Heterogeneous Graph Attention Network

Xiao Wang, Houye Ji, Chuan Shi*, Bai Wang, Peng Cui, P. Yu, and Yanfang Ye. 2019. Heterogeneous Graph Attention Network. In Proceedings of WWW 2019, Jennifer B. Sartor, Theo D'Hondt, and Wolfgang De Meuter (Eds.). ACM, New York, NY, USA, Article 4, 11 pages. https://doi.org/10.475/123_4

Abstract

Abstract

- Heterogeneous graph
  - Contains different types of nodes and links(=edges)
  - Has not been considered. But so important.
- Challenges
  - Heterogeneity
  - Semantic information
- Recently, the most exciting mechanism: Attention mechanism
- 따라서 hierarchical attention에 기반한 novel heterogeneous GNN 소개
  - Node level과 semantic level attention 모두 포함하는 network
  - Node level = meta-path based neighbors의 중요도를 학습
  - Semantic level = meta-paths의 차이점을 학습
- 우리가 제안한 모델은 heterogeneous graph에서 뛰어난 성능 뿐만 아니라 그래프 분석에 있어 해석 가능성의 잠재력을 보여준다.

Introduction

Introduction

- Attention mechanism
  - Deals with variable sized data(가변량 데이터를 다룸)
  - Focus on the most salient parts of data(데이터의 가장 두드러진 부분에 집중)
  - 따라서 다양한 applications에 적용

- Meta-path?
  - a composite relation connecting two objects
  - to capture the semantics.

- Meta-path 예시
  - Movie, Actor, Director의 관계(=links, edges)를 설명해보자.
  - 두 영화에 동일한 배우가 출연했다면 Movie-Actor-Movie로 관계를 표현할 수 있다.
  - 그렇다면 Movie-Director-Movie 관계는 무엇을 의미할까?
  - 두 영화를 같은 감독이 제작했다는 것을 알 수 있다.
  - 따라서 다른 의미를 가질 수 있는 edges를 meta-path라고 정의한다.

Introduction

- 주요 challenge : heterogeneity
  - Actor node has age, sex, and nationality features.
  - Movie node has story and actors features.

- How to handle such complex structural information ?

Introduction

❑ Node-level attention

- aims to learn the importance of meta-path "based neighbors" and assign different attention values to them.
- Meta-path로 생성된 neighbors 중 유익한 neighbors를 구분할 수 있어야 한다.
- 예시) Movie-Director-Movie, Movie-Director-Movie 라는 두 meta-path 존재 (same director)
  - 따라서 만약 두 영화가 타이타닉과 터미네이터라면, 같은 감독이 제작했기 때문에 edges 형성.
  - 하지만 목적이 터미네이터2 영화 장르 구분이라면, 타이타닉보단 터미네이터가 더 큰 weight(=attention values)를 가져야 한다.

❑ Semantic-level attention

- aims to learn the importance of each "meta-path" and assign proper weights to them
- 예시) 두 영화가 존재(node), 두 영화 모두 주연배우가 동일 & 동일한 년도에 제작됨.
  - 따라서 두 영화의 관계는 (Movie-Actor-Movie) 이나 (Movie-Year-Movie) 의 meta-paths에 의해 설명될 수 있다.
  - 하지만 목적이 영화의 장르 구분이라면 제작년도 보다는 출연한 배우가 조금 더 중요한 가중치를 가진다. 따라서 Movie-Actor-Movie meta path가 더 큰 weights를 가져야한다.

Introduction

❑ 학습방법

- Input : node features
- Type-specific transformation matrix를 사용하여 same space에 project
- Node-level attention 진행(attention values 학습. Neighbors를 기반으로)
- Semantic-level attention 진행(attention values 학습. Different math-path 기반)
- 최적화 진행 및 backpropagation 진행

❑ Brief conclusion:

- heterogeneous graph를 다루고, variable data size를 다룰 때 node feature data가 있다면 해당 모델 사용가능!
- 다만 사전 정의된 node type이 있어야하고, 해당 type에 맞는 type-specific transformation matrix가 필요함.

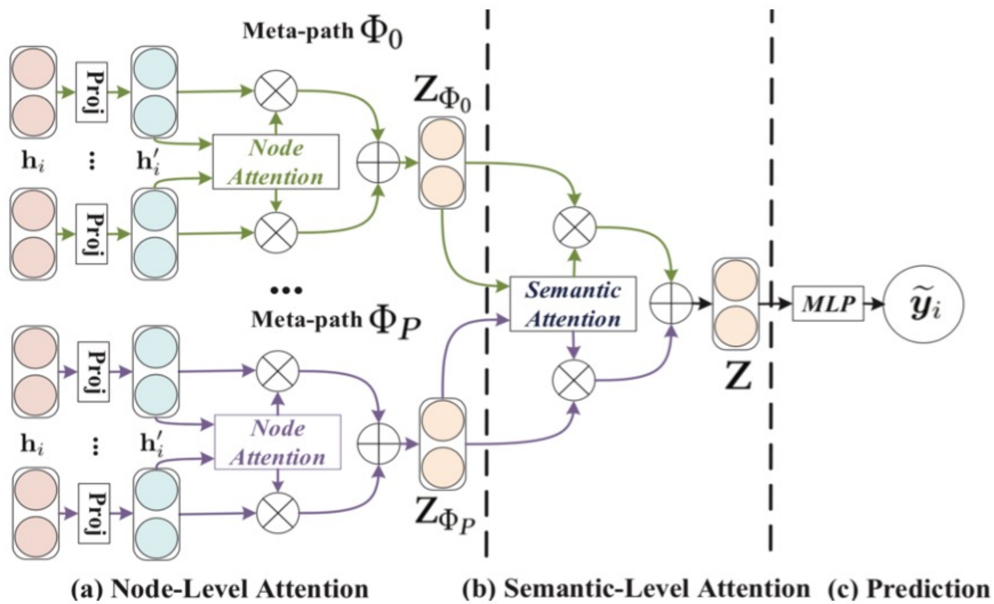
PRELIMINARY

Notations

- Heterogeneous Graph
  - $G = (V, E), V = \ object\ set, E = \ link\ set$
  - $\varphi : mapping\ function, node\ to\ predefined\ node\ type$
    - $\varphi : V \ \rightarrow \ \mathcal{A}\ , \mathcal{A} = predefined\ node\ type$
  - $\psi : mapping\ function, link\ to\ predefined\ link\ type$
    - $\psi : E \ \rightarrow \ \mathcal{R}\ , \mathcal{R} = \text{predefined link type}$
- Meta-path(Φ)
  - $\Phi : A_1 \xrightarrow{R_1} A_2 \xrightarrow{R_2} \dots \xrightarrow{R_n} A_n$ 와 같은 하나의 path를 의미(not edge, link), A는 object
  - Different meta-paths always reveal different semantics
- Meta-path based neighbors(N_i^{Φ})
  - $Given\ node\ i, meta\ path\ \Phi \ \rightarrow \ N_i^{\Phi}$
  - Set of nodes which connect with node i via meta-path Φ (includes itself)

4. PROPOSED MODEL

❑ The overall framework of the proposed model



h_i : $input(= node\ feature)$

h_i' : $projected\ node\ feature$

- Node level attention : Meta-path based neighbors의 weight를 학습
  - Output으로 semantic-specific node embedding을 얻음.
- Semantic-level attention : task에 최적화된 node embedding 조합을 얻을 수 있음.
  - Meta-paths의 차이점을 알아낼 수 있다.(semantic-level)

4.1 Node-level Attention

Node-level Attention

- Goal:
  - Meta-path based neighbors가 특정 task에 따라 다른 중요성을 가짐을 알아내는 것

- Challenge:
  - Different type of nodes have different feature spaces.
  - Solution: use type-specific transformation matrix(M_{ϕ_i})
    - Project the features of different types of nodes into the same feature space.
    - $h_i' = M_{\phi_i} * h_i$

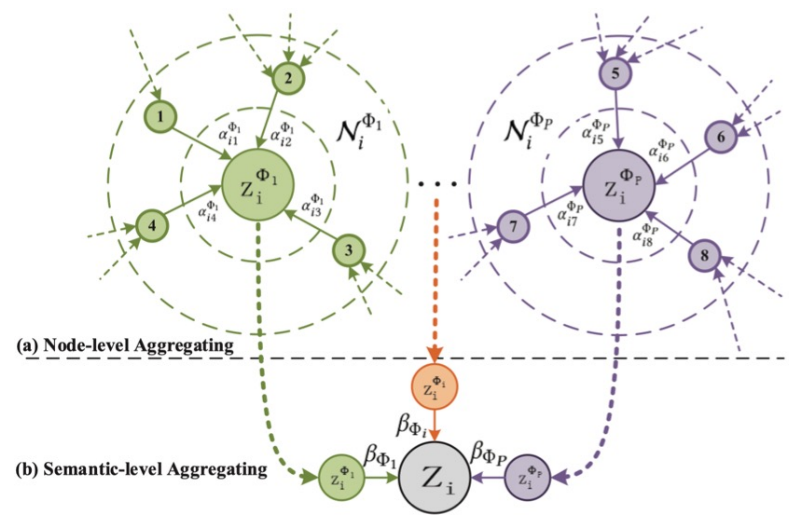
- 모든 feature가 same space에 존재하도록 project 후, importance($e_{i,j}^{\Phi}$) 학습
  - $e_{i,j}^{\Phi}$: meta-path Φ 로 연결된 두 object(=node i,j)의 importance
  - 즉 node i에 대해 node j 가 미치는 영향력.
  - $e_{i,j}^{\Phi}$ = $att_{node}(h_i', h_j'; \Phi)$
  - Asymmetric(node i에 node j가 미치는 영향력은 j 에 i가 미치는 영향력과는 다르다)
    - Critical property of heterogeneous graph

Node-level Attention

- Normalize $e_{i,j}^{\Phi}$ by softmax function
  - $\alpha_{i,j}^{\Phi} = softmax(e_{i,j}^{\Phi})$
  - Asymmetric.

- Generate embedding node i (z_i^{Φ})
  - $z_i^{\Phi} = \sigma(\sum_{j \in N_i^{\Phi}} \alpha_{i,j}^{\Phi} * h_j')$
  - z_i^{Φ} 는 node i의 $n$eighbors(node j)의 noramlize된 importance$(\alpha_{i,j}^{\Phi})$를 aggregate.

Heterogeneous Graph Attention Network



이때 $\alpha_{i,j}^{\Phi}$는 generated by single meta − path. 따라서 semantic information 포함

4.2. Semantic-level Attention

Semantic-level Attention

- Challenge:
  - Heterogeneous graph contains multiple types of semantic information.

- Goal:
  - To learn a more comprehensive node embedding.
  - Solution: use meta-paths & novel semantic-level attention

- $\left(\beta_{\Phi_1}, \ldots, \beta_{\Phi_p}\right) = att_{sem}(Z_{\Phi_1}, \ldots, Z_{\Phi_p})$
  - Z_{Φ_1} : node embedding(Z) p개
  - att_{sem} : semantic-level attention score를 계산하는 deep neural network
  - β_{Φ_1} : learned weights of each meta-path

Semantic-level Attention

- MLP와 같은 transformation을 이용하여 semantic-specific embedding으로 변형
  - semantic-level attention Vector(q)를 사용하여 embedding끼리의 similarity 계산
  - 이는 곧 특정 embedding의 중요도(= 높은 similarity)

- 이후 embedding similarity의 average 계산.
  - Each meta-path importance를 설명하기 위해 비교하는 기준점(=average)으로 활용

- 이후 구한 each meta-path importance를 normalize 진행(via softmax function)
  - 이는 곧, 각 특정 task에 대해 meta-path의 기여도(=importance)이다.
  - Task에 따라 해당 importance는 다르게 계산될 수 있다.

- Final embedding(Z)
  - $Z = \sum_{p=1}^{P} \beta_{\Phi_p} * Z_{\Phi_p}$, node embedding(Z_{Φ_p})에 각 meta-path의 importance(β_{Φ_p}) aggregate

- 결과적으로 node i 에 대한 node j(neighbors)의 importance와, 각 meta-path의 importance의 정보가 포함된 embedding 생성.

Semantic-level Attention

- Loss function은 task마다 다르게 정의
  - 본 예시에서는 node classification에서 minimize the cross-entropy 사용.

- 학습은 semi-supervised learning으로 labeled data를 사용하여 back propagation 진행.

4.3. Analysis of the Proposed Model

Analysis of the Proposed Model

- Benefit:
  - 다양한 유형의 node embedding으로 상호 통합 혹은 상호 향상이 될 수 있다.
  - Highly efficient and can be easily parallelized.
    - Attention 연산은 모든 node 혹은 meta-path에서 개별적으로 수행 가능
  - Type-specific matrix shared parameters for the whole heterogeneous graph
    - Parameter 수가 graph scale에 의존하지 않음.
  - good interpretability
    - Meta-path based neighbors
    - Meta-path importance 표현
    - Large value = more important for our task

5. EXPERIMENTS

EXPERIMENTS

❑ Datasets

- 1. DBLP
  - Node : papers(P), authors(A), conferences(C), terms(T)
    - Authors are divided into fours area : database, data mining, machine learning, information retrieval
    - Hence, meta-path set : APA, APCPA, APTPA
- 2. ACM
  - Node : papers(P), authors(A) and subjects(S)
    - Papers are divided into three class : database, wireless communication, data mining
    - Hence, meta-path set : PAP, PSP
- 3. IMDB

❑ Compare with baselines

- 동일한 datasets에 대해 본 논문에서 소개한 모델과 기존 방법들을 비교

EXPERIMENTS

❑ Compare with baselines (fair comparison : embedding dim(=64))

Table 3: Qantitative results (%) on the node classification task.

Datasets	Metrics	Training	DeepWalk	ESim	metapath2vec	HERec	GCN	GAT	HAN_{nd}	HAN_{sem}	HAN
ACM	Macro-F1	20%	77.25	77.32	65.09	66.17	86.81	86.23	88.15	89.04	**89.40**
		40%	80.47	80.12	69.93	70.89	87.68	87.04	88.41	89.41	**89.79**
		60%	82.55	82.44	71.47	72.38	88.10	87.56	87.91	**90.00**	89.51
		80%	84.17	83.00	73.81	73.92	88.29	87.33	88.48	90.17	**90.63**
	Micro-F1	20%	76.92	76.89	65.00	66.03	86.77	86.01	87.99	88.85	**89.22**
		40%	79.99	79.70	69.75	70.73	87.64	86.79	88.31	89.27	**89.64**
		60%	82.11	82.02	71.29	72.24	88.12	87.40	87.68	**89.85**	89.33
		80%	83.88	82.89	73.69	73.84	88.35	87.11	88.26	89.95	**90.54**
DBLP	Macro-F1	20%	77.43	91.64	90.16	91.68	90.79	90.97	91.17	92.03	**92.24**
		40%	81.02	92.04	90.82	92.16	91.48	91.20	91.46	92.08	**92.40**
		60%	83.67	92.44	91.32	92.80	91.89	90.80	91.78	92.38	**92.80**
		80%	84.81	92.53	91.89	92.34	92.38	91.73	91.80	92.53	**93.08**
	Micro-F1	20%	79.37	92.73	91.53	92.69	91.71	91.96	92.05	92.99	**93.11**
		40%	82.73	93.07	92.03	93.18	92.31	92.16	92.38	93.00	**93.30**
		60%	85.27	93.39	92.48	93.70	92.62	91.84	92.69	93.31	**93.70**
		80%	86.26	93.44	92.80	93.27	93.09	92.55	92.69	93.29	**93.99**
IMDB	Macro-F1	20%	40.72	32.10	41.16	41.65	45.73	49.44	49.78	**50.87**	50.00
		40%	45.19	31.94	44.22	43.86	48.01	50.64	52.11	50.85	**52.71**
		60%	48.13	31.68	45.11	46.27	49.15	51.90	51.73	52.09	**54.24**
		80%	50.35	32.06	45.15	47.64	51.81	52.99	52.66	51.60	**54.38**
	Micro-F1	20%	46.38	35.28	45.65	45.81	49.78	55.28	54.17	55.01	**55.73**
		40%	49.99	35.47	48.24	47.59	51.71	55.91	56.39	55.15	**57.97**
		60%	52.21	35.64	49.09	49.88	52.29	56.44	56.09	56.66	**58.32**
		80%	54.33	35.59	48.81	50.99	54.61	56.97	56.38	56.49	**58.51**

GAT : 기존 homogeneous graph를 위해 제작된 model
HAN : 본 논문에서 소개한 heterogeneous graph를 위한 GAT 모델

EXPERIMENTS

- Details for HAN
  - Initialize parameters and optimize : Adam
  - Learning rate : 0.005
  - The dimension of the semantic-level attention vector : 128
  - Number of attention head : 8
  - Dropout of attention : 0.6
  - Early stopping patience : 100
    - Validation loss doesn't decrease

- GCN, GAT, HAN : same training set, validation set and test set

- DeepWalk, Esim, metapath2vec and HERec (based random walk based methods)
  - Window size : 5
  - Walk length : 100
  - Walks per node : 40
  - Number of negative samples : 5

EXPERIMENTS

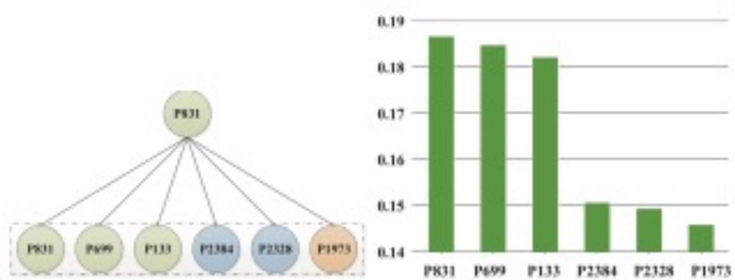
❑ Clustering

Table 4: Qantitative results (%) on the node clustering task.

Datasets	Metrics	DeepWalk	ESim	metapath2vec	HERec	GCN	GAT	HAN_{nd}	HAN_{sem}	HAN
ACM	NMI	41.61	39.14	21.22	40.70	51.40	57.29	60.99	61.05	**61.56**
	ARI	35.10	34.32	21.00	37.13	53.01	60.43	61.48	59.45	**64.39**
DBLP	NMI	76.53	66.32	74.30	76.73	75.01	71.50	75.30	77.31	**79.12**
	ARI	81.35	68.31	78.50	80.98	80.49	77.26	81.46	83.46	**84.76**
IMDB	NMI	1.45	0.55	1.20	1.20	5.45	8.45	9.16	10.31	**10.87**
	ARI	2.15	0.10	1.70	1.65	4.40	7.46	7.98	9.51	**10.01**

- 역시나 기존 다른 방법들보다 성능이 뛰어남을 확인할 수 있다.

❑ Example of ACM



(a) Meta-path based neighbors of P831 (b) Attention values of P831's neighbors

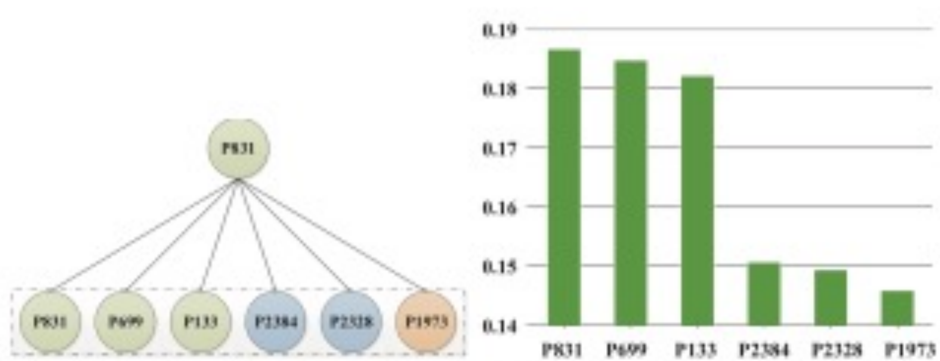
Figure 4: Meta-path based neighbors of node P831 and corresponding attention values (Different colors mean different classes, e.g., *green* means Data Mining, *blue* means Database, *orange* means Wireless Communication).

-> Good interpretability

EXPERIMENTS

❑ Analysis of Hierarchical Attention Mechanism

- 1. Analysis of node-level attention
  - Large attention values = important neighbors for the specific task



(a) Meta-path based neighbors of P831 (b) Attention values of P831's neighbors

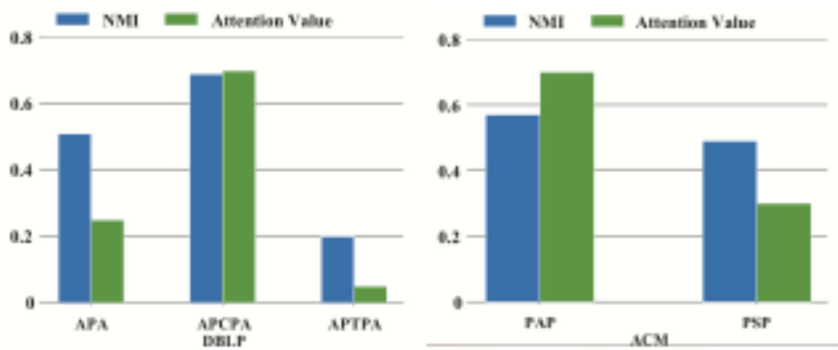
Figure 4: Meta-path based neighbors of node P831 and corresponding attention values (Different colors mean different classes, e.g., *green* means Data Mining, *blue* means Database, *orange* means Wireless Communication).

- PAP meta-path가 주어졌을 때
- Measure attention values(P831 이웃 nodes)
- Itself attention value를 제외하면 P699, P133
  - Green color(=Data mining)
- 두 paper를 모두 제작한 author 라는 meta-path에 based neighbors의 중요도는 당연히 비슷한 장르(Data mining)의 이웃 노드 P699, P133이 높은 importance를 가진다.

EXPERIMENTS

❑ Analysis of Hierarchical Attention Mechanism

- 2. Analysis of semantic-level attention.



(a) NMI values on DBLP (b) NMI values on ACM

Figure 5: Performance of single meta-path and corresponding attention value.

- NMI 는 clustering 결과
- Attention value는 meta-path의 importance
- Task in DBLP : author's research area
  - APCPA meta-path is most important
  - C(conference)

EXPERIMENTS

- Visualization

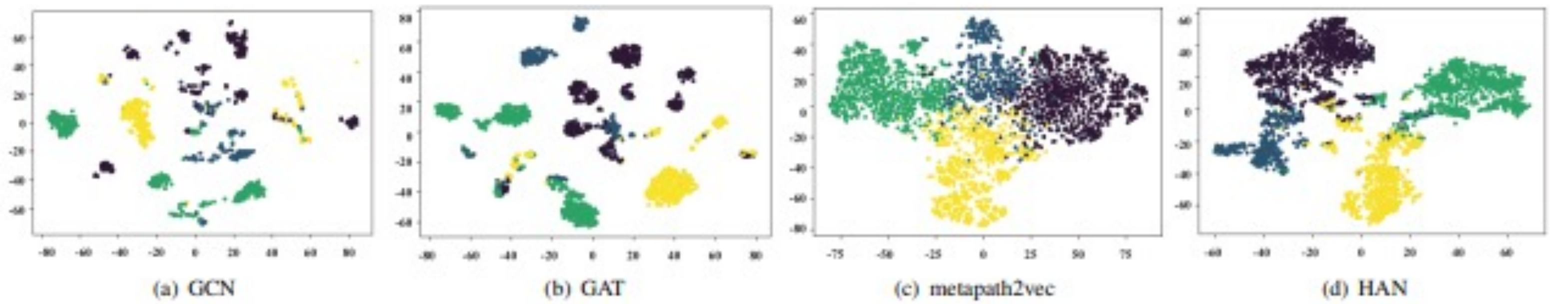
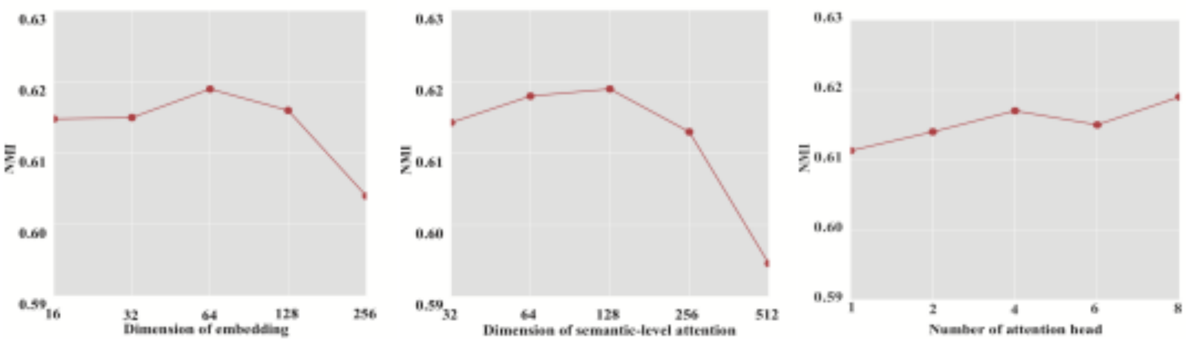


Figure 6: Visualization embedding on DBLP. Each point indicates one author and its color indicates the research area.

EXPERIMENTS

Y축 : NMI
= results of clustering



(a) Dimension of the final embedding **Z** (b) Dimension of the semantic-level attention vector **q** (c) Number of attention head K.

Figure 7: Parameter sensitivity of HAN w.r.t. Dimension of the final embedding Z, Dimension of the semantic-level attention vector q and Number of attention head K.

❑ Parameters Experiments

- Dimension of the final embedding Z(=64)
  - HAN needs a suitable dimension to encode the semantics information
  - Larger dimension may introduce additional redundancies
- Dimension of semantic-level attention vector(=128)
  - Test various dimension. The result is 128.
- Number of attention head(K)
  - Change of performance is slightly

6. CONCLUSION

CONCLUSION

- In this paper, propose heterogeneous GNN based on attention network(HAN)
- HAN can capture complex structures and rich semantics
- 계층구조. Node-level attention -> semantic-level attention
- Experimental results :
  - Classification good
  - Clustering good
- HAN has good interpretability
- **Heterogeneous graph & variable data size -> HAN**
  - **Then, we need type-specific transformation matrix**

Thank you