KDD 2020 Research Track Paper

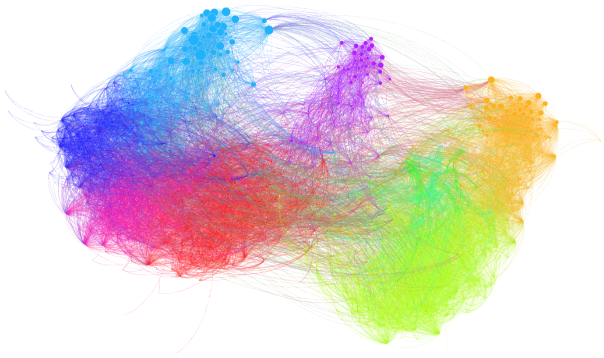
Unsupervised Differentiable Multi-aspect Network Embedding

Chanyoung Park, Carl Yang, Qi Zhu, Donghyun Kim, Hwanjo Yu, Jiawei Han

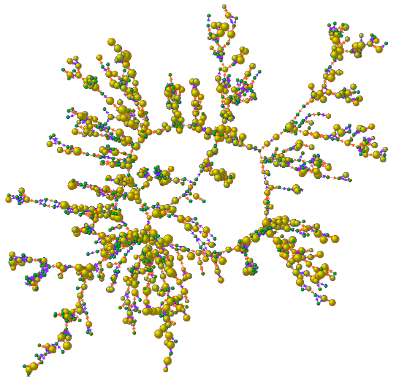
ILLINOIS EMORY UNIVERSITY YAHOO! RESEARCH POSTECH

Network is Everywhere

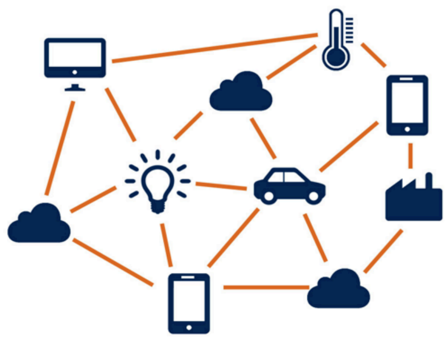
- A ubiquitous data structure to model the relationships between entities
- Many types of data can be flexibly formulated as networks



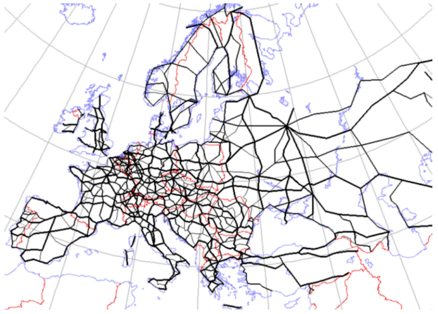
Social Network



Biological Network

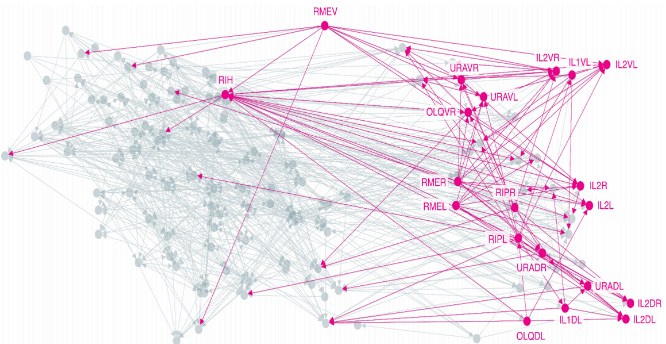


Internet-of-Things

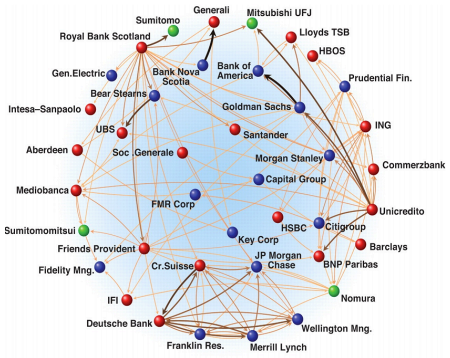


Road network

Chemical Network



Network of neurons



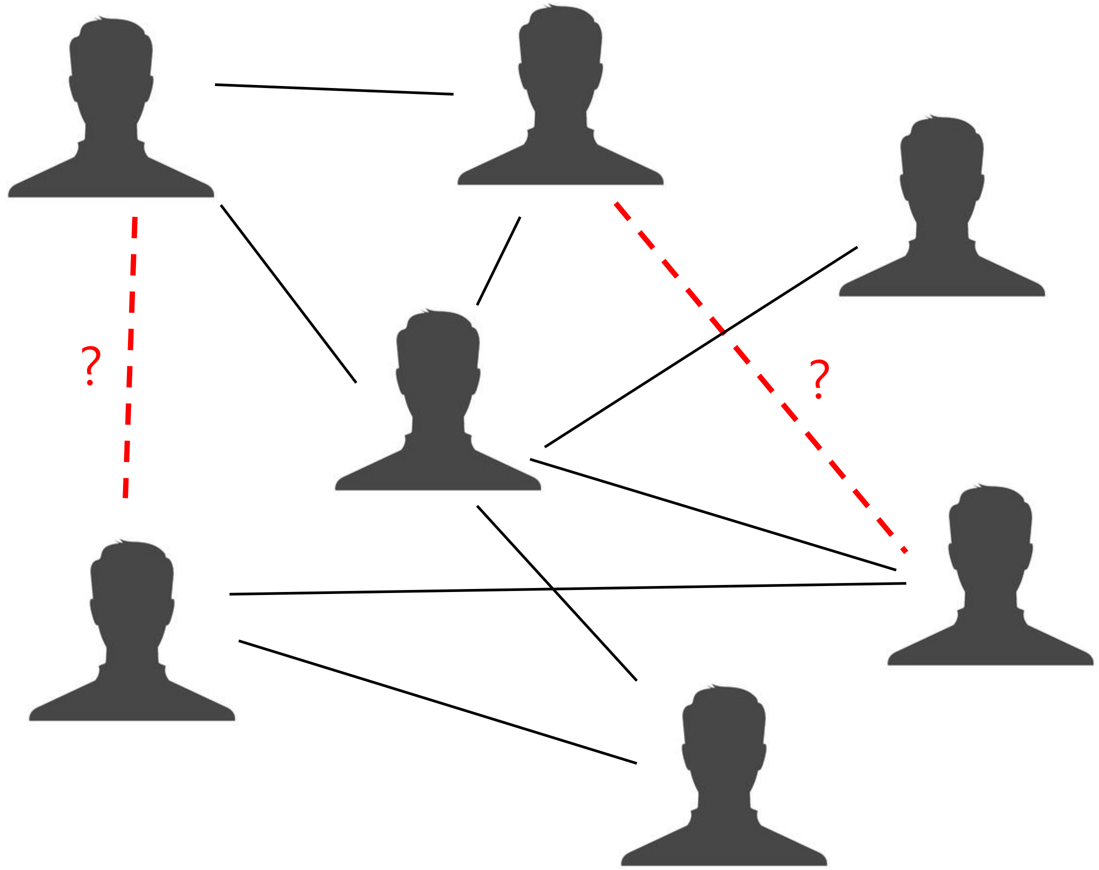
Financial network

Logistic network

Classical Tasks in Networks

- Node classification
  - Predict the type of a given node
- Link prediction
  - Predict whether two nodes are linked
- Community detection
  - Identify densely linked clusters of nodes
- Network similarity
  - How similar are two (sub)networks

Example: Link Prediction
(Friend Recommendation)

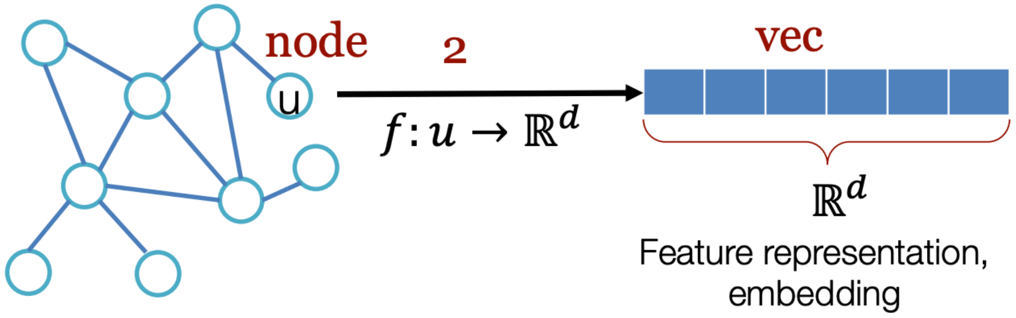


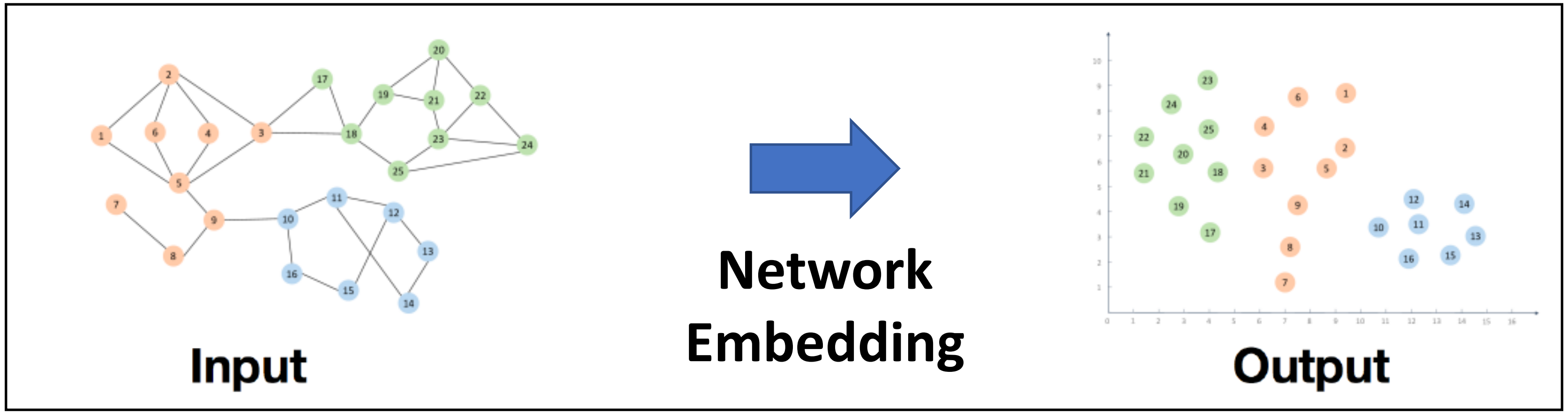
How do we solve these network-related tasks?

Network Embedding!

What is Network Embedding?

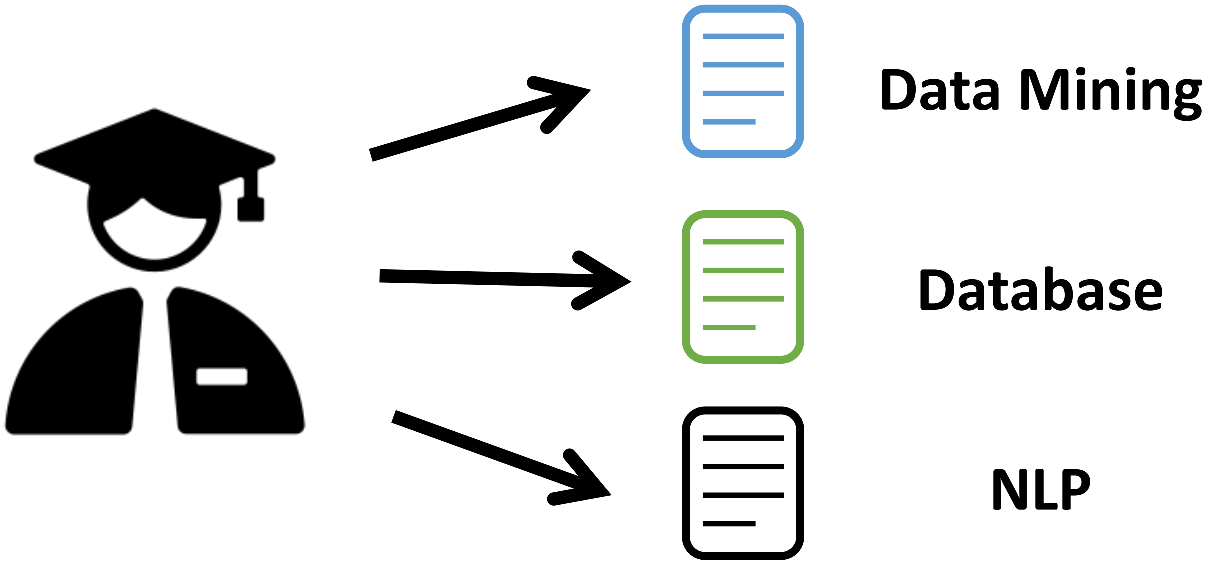
- Encode nodes so that **similarity in the embedding space** approximates **similarity in the original network**
- **Similar nodes in a network have similar vector representations**





Is a Single Vector Enough?

- Nodes (e.g., authors) in an academic publication network belong to multiple research communities
- Modeling each node with a single vector entails information loss



Multi-aspect of each node should be captured

Is Multi-aspect Enough?

- Authors can belong to multiple research communities
- **These communities interact with one another**

Database	$\xleftrightarrow{\text{More Related}}$	Data Mining	$\xleftrightarrow{\text{Less related}}$	Computer Architecture

Interactions among aspects should be captured

Research Question

1. **Is a Single Vector Enough?**
    - Solution: Multi-aspect Network Embedding

2. **Is Multi-aspect Enough?**
    - Solution: Aspect Regularization Framework

Research Question

1. **Is a Single Vector Enough?**
   - Solution: Multi-aspect Network Embedding

2. **Is Multi-aspect Enough?**
   - Solution: Aspect Regularization Framework

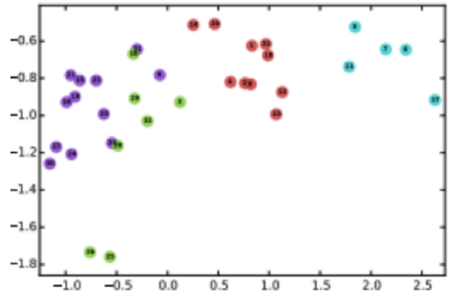
Previous work: Clustering-based aspect assignment

☹ 1. Each node always has the same **fixed aspect** regardless of its current context

☹ 2. Final network embedding **quality depends on the performance of clustering**

- Training **cannot be done end-to-end**

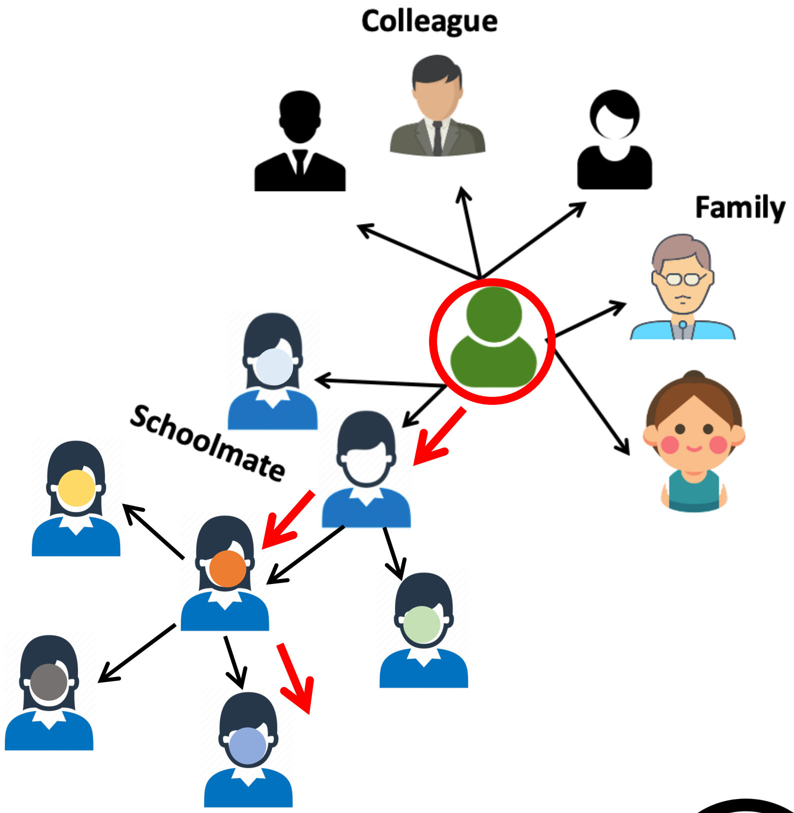
Start network embedding



Done!

[WWW19 Epasto et al]
[KDD19 Liu et al]

This work: Context-based aspect assignment



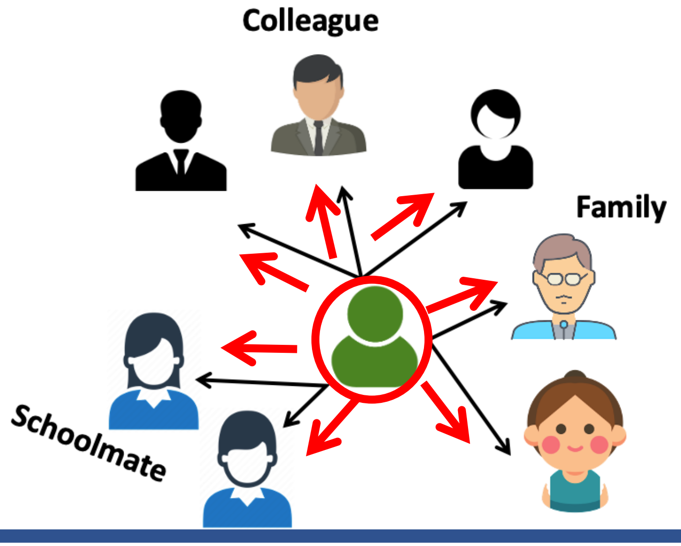
Considers
multi-hop neighbors

More effective for **capturing**
multi-aspect user behavior

Context: Schoolmate

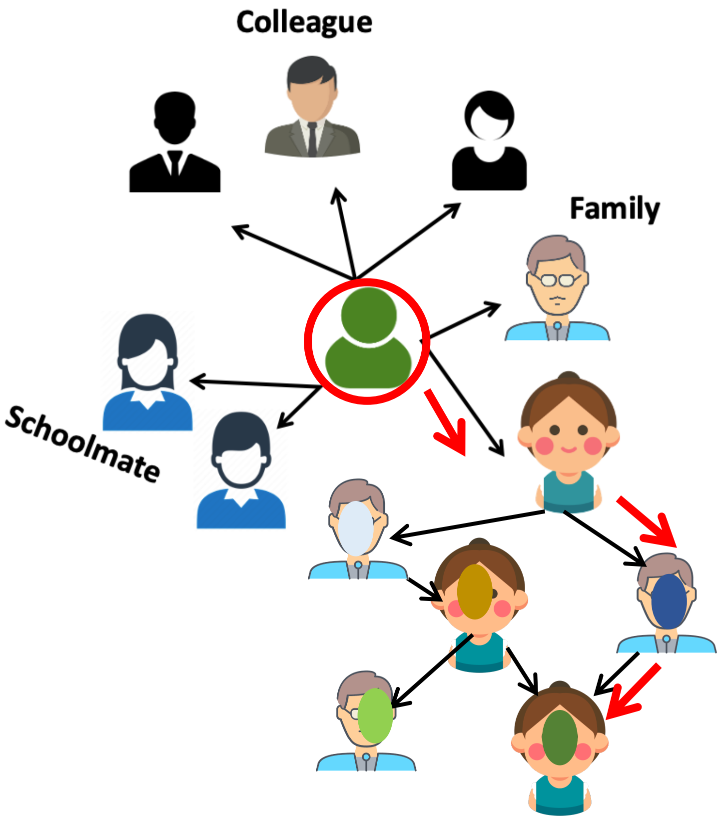
Assign "Schoolmate" aspect

Previous clustering-based method

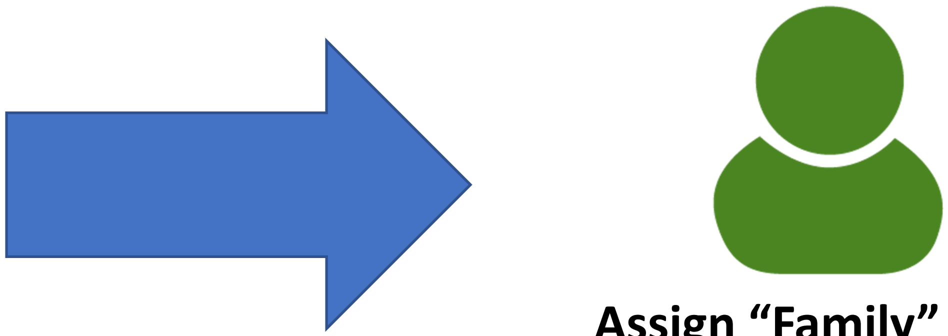


Only considers
one-hop neighbors

This work: Context-based aspect assignment



Context: Family



Assign "Family" aspect

Assign a single aspect for each node based on the context

This assignment process is non-differentiable

Gumbel-Softmax based Aspect Selection

- Adopt the **Gumbel-softmax trick** to **dynamically sample aspects based on the context**

Non-differentiable assignment → **Differentiable** assignment

☹ ☺

Gumbel-Softmax Trick (Jang et al, 2017)

- A simple way to draw a one-hot sample $\boldsymbol{z}$ from the **categorical distribution**
- **Given**: A K-dimensional **categorical distribution** with class probability $\pi_1, \pi_2, \dots, \pi_K$

Non-differentiable

Gumbel noise drawn from Gumbel(0,1)

$$g_i = -\log(-\log(u_i))$$
$$u_i \sim Uniform(0, 1)$$

☹ Gumbel-max

$$z = \text{one-hot}\left(\operatorname{argmax}_i \left[\log \pi_i + g_i\right]\right)$$

Differentiable

☺ Gumbel-softmax

$$z_i = \text{softmax}\left[\log \pi_i + g_i\right] = \frac{\exp\left(\left(\log \pi_i + g_i\right)/\tau\right)}{\sum_{j=1}^{K} \exp\left(\left(\log \pi_j + g_j\right)/\tau\right)} \quad \text{for } i = 1, \dots, K$$

Temperature parameter

As $\tau \to 0$, samples from the Gumbel-Softmax distribution become one-hot

Continuous relaxation of discrete random variable

(Jang et al, 2017) Categorical reparameterization with gumbel-softmax, ICLR 2017

Gumbel-Softmax based Aspect Selection

- Adopt the **Gumbel-softmax trick** to **dynamically sample aspects based on the context**

$$\mathrm{Readout}^{(s)}(\mathcal{N}(v_i)) = \frac{1}{|\mathcal{N}(v_i)|} \sum_{v_j \in \mathcal{N}(v_i)} \mathrm{Q}_j^{(s)} = \bar{\mathrm{Q}}_{\mathcal{N}(v_i)}^{(s)}$$

Gumbel-softmax

Embedding of v_i

Embedding of $N(v_i)$ regarding aspect s

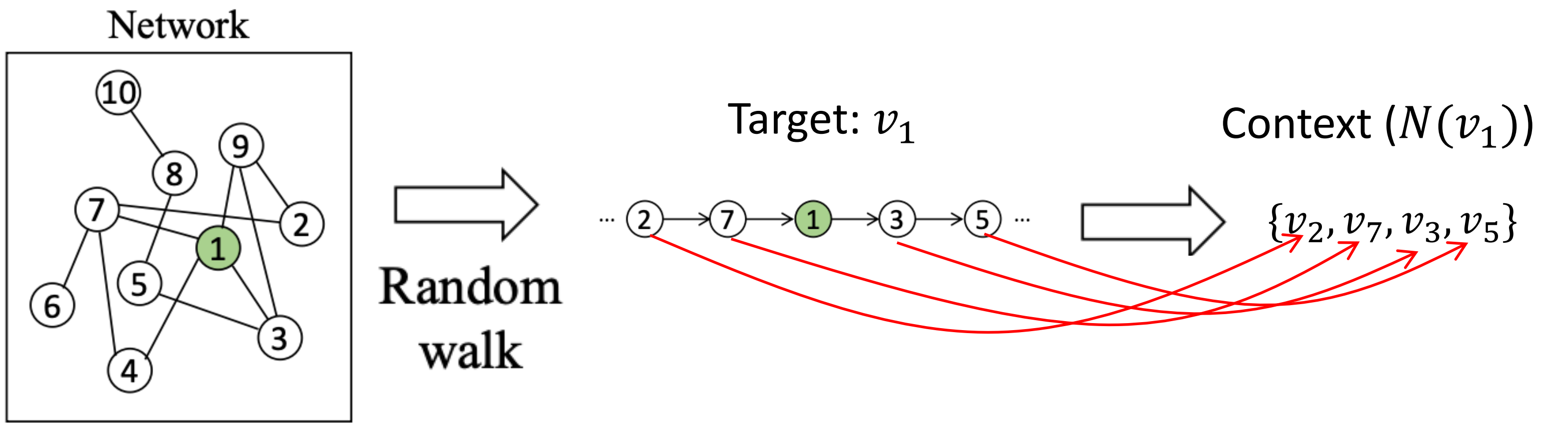
Aspect of node v_i

$$p(\delta(v_i) = s \,|\, \mathcal{N}(v_i)) = \boxed{\frac{\exp[(\log \langle \mathrm{P}_i, \mathrm{Readout}^{(s)}(\mathcal{N}(v_i))\rangle + g_s)/\tau]}{\sum_{s'=1}^{K} \exp[(\log \langle \mathrm{P}_i, \mathrm{Readout}^{(s')}(\mathcal{N}(v_i))\rangle + g_{s'})/\tau]}}$$

Sample the aspect that gives the highest value

Local context of v_i

Probability of v_i being selected as aspect s given its context $N(v_i)$



Single-aspect → Multi-aspect

Single-aspect $\mathcal{J}_{\mathrm{DW}}^{(\mathbf{w})} = \sum_{v_i \in \mathbf{w}} \sum_{v_j \in \mathcal{N}(v_i)} \log \boxed{p(v_j|v_i)}$

$$\frac{\exp(\langle \mathbf{P}_i, \mathbf{Q}_j \rangle)}{\sum_{v_{j'} \in \mathcal{V}} \exp(\langle \mathbf{P}_i, \mathbf{Q}_{j'} \rangle)}$$

$$\frac{\exp[(\log \langle \mathbf{P}_i, \mathrm{Readout}^{(s)}(\mathcal{N}(v_i)) \rangle + g_s)/\tau]}{\sum_{s'=1}^{K} \exp[(\log \langle \mathbf{P}_i, \mathrm{Readout}^{(s')}(\mathcal{N}(v_i)) \rangle + g_{s'})/\tau]}$$

Multi-aspect $\mathcal{J}_{\mathrm{asp2vec}}^{(\mathbf{w})} = \sum_{v_i \in \mathbf{w}} \sum_{v_j \in \mathcal{N}(v_i)} \sum_{s=1}^{K} \boxed{p(\delta(v_i) = s|\mathcal{N}(v_i))} \log \boxed{p(v_j|v_i, p(\delta(v_i) = s))}$

Aspect selection probability

$$\frac{\exp(\langle \mathbf{P}_i, \mathbf{Q}_j^{(s)} \rangle)}{\sum_{v_{j'} \in \mathcal{V}} \exp(\langle \mathbf{P}_i, \mathbf{Q}_{j'}^{(s)} \rangle)}$$

Final objective function $\mathcal{L}_{\mathrm{asp2vec}} = -\sum_{\mathbf{w} \in \mathcal{W}} \mathcal{J}_{\mathrm{asp2vec}}^{(\mathbf{w})}$

Research Question

1. ~~Is a Single Vector Enough?~~
    - ~~Solution: Multi-aspect Network Embedding~~

2. **Is Multi-aspect Enough?**
    - Solution: Aspect Regularization Framework

Aspect Regularization Framework

- **Interactions among aspects should be captured**
  - **More related:** Data Mining (DM) ↔ Database (DB)
  - **Less related:** Data Mining (DM) ↔ Computer Architecture (CA)

- Goal: Aspect embeddings should be
  1. **Related to each other (Relatedness)**
     - To capture some common information shared among aspects (e.g., DM ↔ DB)
  2. **Diverse from each other (Diversity)**
     - To independently capture the inherent properties of individual aspects (e.g., DM ↔ CA)

How to capture both relatedness and diversity among aspects?

Capturing Diversity

- Minimize similarity among aspect embeddings (= maximize diversity)

$$\text{reg}_{\text{asp}} = \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} \boxed{\text{A-Sim}(Q_*^{(i)}, Q_*^{(j)})}$$

Aspect similarity between aspect i and j

$$Q_*^{(i)} \in \mathbb{R}^{n \times d}$$

Aspect embedding matrix w.r.t. aspect i

$$\text{A-Sim}(Q_*^{(i)}, Q_*^{(j)}) = \sum_{h=1}^{|V|} f(Q_h^{(i)}, Q_h^{(j)})$$

$$f(Q_h^{(i)}, Q_h^{(j)}) = \frac{\langle Q_h^{(i)}, Q_h^{(j)} \rangle}{\|Q_h^{(i)}\| \|Q_h^{(j)}\|}, \quad -1 \leq f(Q_h^{(i)}, Q_h^{(j)}) \leq 1$$

Cosine similarity

What about relatedness?

Capturing Relatedness

- Allow similarity among aspects **to some extent**

Binary mask

$$\text{A-Sim}(Q_*^{(i)}, Q_*^{(j)}) = \sum_{h=1}^{|V|} f(Q_h^{(i)}, Q_h^{(j)}) \quad \Rightarrow \quad \text{A-Sim}(Q_*^{(i)}, Q_*^{(j)}) = \sum_{h=1}^{|V|} \boxed{w_{i,j}^h} f(Q_h^{(i)}, Q_h^{(j)})$$

Maximize diversity

Maximize diversity + allow some similarity

$$w_{i,j}^h = \begin{cases} 1, & \left| f(Q_h^{(i)}, Q_h^{(j)}) \right| \geq \epsilon \\ 0, & \text{otherwise} \end{cases}$$

- Enforce loss if similarity is larger than ϵ
  - Allow similarity as much as ϵ

Final Objective Function

$$\mathcal{L} = \underbrace{\mathcal{L}_{\text{asp2vec}}}_{\text{Multi-aspect embedding}} + \lambda \underbrace{\text{reg}_{\text{asp}}}_{\text{Aspect regularization}}$$

Question 1 → Multi-aspect embedding

Question 2 → Aspect regularization

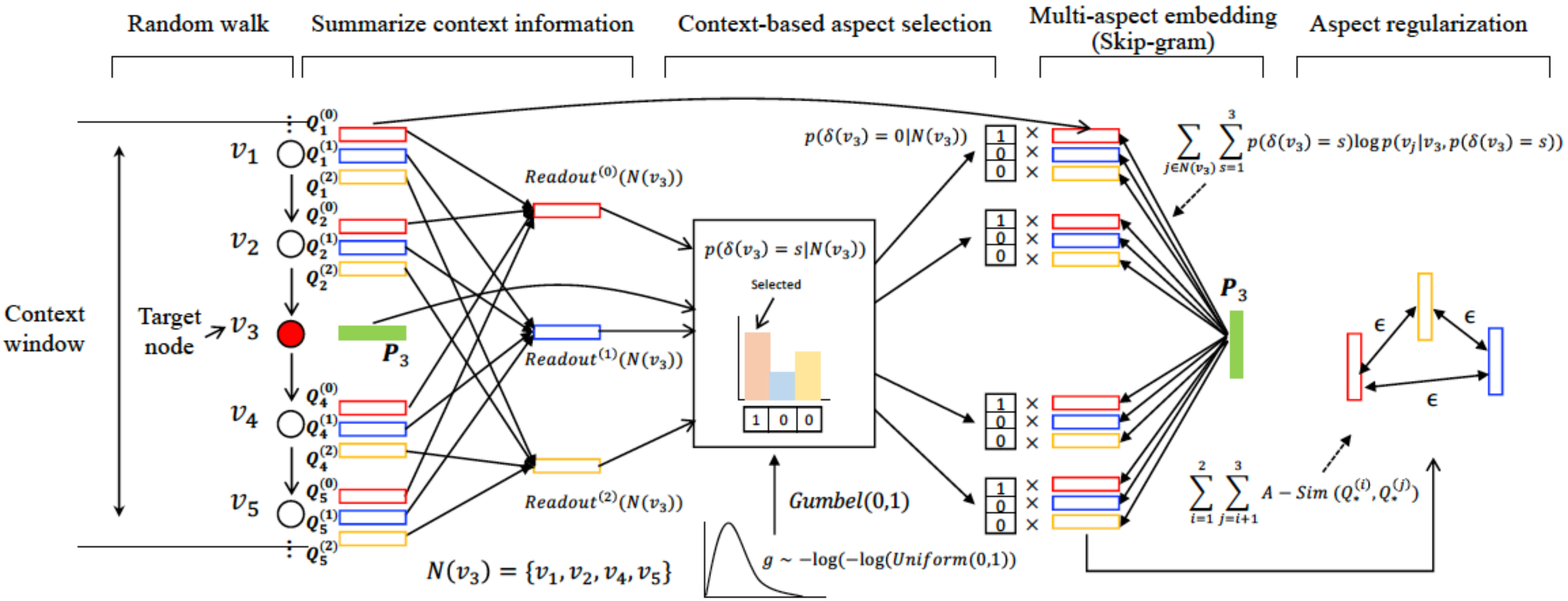
$$\mathcal{L}_{\text{asp2vec}} = -\sum_{\mathbf{w}\in\mathcal{W}} \mathcal{J}^{(\mathbf{w})}_{\text{asp2vec}}$$

$$\mathcal{J}^{(\mathbf{w})}_{\text{asp2vec}} = \sum_{v_i\in\mathbf{w}} \sum_{v_j\in\mathcal{N}(v_i)} \sum_{s=1}^{K} p(\delta(v_i)=s|\mathcal{N}(v_i)) \log p(v_j|v_i, p(\delta(v_i)=s))$$

$$\text{reg}_{\text{asp}} = \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} \text{A-Sim}(\mathbf{Q}_*^{(i)}, \mathbf{Q}_*^{(j)})$$

$$\text{A-Sim}(\mathbf{Q}_*^{(i)}, \mathbf{Q}_*^{(j)}) = \sum_{h=1}^{|V|} w_{i,j}^{h} f(\mathbf{Q}_h^{(i)}, \mathbf{Q}_h^{(j)})$$

Overall Architecture: asp2vec



Experiments: Dataset

Table 2: Statistics of the datasets. (Dir.: directed graph.)

Dataset			Num. nodes	Num. edges
Homogeneous Network	Social Network	Filmtrust (Dir.)	1,642	1,853
		Wiki-vote (Dir.)	7,066	103,689
		CiaoDVD (Dir.)	7,375	111,781
		BlogCatalog	10,312	333,983
		Epinions (Dir.)	49,290	487,181
		Flickr	80,513	5,899,882
	PPI		3,890	76,584
	Wikipedia (Word co-occurrence)		4,777	184,812
	Academic Network	Cora	2,708	5,429
		ca-HepTh	9,877	25,998
		ca-AstroPh	18,772	198,110
		4area	27,199	66,832

Result: Link Prediction

Table 1: The overall performance for link prediction in terms of AUC-ROC (OOM: Out of memory).

dim ($d \times K$)	100 ($d = 20$, $K = 5$)					200 ($d = 40$, $K = 5$)					500 ($d = 100$, $K = 5$)				
	DW	DGI	PolyDW	Splitter	asp2vec	DW	DGI	PolyDW	Splitter	asp2vec	DW	DGI	PolyDW	Splitter	asp2vec
Filmtrust	0.6850	0.6973	0.6953	0.6128	**0.7426**	0.7399	0.7094	0.6841	0.6111	**0.7460**	0.7415	0.7215	0.6643	0.6097	**0.7501**
Wiki-vote	0.6273	0.5860	0.5557	0.5190	**0.6478**	0.6277	0.5741	0.5179	0.5085	**0.6464**	0.6260	**0.6540**	0.5161	0.5048	0.6507
CiaoDVD	0.7136	0.6809	0.6528	0.5978	**0.7430**	0.7014	0.6696	0.6263	0.5881	**0.7447**	0.7140	0.6897	0.6058	0.5819	**0.7450**
BlogCatalog	0.8734	0.9191	0.7505	0.8441	**0.9503**	0.9220	0.9083	0.6944	0.8199	**0.9548**	0.9331	OOM	0.6249	0.7876	**0.9429**
Epinions	0.7188	0.6684	0.7038	0.6880	**0.7416**	0.7223	0.6711	0.6884	0.6733	**0.7441**	0.7312	OOM	0.6720	0.6581	**0.7459**
Flickr	0.9506	0.9214	0.9146	0.9528	**0.9584**	0.9580	OOM	0.8862	0.8582	0.9571	0.9570	OOM	0.8582	0.9299	**0.9678**
PPI	0.8236	0.8087	0.7286	0.8372	**0.8887**	0.8237	0.8341	0.6995	0.8346	**0.8947**	0.8214	0.8593	0.6693	0.8336	**0.8991**
Wikipedia	0.7729	0.8984	0.6259	0.6897	**0.9049**	0.8677	0.8927	0.5920	0.6939	**0.9040**	0.8414	**0.9029**	0.5218	0.7018	0.9011
Cora	**0.9181**	0.8223	0.8504	0.8357	0.8814	**0.9110**	0.8300	0.8416	0.8361	0.9056	0.8814	**0.9475**	0.8393	0.8412	0.9181
ca-HepTh	**0.9080**	0.8661	0.8806	0.8827	0.8989	**0.9160**	0.8787	0.8812	0.9076	0.9119	**0.9219**	0.7402	0.8831	0.9058	0.9185
ca-AstroPh	**0.9784**	0.9144	0.9661	0.9731	0.9734	0.9803	0.9690	0.9734	0.9791	**0.9821**	0.9775	OOM	0.9754	0.9827	**0.9842**
4area	**0.9548**	0.9253	0.9441	0.9355	0.9503	0.9551	0.9349	0.9449	0.9496	**0.9587**	0.9553	OOM	0.9463	0.9550	**0.9627**

- asp2vec generally performs well on all datasets
- Especially superior on social networks, PPI and Wikipedia networks
  - asp2vec performs better on networks that inherently exhibit multiple aspects

Result: Benefit of Gumbel-softmax based Aspect Selection

$d = 20, K = 5$	Softmax	Gumbel-Softmax	Improvement
Filmtrust	0.6421	**0.7426**	15.65%
Wiki-vote	0.6165	**0.6478**	5.08%
CiaoDVD	0.6162	**0.7430**	20.58%
BlogCatalog	0.7323	**0.9503**	29.77%
Epinions	0.6693	**0.7416**	10.80%
Flickr	0.8956	**0.9584**	7.01%
PPI	0.6919	**0.8887**	28.44%
Wikipedia	0.8269	**0.9049**	9.43%
Cora	0.8605	**0.8814**	2.43%
ca-HepTh	0.8890	**0.8989**	1.11%
ca-AstroPh	0.9116	**0.9734**	6.78%
4area	0.9286	**0.9503**	2.34%

Gumbel-Softmax is beneficial

Improvements: Social networks, PPI >> Academic networks

- Aspect modeling is more effective for networks with inherently diverse aspects
  - Aspect diversity: ex) User in social network vs. author in academic network

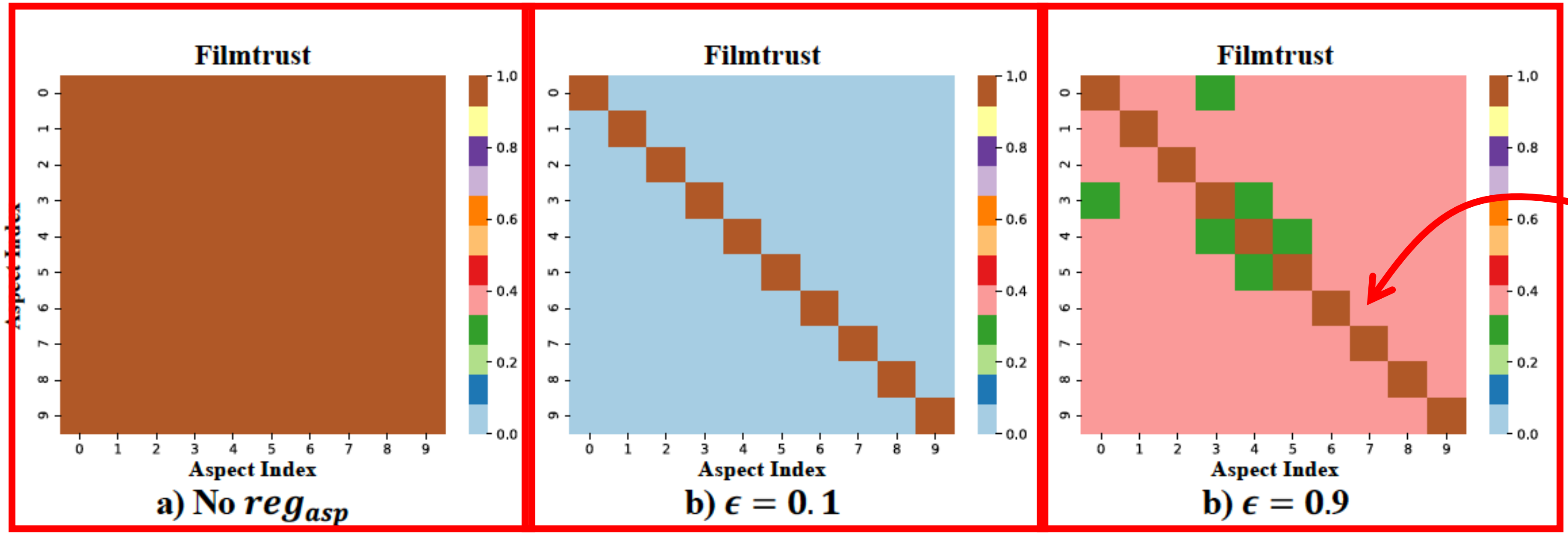
Result: Benefit of Aspect Regularization

Table 4: Link prediction performance (AUC-ROC) without reg_{asp}, and over various thresholds (ϵ).

dim = 100 (d = 20, K = 5)	No reg_{asp}	Threshold (ϵ) 0.9	0.7	0.5	0.3	0.1	best vs. No reg_{asp}
Filmtrust	0.660	**0.743**	0.742	0.740	0.738	0.735	12.58%
Wiki-vote	0.616	0.647	**0.648**	0.647	0.647	0.645	5.15%
CiaoDVD	0.617	**0.743**	0.742	0.742	0.738	0.735	20.37%
BlogCatalog	0.791	0.948	**0.950**	0.949	0.939	0.869	20.11%
Epinions	0.684	**0.742**	0.741	0.738	0.731	0.693	8.37%
Flickr	0.897	0.955	**0.958**	0.954	0.954	0.929	6.85%
PPI	0.729	0.880	0.885	**0.889**	0.881	0.819	21.97%
Wikipedia	0.841	0.896	0.904	**0.905**	0.880	0.850	7.60%
Cora	0.879	0.881	0.880	**0.881**	0.862	0.857	0.23%
ca-HepTh	0.879	**0.899**	0.896	0.898	0.893	0.864	2.30%
ca-AstroPh	0.921	**0.973**	0.973	0.971	0.967	0.939	5.56%
4area	0.919	**0.950**	0.949	0.946	0.940	0.915	3.44%

1) Performance drops significantly when the aspect regularization framework is not incorporated
2) Aspect regularization framework is less effective on the academic networks
   - Academic networks inherently have less diverse aspects

Result: How are the aspect embeddings learned?



Average pairwise similarity between aspect embedding vectors for all nodes

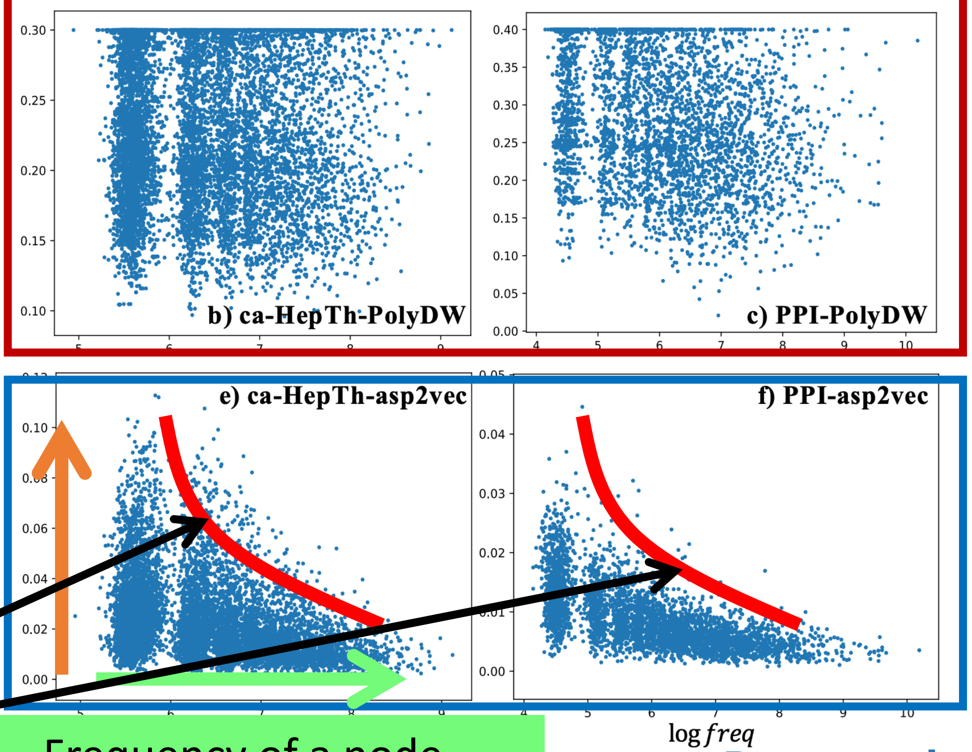
- Aspect embeddings are trained to be highly similar to each other without reg_{asp}
  - Verifies the necessity of aspect regularization
- Small ϵ encourages the aspect embeddings to be diverse
- Large ϵ allows more flexibility in learning the aspect embeddings

Result: How are aspects assigned?

How does the real data look like?

Frequently appearing node → Popular
→ Likely to have diverse aspects
→ Aspects are relatively evenly distributed
→ Variance of aspect distribution is small

Previous work
(offline clustering-based aspect selection)



Variance of aspect distribution

Frequency of a node appearing in random walks (Node popularity)

Proposed work
(Gumbel-softmax based aspect selection)

The results reflect the real-world data

Conclusion

- Proposed a novel multi-aspect network embedding method
  - Dynamically determines the aspect based on the context information
- **Aspect selection module** (based on Gumbel-softmax trick)
  - Approximate the discrete sampling of the aspects
  - End-to-end training
- **Aspect regularization framework**
  - Encourage the learned aspect embeddings to be diverse, but to some extent related to each other
- Also easily extended to heterogeneous network (See paper)

Thank You!

For more information, please check our paper and code!

- Paper: https://arxiv.org/abs/2006.04239
- Code & Datasets: https://github.com/pcy1302/asp2vec
- Contact: cy.park424@gmail.com