corruption function $\mathcal{C}$
  2) Use the encoder $\mathcal{E}$ to obtain node embeddings $\mathrm{h_n}$ and $\tilde{\mathrm{h}}_n$.
  3) Use the readout function $\mathcal{R}$ to obtain the summary vector s.
  4) Use the discriminator $\mathfrak{D}_E$ to discriminate $\mathrm{h_n}$ and $\tilde{\mathrm{h}}_n$.
  5) Maximize $I(\mathrm{h_n}; \mathrm{s})$ via: $\mathcal{L} = \sup_{\Theta} \mathbb{E}[\log \mathcal{D}(\mathbf{h}_n; \mathbf{s})] + \mathbb{E}[\log(1 - \mathcal{D}(\tilde{\mathbf{h}}_n; \mathbf{s}))]$



Outline

- Introduction
- Preliminary
- ➢ **Methodology**
- Experiments
- Conclusion

Methodology: High-order Deep Infomax

- High-order Deep Infomax (HDI):
  - Capture extrinsic and intrinsic signals via high-order mutual information $I(\mathrm{h_n}; \mathrm{s}; \mathrm{f_n})$.

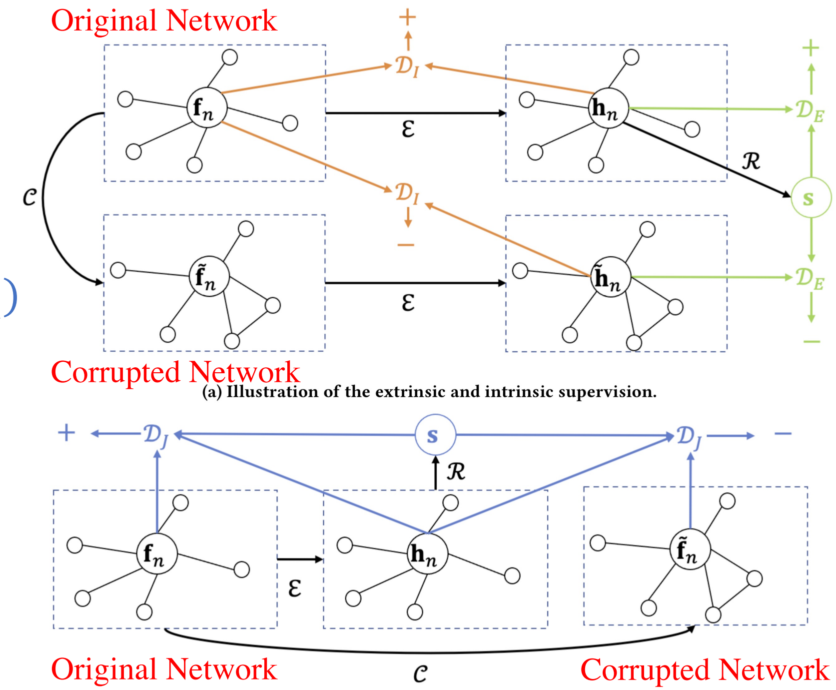
summary
embedding
attribute

- High-order mutual information:

$$I(\mathrm{h_n}; \mathrm{s}; \mathrm{f_n}) = \underbrace{I(\mathrm{h_n}; \mathrm{s})}_{\text{Extrinsic}} + \underbrace{I(\mathrm{h_n}; \mathrm{f_n})}_{\text{Intrinsic}} - \underbrace{I(\mathrm{h_n}; \mathrm{s}, \mathrm{f_n})}_{\text{Joint}}$$

- If directly maximize $I(\mathrm{h_n}; \mathrm{s}; \mathrm{f_n})$
  - Must max $I(\mathrm{h_n}; \mathrm{s}) + I(\mathrm{h_n}; \mathrm{f_n})$ and min $I(\mathrm{h_n}; \mathrm{s}, \mathrm{f_n})$
  - Maximizing $I(\mathrm{h_n}; \mathrm{s}, \mathrm{f_n})$ improves performance
- Jointly maximize three mutual information:
  - $\mathcal{L} = \lambda_E I(\mathrm{h_n}; \mathrm{s}) + \lambda_I I(\mathrm{h_n}; \mathrm{f_n}) + \lambda_J I(\mathrm{h_n}; \mathrm{s}, \mathrm{f_n})$
  - Final objective function will be:
  - $\mathcal{L} = \lambda_E \mathcal{L}_E + \lambda_I \mathcal{L}_I + \lambda_J \mathcal{L}_J$

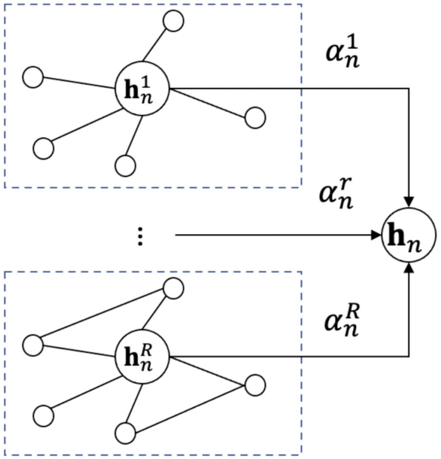
$\mathcal{L}_E$, $\mathcal{L}_I$ and $\mathcal{L}_J$ are the BCE losses of the discriminators.



(b) Illustration of the joint supervision.

Methodology: Fusion Module

- Extend HDI to multiplex graphs.
  - How to combine different layers?
- Fusion module is attention-based.
  - Different layers have different weights.
- Training the fusion module:
  - Apply HDI on top of the fused embedding.



- Full model: High-order Deep Multiplex Infomax (HDMI)
  - Objective: $\mathcal{L} = \lambda_M \mathcal{L}_M + \sum_r \lambda_r \mathcal{L}_r$

fusion module different layers

Outline

- Introduction
- Preliminary
- Methodology
- ➢ **Experiments**
- Conclusion

Experiments: Setup

- Datasets

Datasets	# Nodes	Relation Types	# Edges	# Attributes	# Labeled Data	# Classes
ACM	3,025	Paper-Subject-Paper (PSP) Paper-Author-Paper (PAP)	2,210,761 29,281	1,830 (Paper Abstract)	600	3
IMDB	3,550	Movie-Actor-Movie (MAM) Movie-Director-Movie (MDM)	66,428 13,788	1,007 (Movie plot)	300	3
DBLP	7,907	Paper-Author-Paper (PAP) Paper-Paper-Paper (PPP) Paper-Author-Term-Author-Paper (PATAP)	144,783 90,145 57,137,515	2,000 (Paper Abstract)	80	4
Amazon	7,621	Item-AlsoView-Item (IVI) Item-AlsoBought-Item (IBI) Item-BoughtTogether-Item (IOI)	266,237 1,104,257 16,305	2,000 (Item description)	80	4

- Questions
  - How will HDI and HDMI improve the quality?
  - Will fusion module assign appropriate attention scores?
- Tasks & Metrics
  - Node classification: Macro-F1 & Micro-F1
  - Node clustering: NMI, Similarity Search (Sim@5)
- Baselines
  - Network embedding: Deepwalk, DGI etc.
  - Multiplex network embeddings: HAN, DMGI etc.

Perozzi, Bryan, Rami Al-Rfou, and Steven Skiena. "Deepwalk: Online learning of social representations." *KDD*. 2014.
Velickovic, Petar, et al. "Deep Graph Infomax." *ICLR* . 2019.
Wang, Xiao, et al. "Heterogeneous graph attention network." *WWW*. 2019.
Park, Chanyoung, et al. "Unsupervised attributed multiplex network embedding." *AAAI*. 2020.

Experiments: Node Classification

- HDMI performs the best.
- HDI is better than baselines.

Dataset	ACM		IMDB		DBLP		Amazon	
Metric	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1
DeepWalk	0.739	0.748	0.532	0.550	0.533	0.537	0.663	0.671
node2vec	0.741	0.749	0.533	0.550	0.543	0.547	0.662	0.669
GCN/GAT	0.869	0.870	0.603	0.611	0.734	0.717	0.646	0.649
DGI	0.881	0.881	0.598	0.606	0.723	0.720	0.403	0.418
ANRL	0.819	0.820	0.573	0.576	0.770	0.699	0.692	0.690
CAN	0.590	0.636	0.577	0.588	0.702	0.694	0.498	0.499
DGCN	0.888	0.888	0.582	0.592	0.707	0.698	0.478	0.509
CMNA	0.782	0.788	0.549	0.566	0.566	0.561	0.657	0.665
MNE	0.792	0.797	0.552	0.574	0.566	0.562	0.556	0.567
mGCN	0.858	0.860	0.623	0.630	0.725	0.713	0.660	0.661
HAN	0.878	0.879	0.599	0.607	0.716	0.708	0.501	0.509
DMGI	0.898	0.898	0.648	0.648	0.771	0.766	0.746	0.748
$\text{DMGI}_{\text{attn}}$	0.887	0.887	0.602	0.606	0.778	0.770	0.758	0.758
HDI	**0.901**	0.900	0.634	0.638	0.814	0.800	0.804	0.806
HDMI	**0.901**	**0.901**	**0.650**	**0.658**	**0.820**	**0.811**	**0.808**	**0.812**

Experiments: Node Clustering

- HDMI performs the best.
- HDI is better than most of the baselines.

Dataset	ACM		IMDB		DBLP		Amazon	
Metric	NMI	Sim@5	NMI	Sim@5	NMI	Sim@5	NMI	Sim@5
DeepWalk	0.310	0.710	0.117	0.490	0.348	0.629	0.083	0.726
node2vec	0.309	0.710	0.123	0.487	0.382	0.629	0.074	0.738
GCN/GAT	0.671	0.867	0.176	0.565	0.465	0.724	0.287	0.624
DGI	0.640	0.889	0.182	0.578	0.551	0.786	0.007	0.558
ANRL	0.515	0.814	0.163	0.527	0.332	0.720	0.166	0.763
CAN	0.504	0.836	0.074	0.544	0.323	0.792	0.001	0.537
DGCN	0.691	0.690	0.143	0.179	0.462	0.491	0.143	0.194
CMNA	0.498	0.363	0.152	0.069	0.420	0.511	0.070	0.435
MNE	0.545	0.791	0.013	0.482	0.136	0.711	0.001	0.395
mGCN	0.668	0.873	0.183	0.550	0.468	0.726	0.301	0.630
HAN	0.658	0.872	0.164	0.561	0.472	0.779	0.029	0.495
DMGI	0.687	0.898	0.196	0.605	0.409	0.766	0.425	0.816
$\text{DMGI}_{\text{attn}}$	**0.702**	**0.901**	0.185	0.586	0.554	0.798	0.412	0.825
HDI	0.650	0.900	0.194	0.605	0.570	0.799	0.487	0.856
HDMI	0.695	0.898	**0.198**	**0.607**	**0.582**	**0.809**	**0.500**	**0.857**

Experiments: Ablation Study

1. Intrinsic (I.) & Joint (J.) MI significantly improve over Extrinsic (E) MI.
2. Fusion (HDMI) improves over simple average pooling.
3. Reconstruction Error (R.) does not imply high quality embedding!

Dataset	ACM				IMDB				DBLP						Amazon					
Layer	PSP		PAP		MDM		MAM		PAP		PPP		PATAP		IVI		IBI		IOI	
Metric	MaF1	MiF1	MaF1	MiF1	MaF1	MiF1	MaF1	MiF1	MaF1	MiF1	MaF1	MiF1	MaF1	MiF1	MaF1	MiF1	MaF1	MiF1	MaF1	MiF1
E.	0.663	0.668	0.855	0.853	0.573	0.586	0.558	0.564	0.804	0.796	0.728	0.717	0.240	0.272	0.380	0.388	0.386	0.410	0.569	0.574
E. + R.	0.668	0.673	0.864	0.847	0.590	0.597	0.560	0.570	0.809	0.801	0.737	0.728	0.240	0.280	0.392	0.398	0.410	0.427	0.579	0.589
E. + I.	0.719	0.732	0.886	0.887	0.617	0.624	0.593	0.600	0.803	0.792	0.742	0.733	0.240	0.276	0.559	0.561	0.517	0.527	0.792	**0.799**
E. + I. + J.	**0.742**	**0.744**	**0.889**	**0.888**	**0.626**	**0.631**	**0.600**	**0.606**	**0.812**	**0.803**	**0.751**	**0.745**	**0.241**	**0.284**	**0.581**	**0.583**	**0.524**	**0.529**	**0.796**	**0.799**

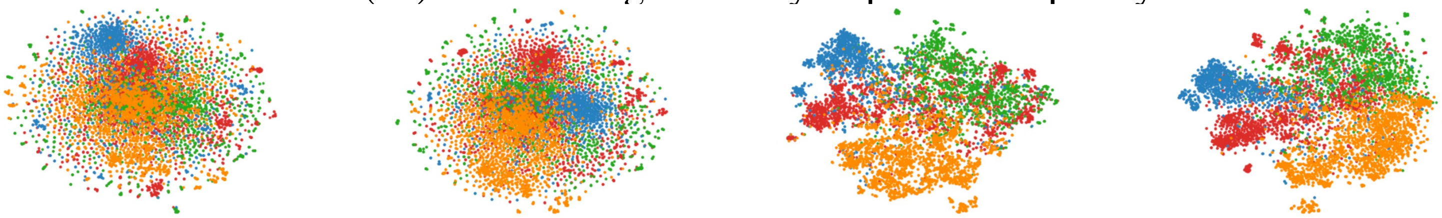
Metric	ACM MaF1	ACM MiF1	IMDB MaF1	IMDB MiF1	DBLP MaF1	DBLP FiF1	Amazon MaF1	Amazon MiF1
HDI	**0.901**	0.900	0.634	0.638	0.814	0.800	0.804	0.806
HDMI	**0.901**	**0.901**	**0.650**	**0.658**	**0.820**	**0.811**	**0.808**	**0.812**

Dataset	ACM				IMDB				DBLP						Amazon					
Layer	PSP		PAP		MDM		MAM		PAP		PPP		PATAP		IVI		IBI		IOI	
Metric	NMI	S@5	NMI	S@5	NMI	S@5	NMI	S@5	NMI	S@5	NMI	S@5	NMI	S@5	NMI	S@5	NMI	S@5	NMI	S@5
E.	0.526	0.698	0.651	0.872	0.145	0.549	0.089	0.495	0.547	0.800	0.404	0.741	0.054	0.583	0.002	0.395	0.003	0.414	0.038	0.701
E. + R.	0.525	**0.728**	0.659	0.874	0.150	0.552	0.079	0.490	0.564	0.804	**0.421**	0.741	0.051	0.568	0.002	0.399	0.003	0.426	0.020	0.660
E. + I.	0.527	0.708	0.656	0.882	0.193	**0.595**	**0.143**	**0.527**	**0.569**	0.802	0.405	0.741	0.053	0.569	0.152	0.512	0.143	0.517	0.401	0.824
E. + I. + J.	**0.528**	0.716	**0.662**	**0.886**	**0.194**	0.592	**0.143**	**0.527**	0.562	**0.805**	0.408	**0.742**	**0.054**	**0.591**	**0.169**	**0.544**	**0.153**	**0.525**	**0.407**	**0.826**

Metric	ACM NMI	ACM Sim@5	IMDB NMI	IMDB Sim@5	DBLP NMI	DBLP Sim@5	Amazon NMI	Amazon Sim@5
HDI	0.650	**0.900**	0.194	0.605	0.570	0.799	0.487	0.856
HDMI	**0.695**	0.898	**0.198**	**0.607**	**0.582**	**0.809**	**0.500**	**0.857**

Experiments: T-SNE Visualization

- Comparison of different signals: IOI layer of the Amazon network
  - Intrinsic (I.) and Joint (J.) MI improve the quality.
  - Reconstruction error (R.) does not significantly improve the quality.



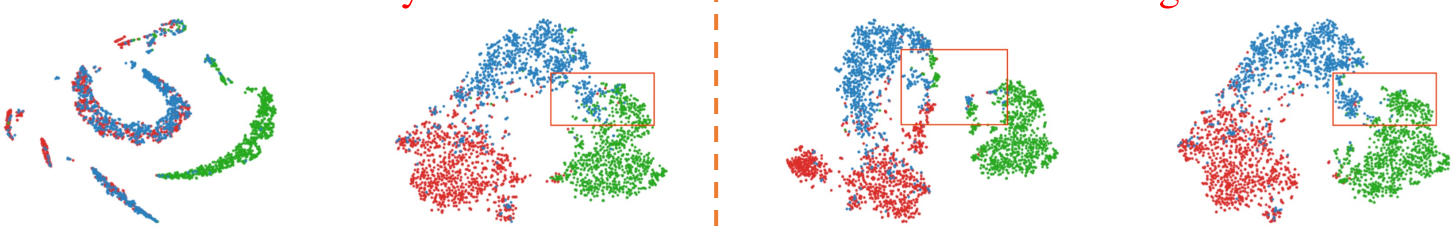
(a) E. (b) E. + R. (c) E. + I. (d) E. + I. + J.

Better

- Comparison of fusion mechanism: ACM network
  - Proposed fusion module is better than average pooling.

Different Layers

Different Fusing Methods

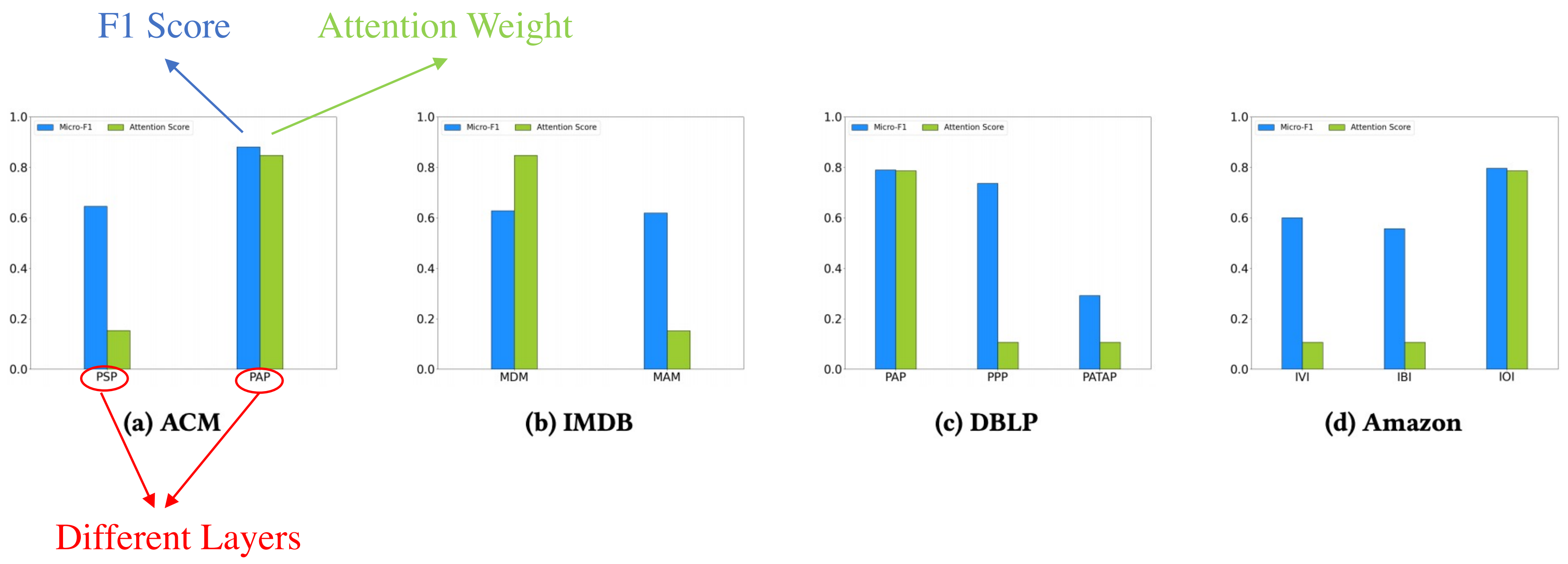


(a) PSP (b) PAP (c) Average (d) Fusion

Better

Experiments: Attention Scores

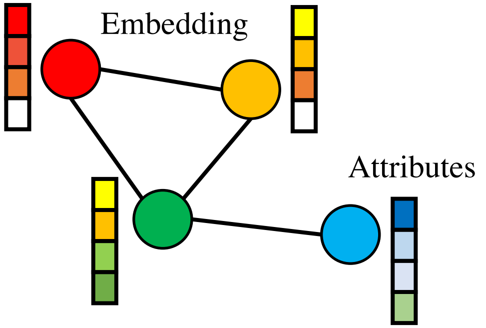
- Appropriate attention scores are assigned to different layers.
  - Higher F1 scores -> Higher attention scores.



Outline

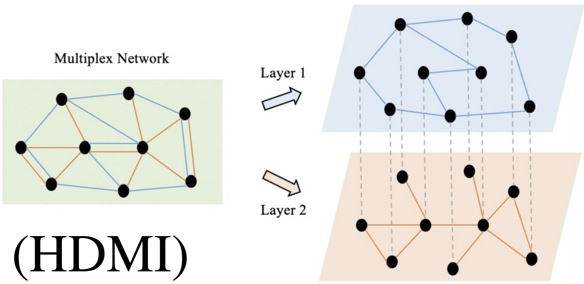
- Introduction
- Preliminary
- Methodology
- Experiments
- ➢ **Conclusion**

Conclusion

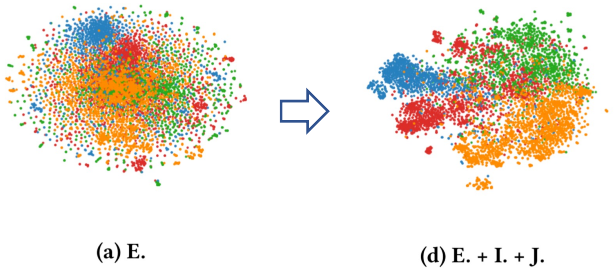


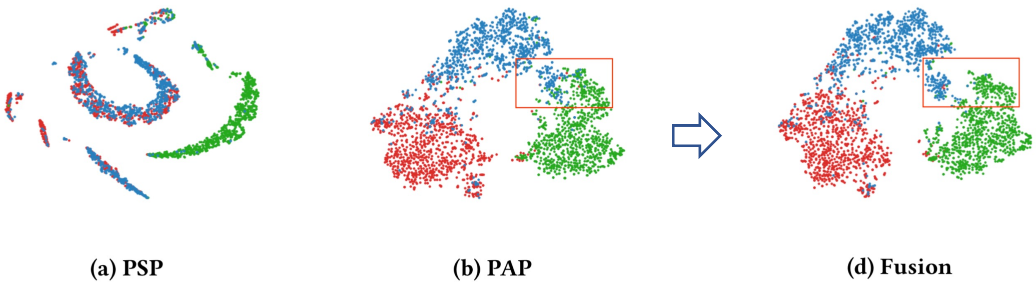
- MI based self-supervised learning for graphs
  - Challenge 1: Jointly capture extrinsic and intrinsic information for graphs.
  - ✓ Solution 1: High-order Deep Infomax (HDI)
  - Challenge 2: Extend HDI to multiplex graphs.
  - ✓ Solution 2: Attention based fusion

    -> High-order Deep Multiplex Infomax (HDMI)



- Results:
  - HDI significantly improves the quality of embeddings.
  - HDMI further improves HDI.





Thank you!