The Dynamics, Uncertainty and Heterogeneity in Network Embedding

Xiangliang Zhang

King Abdullah University of Science and Technology

BSMDMA Workshop @ IJCAI 2019, Aug 11

KAUST
King Abdullah University of Science and Technology

Introduce myself

Experience:

2017-Present, Associate Professor - Computer Science CEMSE Division, KAUST

2011-2017, Assistant Professor – Computer Science CEMSE Division, KAUST

2010-2011, Research Scientist, CEMSE Division, KAUST

2010, ERCIM Research Fellow, NTNU, Norway

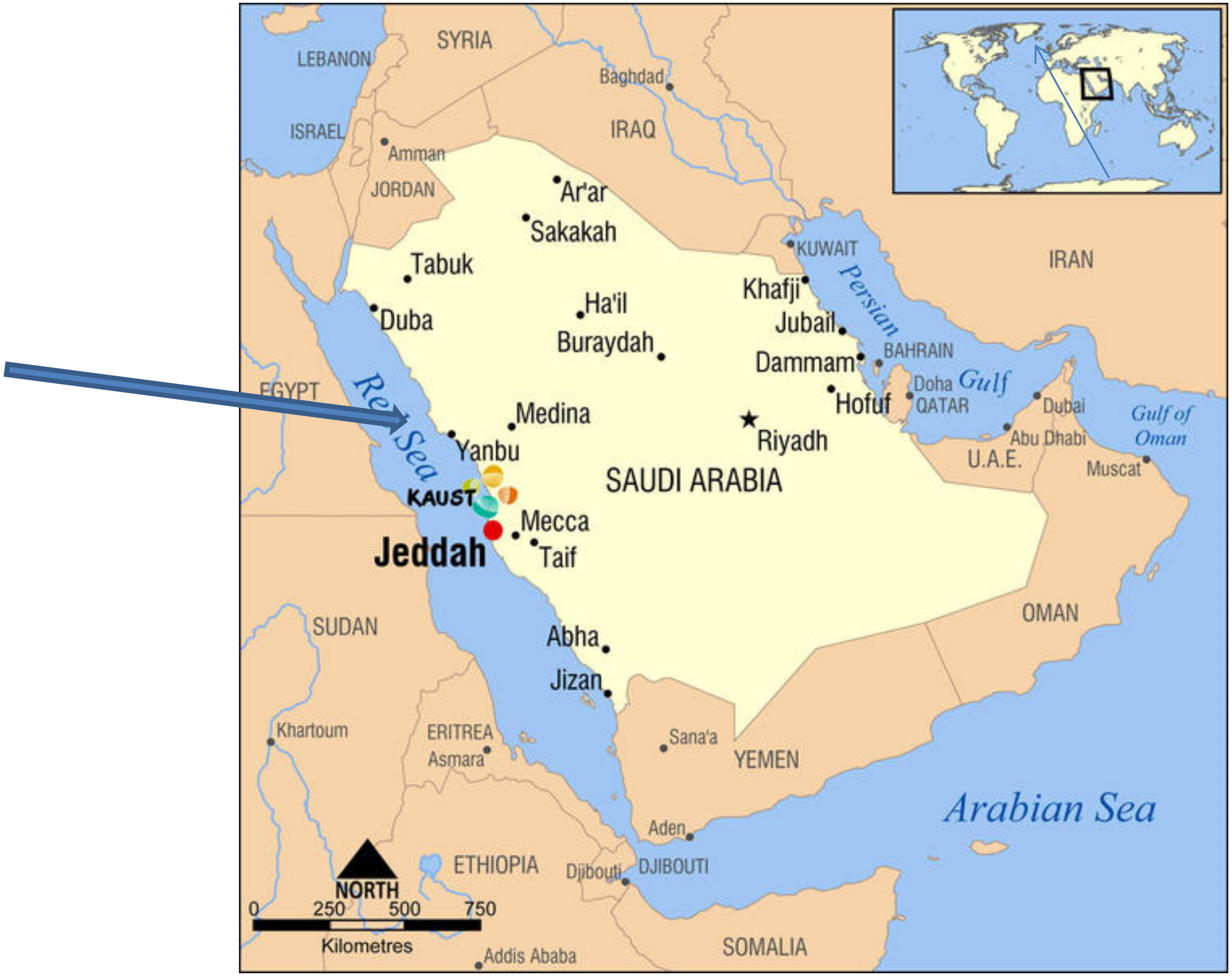
Education:

2006-2010, PhD, INRIA – University Paris Sud 11, France

2003-2006, MS, Xi’an Jiaotong University, China

1999-2003, BS, Xi’an Jiaotong University, China

KAUST in Saudi Arabia



KAUST
King Abdullah University of Science and Technology

KAUST in Numbers

Since September 2009

1000+ students (Masters + Ph.D. only)

~150 faculties

700+ Research scientist and Post-docs

100+ nationalities

3 Divisions, 11 Research Centers

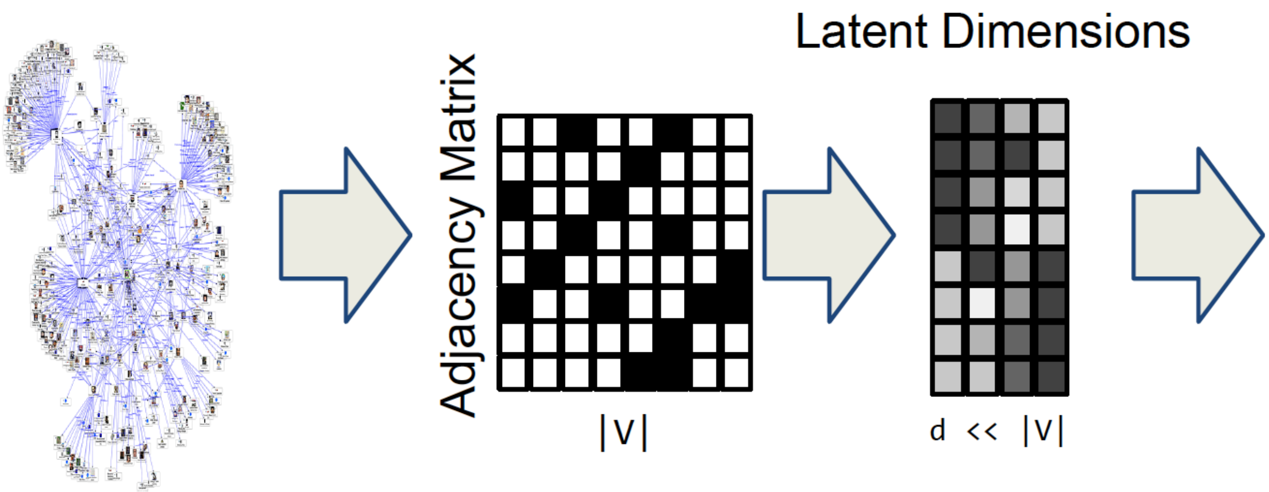
~25% female students

Outline

- Introduction of Network Embedding
- Co-embedding of Attributed Network for User Profiling
- Dynamic Embedding for User Profiling
- Walking with Reinforcement for Semi-supervised Embedding learning from Attributed Network
- Active Heterogenous Network Embedding
- Future work Discussion

Network Embedding

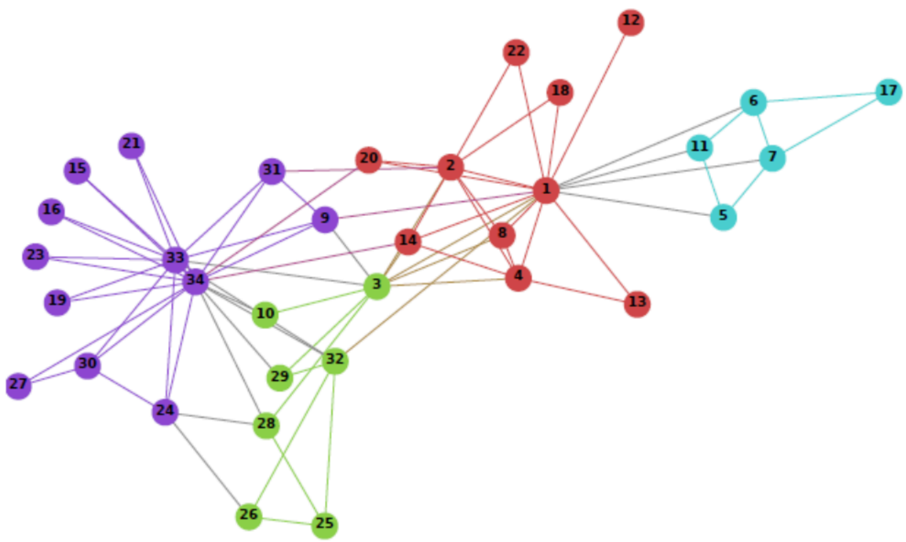
Create features by transforming the graph into a lower dimensional latent representation.



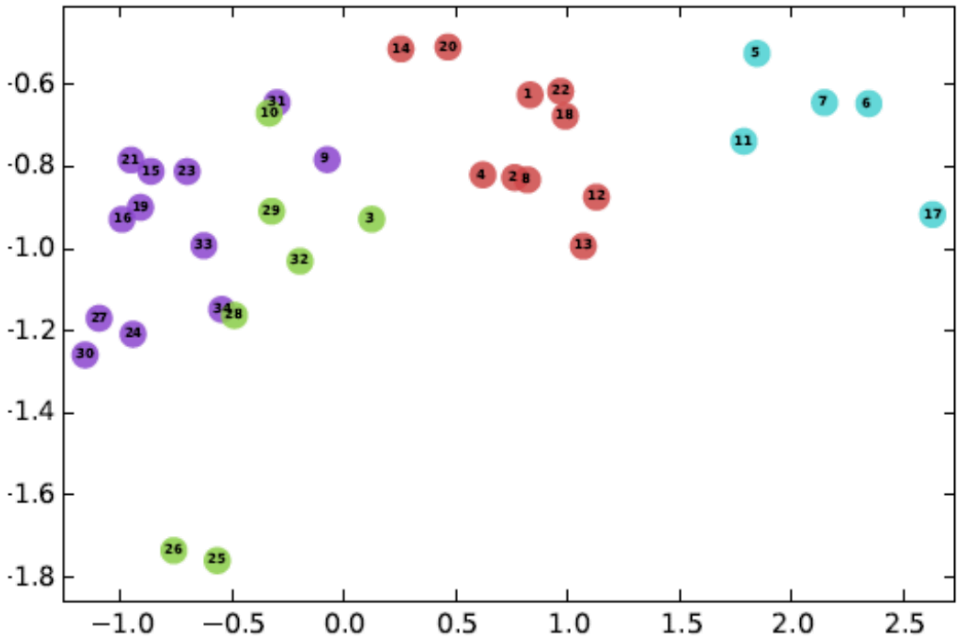
- Node Classification
- Link Prediction
- Community Detection
- Visualization
- etc

Visual Example

On Zachary's Karate Graph:

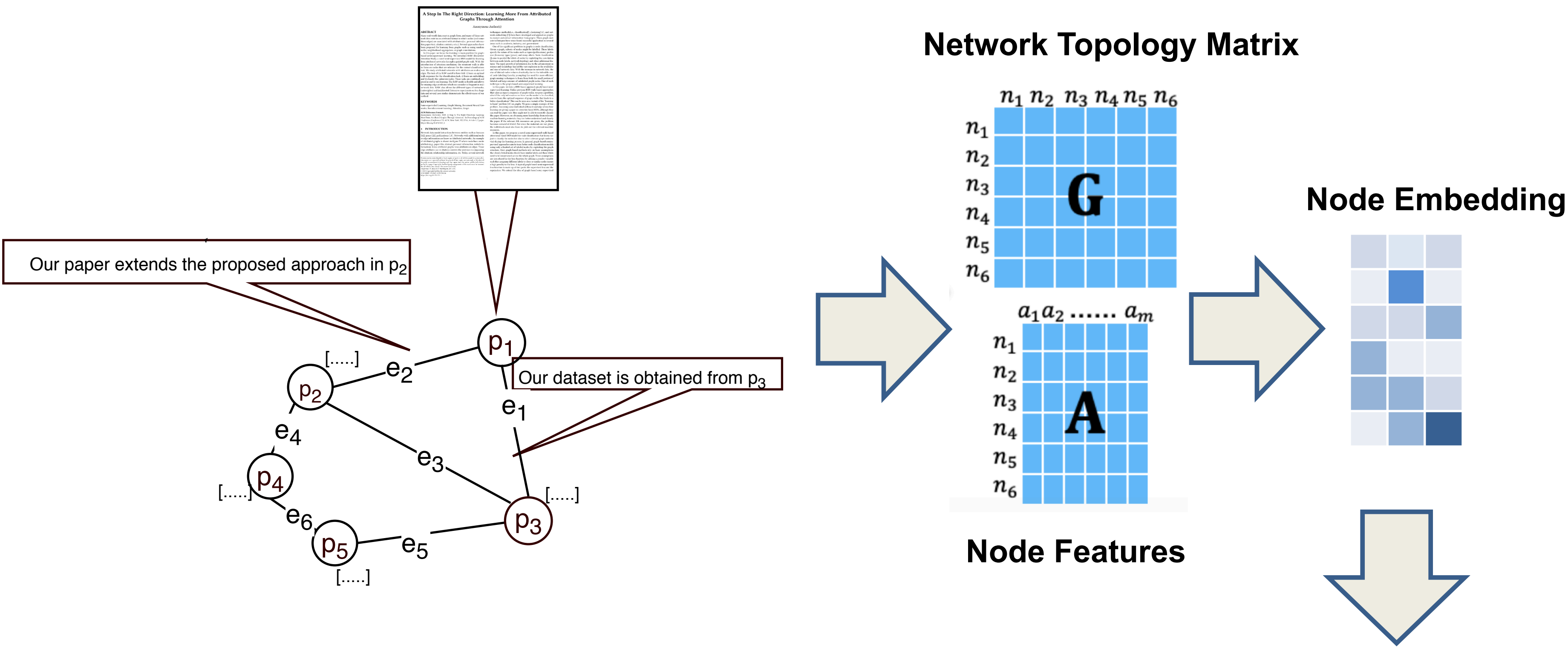


Input



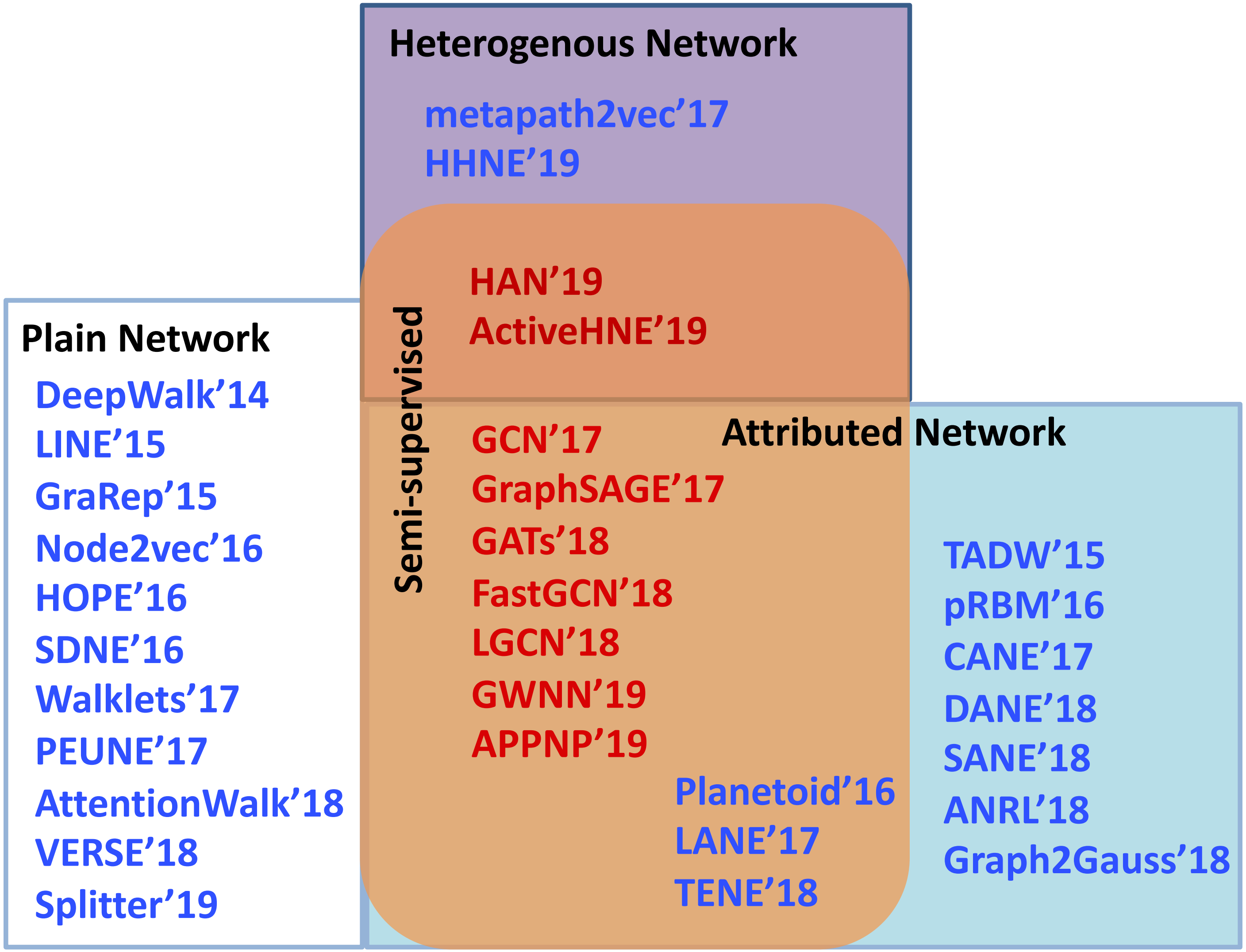
Output

Attributed Network Embedding



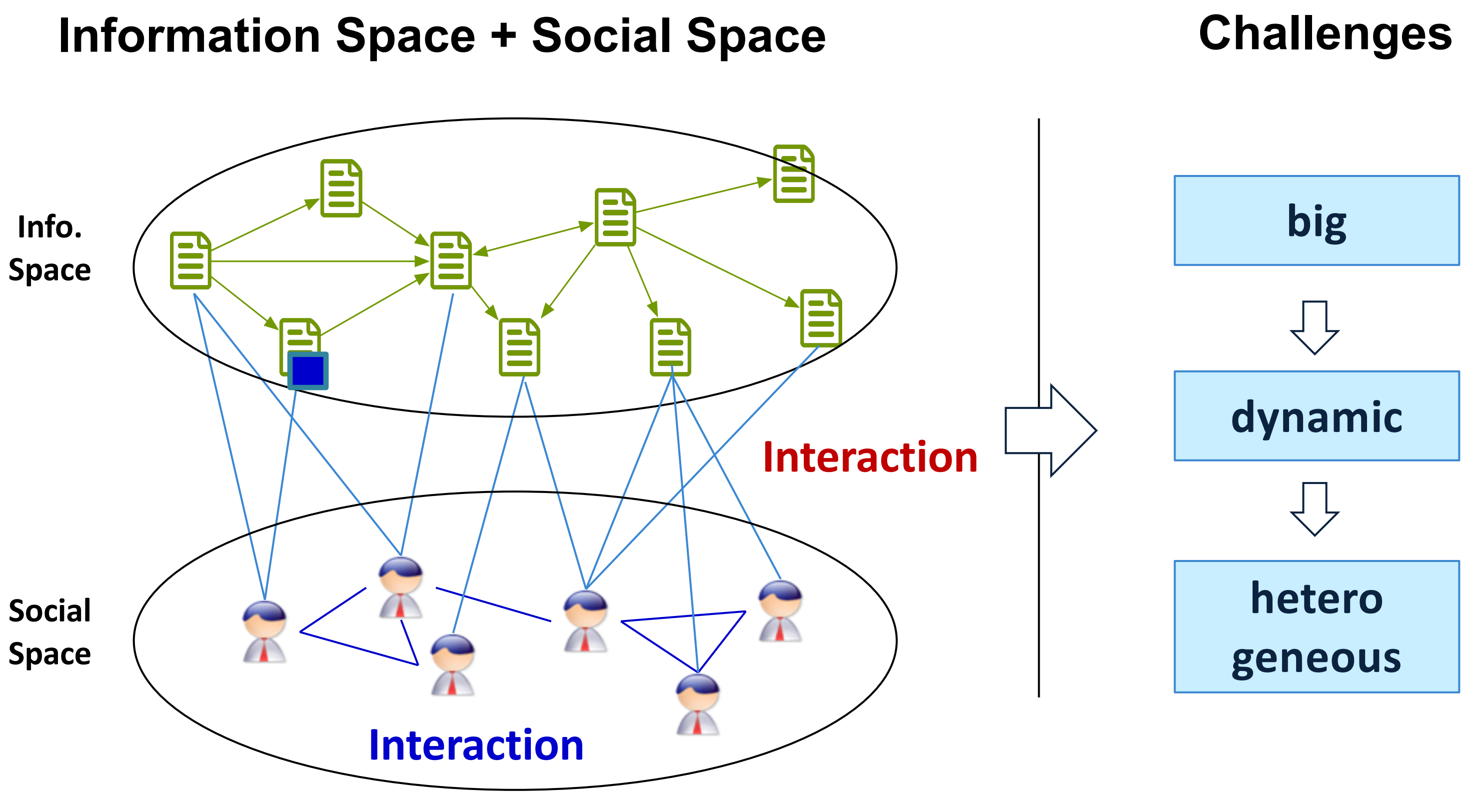
- Node Classification
- Node Profiling
- Link Prediction
- Community Detection
- Visualization, etc

Recent Work of Network Embedding



This is not a complete list. To be updated continuously

Challenges



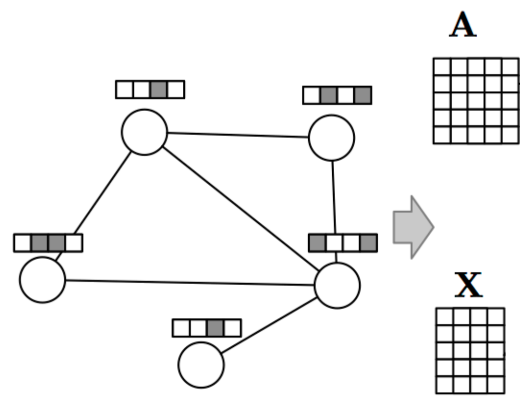
1. J. Scott. (1991, 2000, 2012). Social network analysis: A handbook.
2. D. Easley and J. Kleinberg. Networks, crowds, and markets: Reasoning about a highly connected world. Cambridge University Press, 2010.

Outline

- Introduction of Network Embedding
- Co-embedding of Attributed Network for User Profiling
- Dynamic Embedding for User Profiling
- Walking with Reinforcement for Semi-supervised Embedding learning from Attributed Network
- Active Heterogenous Network Embedding
- Future work Discussion

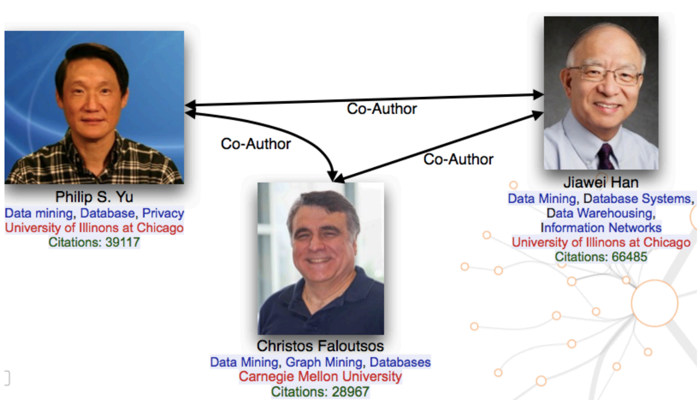
Problem: Attributed Network Co-Embedding

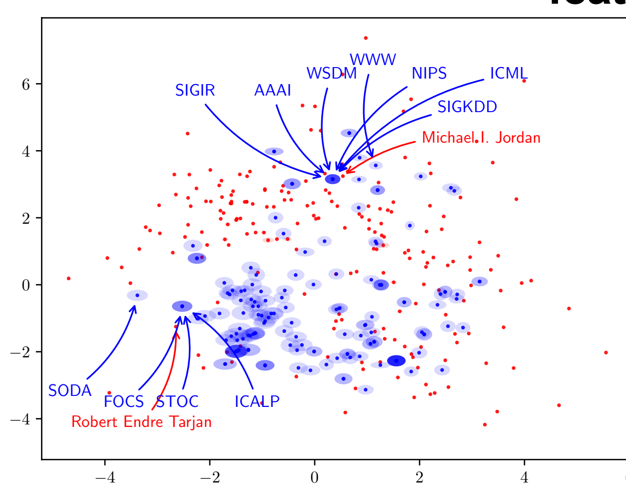
Goal: map nodes and features in the same space



Input: Attributed Network

Output: Node and feature embedding

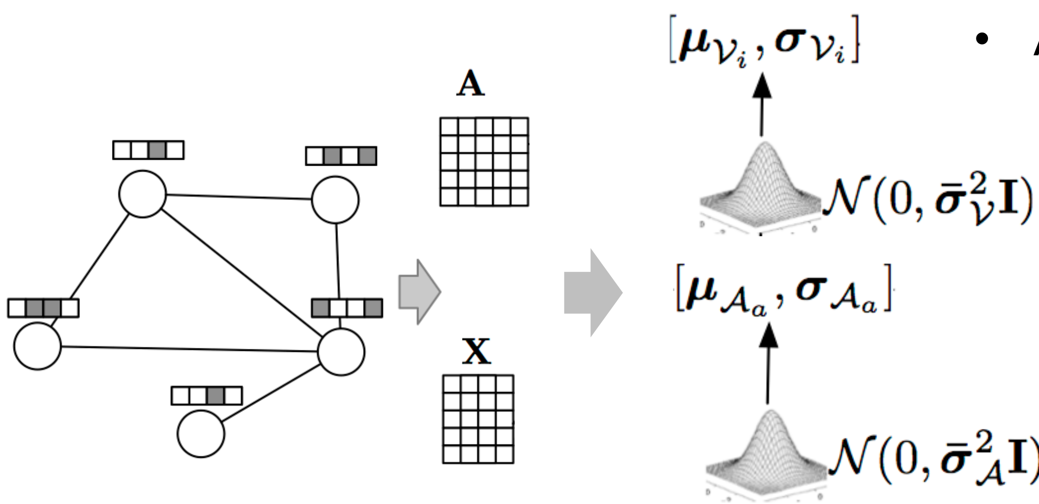




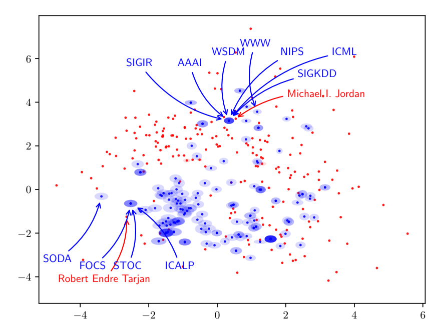
Problem: Attributed Network Co-Embedding

Our Co-embedding Model

- Represent both **node** and **feature** in the **same** space
  → to **quantitatively** measure their relationship
- Represent node/attribute by a **Gaussian distribution**, **not a single point**
- → to capture the **uncertainty**

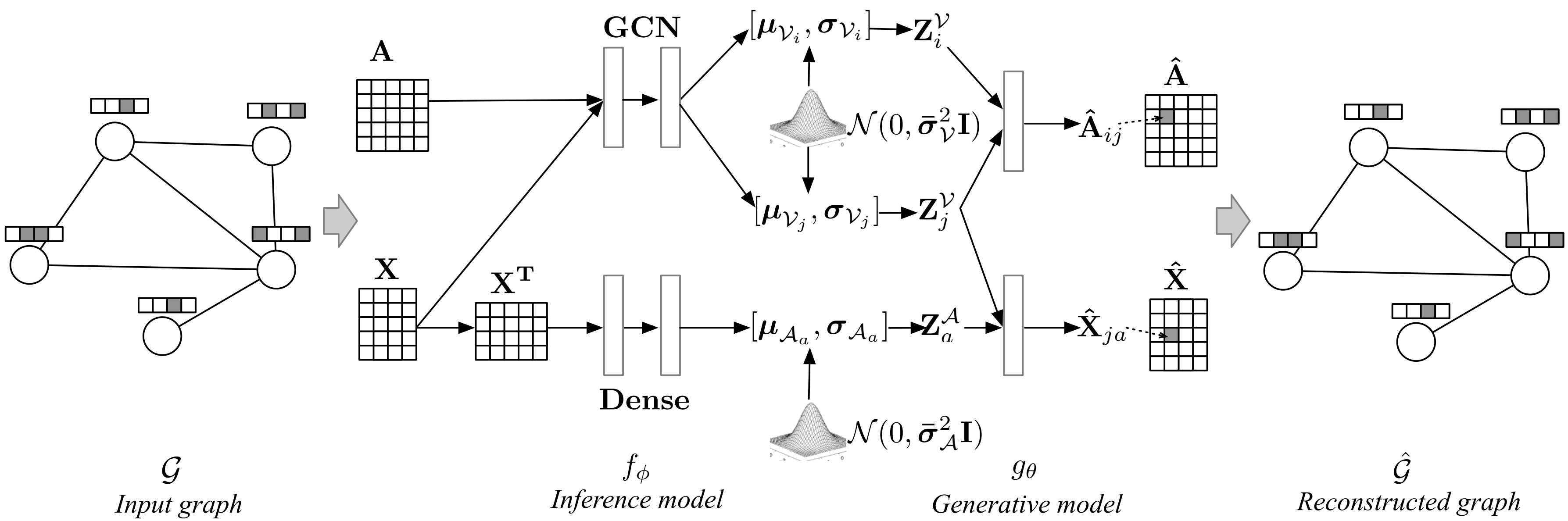


Output:
- Node representation
- Attribute representation



Co-Embedding Model for Attributed Network

The proposed model, CAN, based on VAE



Z. Meng, S. Liang, H. Bao, X. Zhang. "Co-Embedding Attributed Networks". WSDM 2019

Experimental Evaluation, datasets

	Datasets	#Nodes	#Edges	#Attributes	#Labels
Citation	**Cora**	2,708	5,429	1,433	7
	Citeseer	3,312	4,660	3,703	6
	Pubmed	19,717	44,338	500	3
Social	**BlogCatalog**	5,196	171,743	8,189	6
	Flickr	7,575	239,738	12,047	9
	Facebook	4,039	88,234	1,406	-
Co-author	**DBLP**	12,213	131,713	172	-

Applications:

- Node classification
- Link Prediction
- Attribute Inference
- User Profiling

Experimental Evaluation, results

Method	Cora		Citeseer		Facebook		Pubmed		Flickr		BlogCatalog	
	AUC	AP	AUC	AP	AUC	AP	AUC	AP	AUC	AP	AUC	AP
AANE	.767	.720	.785	.765	.842	.834	.783	.754	.709	.697	.711	.714
GaphSAGE	.795	.763	.802	.791	.854	.846	.840	.829	.732	.728	.723	.702
ANRL-WAN	.832	.843	.867	.848	.935	.912	.918	.897	.724	.763	.762	.758
GAE	.914	.926	.908	.920	.980	.979	.944	.947	.828	.827	.821	.821
CAN	**.985**	**.984**	**.950**	**.958**	**.988**	**.986**	**.980**	**.977**	**.914**	**.922**	**.837**	**.837**

Link Prediction

Method	Cora		Citeseer		Facebook		Pubmed		Flickr		BlogCatalog	
	AUC	AP	AUC	AP	AUC	AP	AUC	AP	AUC	AP	AUC	AP
EdgeExp	.682	.690	.707	.714	.671	.687	.586	.576	.678	.685	.684	.744
SAN	.664	.672	.679	.675	.712	.723	.579	.572	.653	.660	.694	.710
BLA	.808	.801	.854	.876	.868	.830	.622	.602	.730	.769	.787	.792
CAN	**.932**	**.916**	**.954**	**.939**	**.974**	**.971**	**.670**	**.652**	**.867**	**.865**	**.868**	**.867**

Attribute Inference

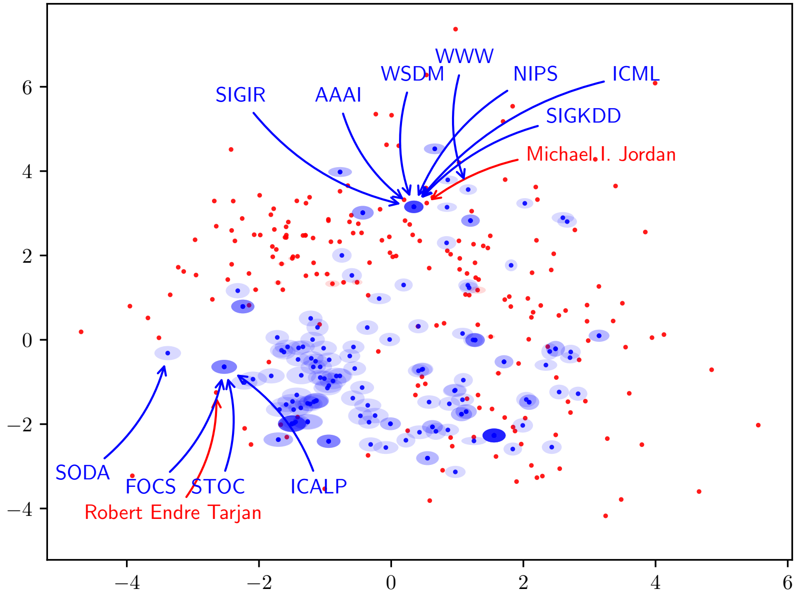
Experimental Evaluation, results

User Profiling

Experts	Top-4 conferences
Michael I. Jordan	NIPS, UAI, ICML, ICASSP
Geoffrey E. Hinton	NIPS, ICML, ICCV, ICASSP
Yann LeCun	NIPS, ICCV, ICML, CVPR
Robert Endre Tarjan	FOCS, STOC, COLT, SODA
Sanjeev Arora	SODA, FOCS, ICALP, COLT

Machine Learning Computer Vision

Theoretical Computer Science



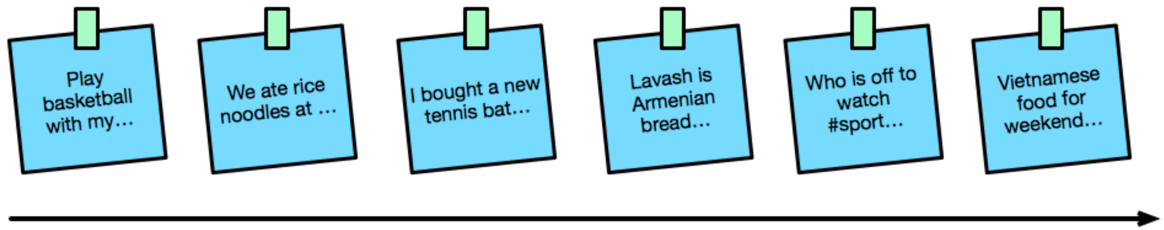
Outline

- Introduction of Network Embedding
- Co-embedding of Attributed Network for User Profiling
- Dynamic Embedding for User Profiling
- Walking with Reinforcement for Semi-supervised Embedding learning from Attributed Network
- Active Heterogenous Network Embedding
- Future work Discussion

The Task

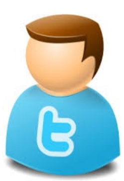
Input: A stream of tweets generated across the time

Twitter Users



Tweets over time

Output: A set of keywords to profile the user at different point in time



Given a user at time t

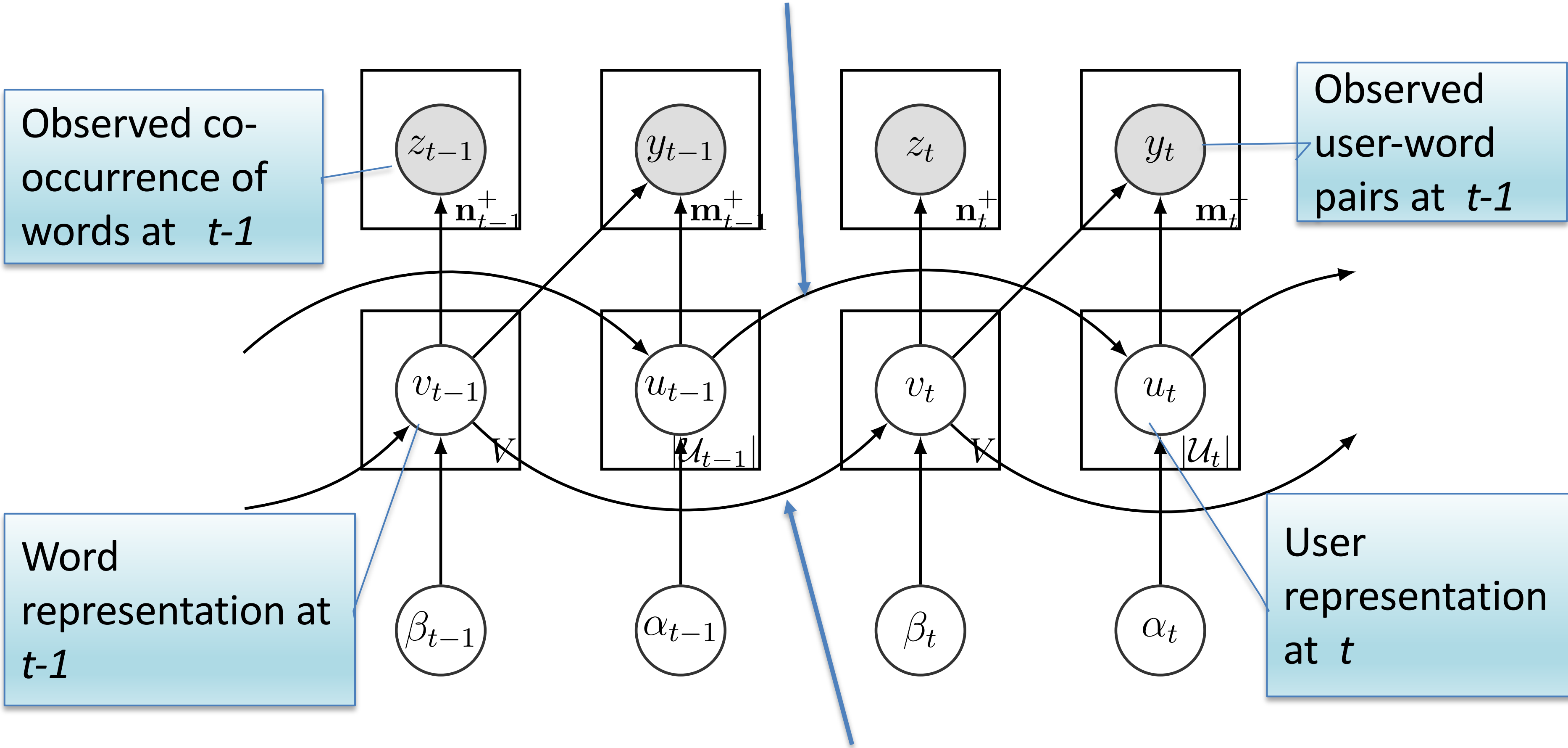
Sport
Food

Vacation?
Holiday?

At time t+1

Dynamic User and Word Embedding

User Diffusion $p(\mathbf{U}_t \mid \mathbf{U}_{t-1}) \propto \mathcal{N}(\mathbf{U}_{t-1}, \boldsymbol{\alpha}_{t-1}^2 \mathbf{I}) \cdot \mathcal{N}(\mathbf{0}, \overline{\boldsymbol{\alpha}}_0^2 \mathbf{I})$



Word Diffusion $p(\mathbf{V}_t \mid \mathbf{V}_{t-1}) \propto \mathcal{N}(\mathbf{V}_{t-1}, \boldsymbol{\beta}_{t-1}^2 \mathbf{I}) \cdot \mathcal{N}(\mathbf{0}, \overline{\boldsymbol{\beta}}_0^2 \mathbf{I})$

S. Liang, X. Zhang, Z. Ren, E. Kanoulas. Dynamic Embeddings for User Profiling in Twitter. KDD 2018

DUWE model inference

- Apply the skip-gram filtering for the inference (Bamler et al. 2017) and the variational inference algorithm to obtain the embeddings
- Posterior distribution over $\mathbf{U}_{\leq t}$ and $\mathbf{V}_{\leq t}$ conditional on the statistics information $\mathbf{m}^{\pm}_{\leq t}$ and $\mathbf{n}^{\pm}_{\leq t}$ as follows:

$$p(\mathbf{U}_{\leq t}, \mathbf{V}_{\leq t} \mid \mathbf{m}^{\pm}_{\leq t}, \mathbf{n}^{\pm}_{\leq t}) = \frac{p(\mathbf{m}^{\pm}_{\leq t}\mathbf{n}^{\pm}_{\leq t}, \mathbf{U}_{\leq t}, \mathbf{V}_{\leq t})}{\iint p(\mathbf{m}^{\pm}_{\leq t}\mathbf{n}^{\pm}_{\leq t}, \mathbf{U}_{\leq t}, \mathbf{V}_{\leq t}) d\mathbf{U}_{\leq t}\, d\mathbf{V}_{\leq t}}$$

positive and negative indicator matrices for all user-to-word pairs

positive and negative indicator matrices for all word-to-word pairs

where we have:

model transition for users

model transition for words

$$p(\mathbf{m}^{\pm}_{\leq t}\mathbf{n}^{\pm}_{\leq t}, \mathbf{U}_{\leq t}, \mathbf{V}_{\leq t}) = \prod_{t'=1}^{t} \Bigg(\boxed{p(\mathbf{U}_{t'} \mid \mathbf{U}_{t'-1})} \boxed{p(\mathbf{V}_{t'} \mid \mathbf{V}_{t'-1})} \times \Big(\prod_{k,l=1}^{V} \boxed{p(\mathbf{n}^{\pm}_{k,l,t'} \mid \mathbf{v}_k, \mathbf{v}_l)} \Big) \cdot \Big(\prod_{i=1}^{|\mathcal{U}_{t'}|} \prod_{k=1}^{V} \boxed{p(\mathbf{m}^{\pm}_{u_i,k,t'} \mid \mathbf{u}_i, \mathbf{v}_k)} \Big) \Bigg)$$

skip-gram model for words

skip-gram model for user and words

An Example User's Dynamic Profiling Results over Time

Top-6 keywords of an example user's dynamic profile, whose interests cover a number of aspects and dramatically change over time, from Sport, fitness, kitchen, exercise, to education.

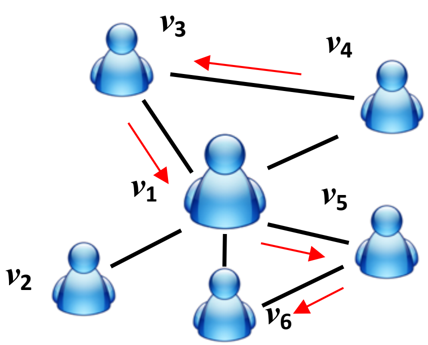
	Apr. 2014 to Jun. 2014	Jul. 2014 to Sep. 2014	Oct. 2014 to Dec. 2014
Ground Truth	badminton leaf basketball flower bicycling root	muscle apple heart kiwi lungs pomelo	freezer fly toaster cockroach cabinet ant
DPDR	badminton sky basketball herb coach grass	heart apple ankle pomelo finger peach	freezer water muffin fly toaster cockroach
DUWE	badminton flower basketball leaf bicycling fruit	heart apple muscle kiwi breath pomelo	freezer ant dishwaster fly toaster cockroach

Outline

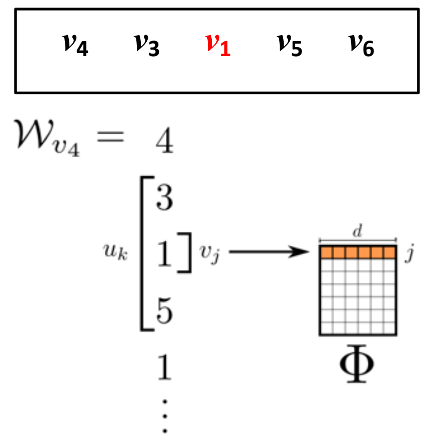
- Introduction of Network Embedding
- Co-embedding of Attributed Network for User Profiling
- Dynamic Embedding for User Profiling
- Walking with Reinforcement for Semi-supervised Embedding learning from Attributed Network
- Active Heterogenous Network Embedding
- Future work Discussion

DeepWalk and its extensions

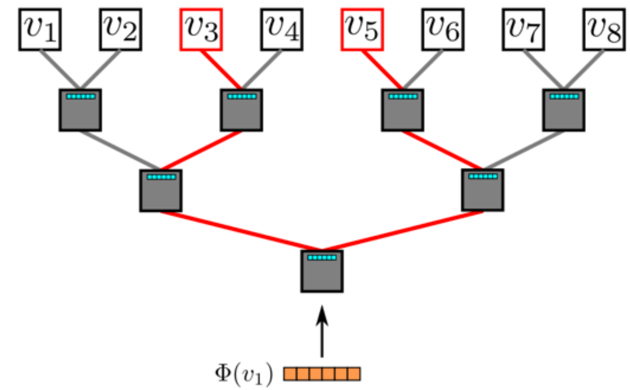
Random walk

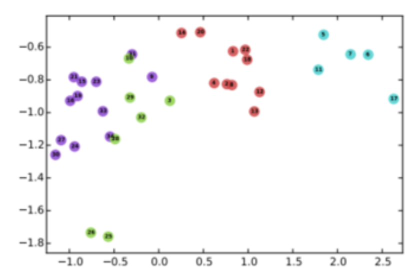


One example RW path



SkipGram with Hierarchical softmax





$$P(\{v_{i-w}, \cdots, v_{i+w}\} \backslash v_i | \Phi(v_i)) = \prod_{j=i-w, j \neq i}^{i+w} P(v_j | \Phi(v_i))$$

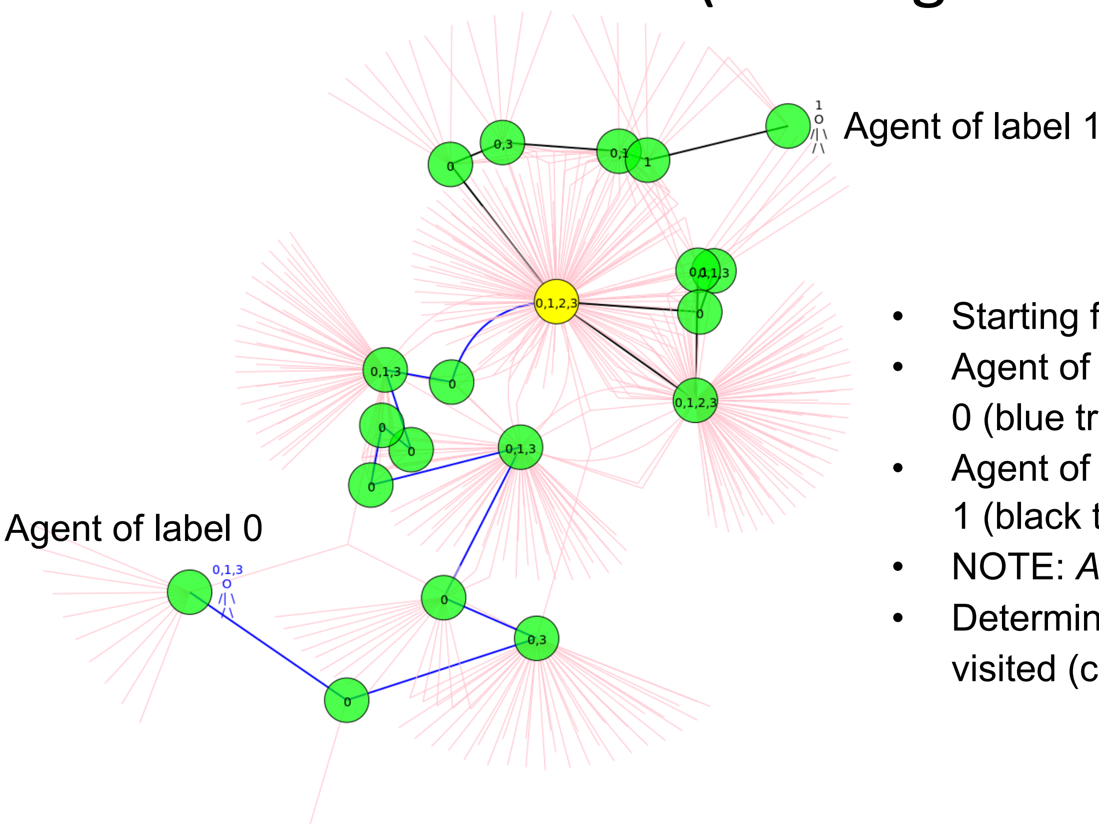
Hierarchical softmax

$$P(v_j | \Phi(v_i)) = \prod_{l=1}^{\log |V|} P(b_l | \Phi(v_i)) = \prod_{l=1}^{\log |V|} 1/(1 + e^{-\Phi(v_i) \cdot \Psi(b_l)})$$

B. Perozzi, R. Al-Rfou, and S. Skiena. 2014. Deepwalk: Online learning of social representations. *KDD*, 701–710.

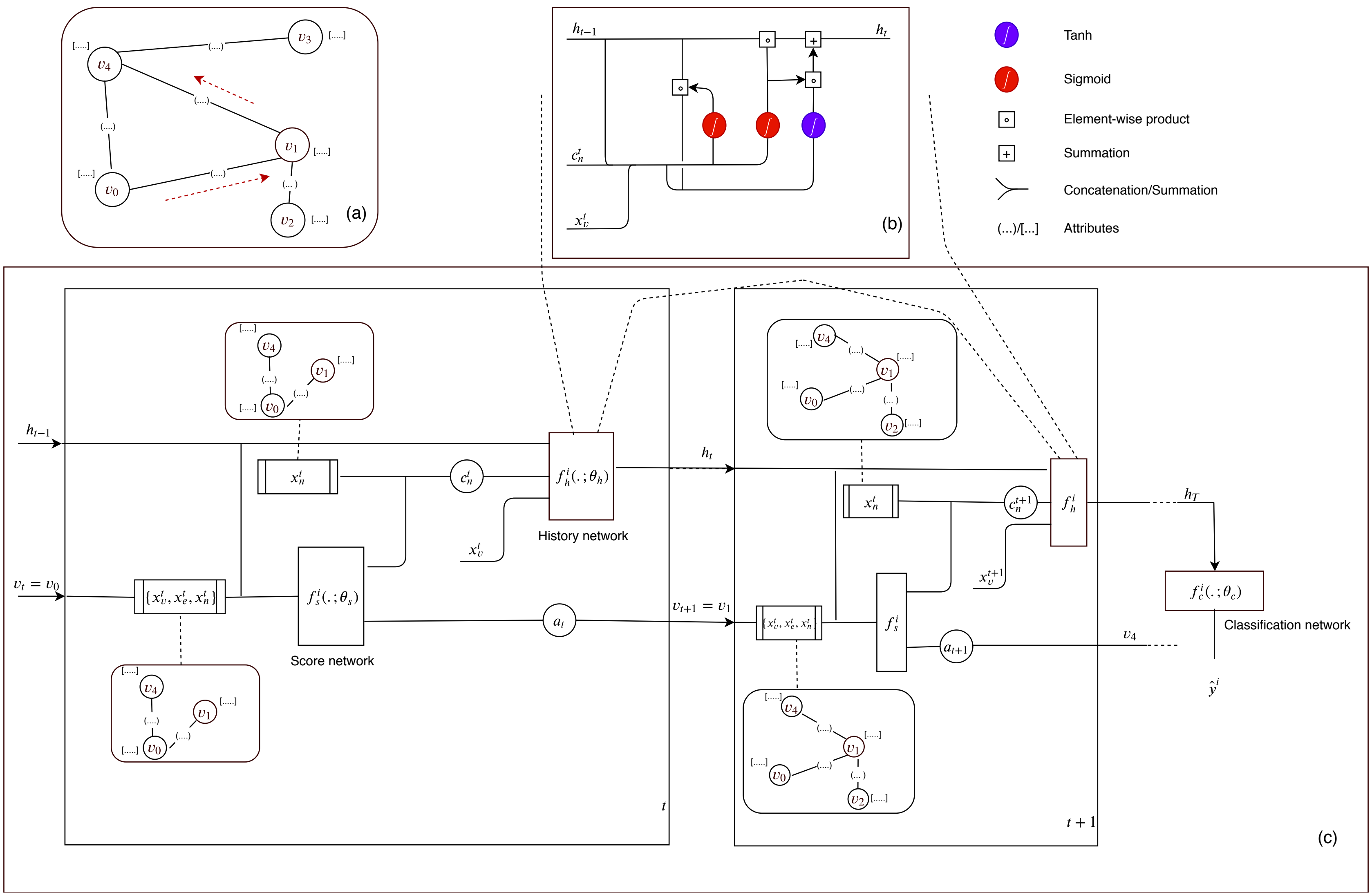
Walking with reinforcement?

- **Task**: walking to classify nodes
- **Solution**: learn node embedding in **semi-supervised** way with **reinforcement** (earning rewards to classify correctly)



- Starting from the yellow node, with labels {0, 1, 2, 3}
- Agent of label 0 visited a number of nodes with label 0 (blue trajectory)
- Agent of label 1 visited a number of nodes with label 1 (black trajectory)
- NOTE: *Agents do not know labels when walking.*
- Determine where to go by knowing what has been visited (content of previously visited nodes)

Our proposed Multi-Label-Graph-Walk (MLGW) approach



U Akujuobi, Y Han, Q Zhang, X Zhang. Collaborative Graph Walk for Semi-supervised Multi-Label Node Classification. ICDM 2019

Evaluation Results

	$\mid V \mid$	$\mid E \mid$	$\mid L \mid$	$\mid V^L \mid$
DBLP	28,702	68,335	4	28,702
Delve-R	1,229,280	4,322,275	20	131,991

	DBLP						Delve-R					
	Tr-1						Tr-1					
	precision		Recall		F1		precision		Recall		F1	
	macro	micro	macro	micro	macro	micro	macro	micro	macro	micro	macro	micro
BR	79.9	81.1	70.0	72.4	74.5	76.5	80.9	87.8	66.5	76.6	72.5	81.8
LP	77.7	79.1	70.5	73.4	73.7	76.1	75.1	83.3	64.2	75.0	69.0	78.9
CC	76.8	77.8	7.02	74.5	74.2	76.1	78.8	86.3	68.5	78.0	72.9	81.9
Rk	77.7	79.1	70.5	73.4	73.7	76.1	75.1	83.3	64.2	75.0	69.0	78.9
MLKNN	68.3	70.7	60.7	63.5	64.2	66.9	65.6	74.0	46.2	58.1	53.2	65.1
MARM	55.6	52.1	55.8	62.8	48.2	56.9	9.4	22.7	5.3	19	2.4	20.7
GF	70.7	73.0	62.4	65.5	66.1	69.0	65.7	74.1	46.3	58.2	53.3	65.2
GraphSAGE_mean	73.7	75.9	73.5	75.1	73.5	75.5	59.6	72.0	74.6	81.3	65.2	76.4
GraphSAGE_GCN	78.3	79.8	65.6	67.5	71.2	73.1	51.3	62.6	66.5	75.4	57.3	68.4
GraphSAGE_maxpool	76.8	78.6	71.6	73.3	74.1	75.8	60.0	72.0	77.0	83.4	66.4	77.3
GraphSAGE_meanpool	72.2	74.4	73.4	74.4	72.4	74.4	59.4	71.2	76.4	82.9	65.9	76.6
GraphSAGE_LSTM	70.5	73.2	73.8	75.4	71.9	74.3	58.1	69.9	75.4	82.4	64.9	75.6
MLGW-I_TRANS	80.4	81.7	74.0	76.2	76.8	78.8	78.0	85.5	75.0	84.3	76.0	84.9
MLGW-I_IND	80.4	81.5	73.9	76.3	76.7	78.8	78.0	85.5	74.9	84.3	76.0	84.9

Outline

- Introduction of Network Embedding
- Co-embedding of Attributed Network for User Profiling
- Dynamic Embedding for User Profiling
- Walking with Reinforcement for Semi-supervised Embedding learning from Attributed Network
- Active Heterogenous Network Embedding
- Future work Discussion

ActiveHNE: active heterogeneous network embedding

- Semi-supervised has promoted the learned embedding for multi-label classification, in ***homogenous*** network.
- **Heterogenous network**? However
  - Labels are difficult to obtain →

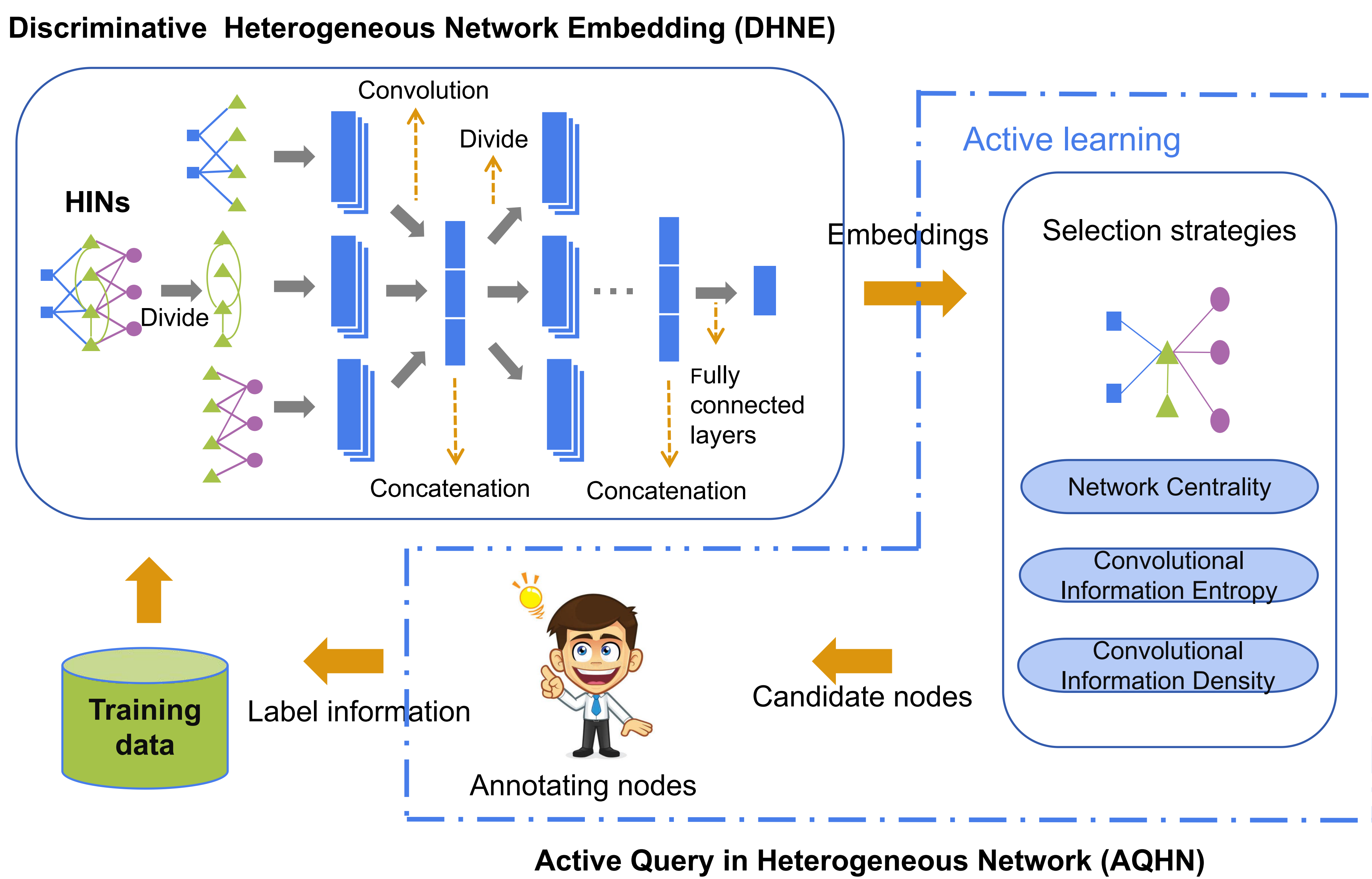
    Active Learning, iteratively query the most valuable batch of nodes to label and add for training
  - Nodes are in different types →

    Decompose the **original HIN** into

    homogeneous networks + bipartite networks

X Chen, G Yu, J Wang, C Domeniconi, Z Li, X Zhang. ActiveHNE: Active Heterogeneous Network Embedding
IJCAI 2019
Tuesday Aug 13, 10:50 - 12:20ML|AL - Active Learning 1

ActiveHNE architecture



Active Query in Heterogenous Network

Three active selection strategies based on **uncertainty and representation:**

- Network Centrality, e.g., Degree Centrality

$$\phi_{nc}(v_i) = |\mathcal{N}_i|,$$

- Convolutional Information Entropy (node representation uncertainty)

$$\phi_{cie}(v_i) = \sum_{v_j \in \{v_i \bigcup \mathcal{N}_i\}} w_j (-\sum_{c=1}^{C} \mathbf{F}_{jc} \log \mathbf{F}_{jc})$$

importance of a node

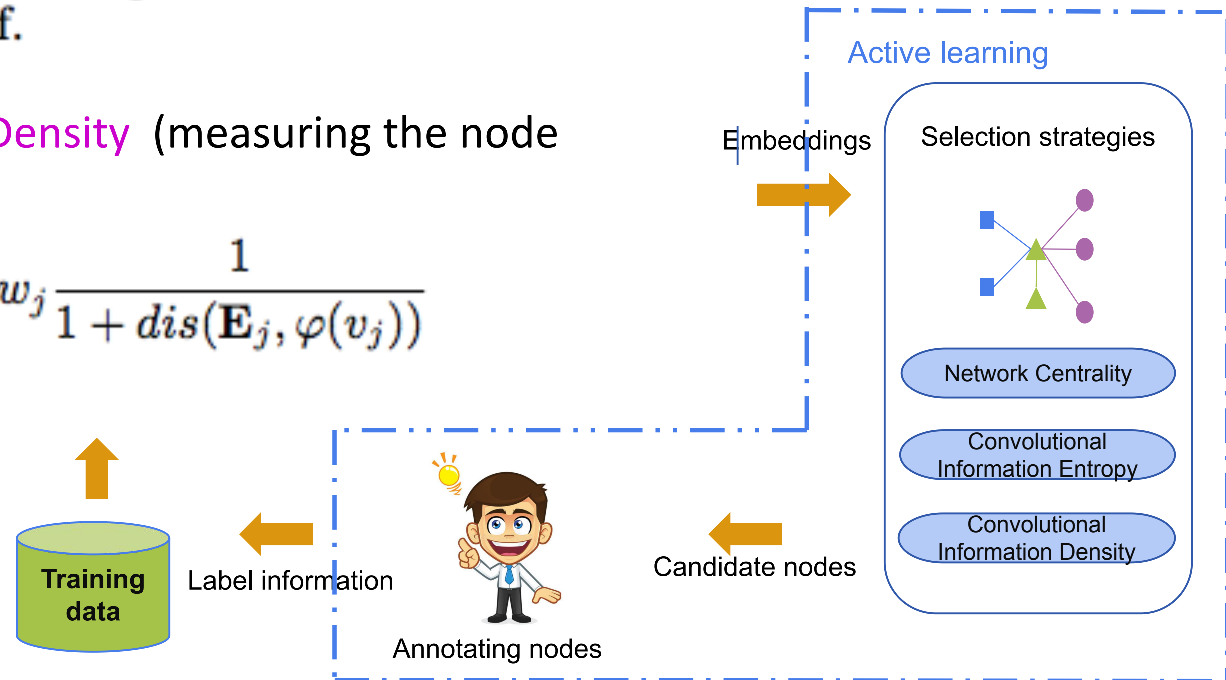
$$w_i = tanh(\frac{n_i}{N} + \frac{m_i}{V_T})$$

The uncertainty of v_i is a weighted sum of the uncertainties of its neighbors and itself.

- Convolutional Information Density (measuring the node representativeness)

$$\phi_{cid}(v_i) = \sum_{v_j \in \{v_i \bigcup \mathcal{N}_i\}} w_j \frac{1}{1 + dis(\mathbf{E}_j, \varphi(v_j))}$$

Which one or which ones to choose？



Active Query in Heterogeneous Network (AQHN)

Selecting candidate nodes

- Multi-armed bandit machine: Choose one or several arms to maximize the cumulative reward
- View each selection strategy as an arm, and approximate the importance of each strategy by estimating the expected reward (i.e., utility) of the corresponding arm
- The empirical reward of arm λ in iteration r is estimated as

$$\hat{\mu}_r(\lambda) = \frac{\Delta_r(\lambda)}{\Delta_r(\bigcup_{\lambda=1}^{\Lambda} \lambda)}$$

where

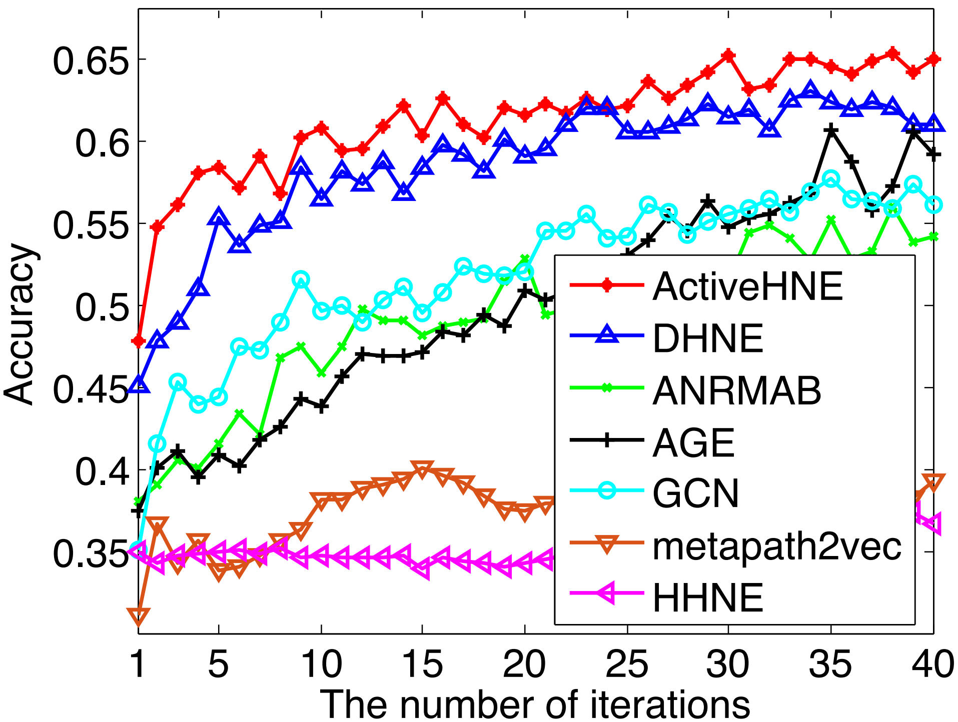
$$\Delta_r(\lambda) = \sum_{v_i \in \mathcal{Q}_r^{\lambda}} \sum_{v_j \in \mathcal{N}(v_i)} dis(\mathbf{E}_j^r, \mathbf{E}_j^{r-1})$$

is the local **embedding changes** caused by arm λ in iteration r.

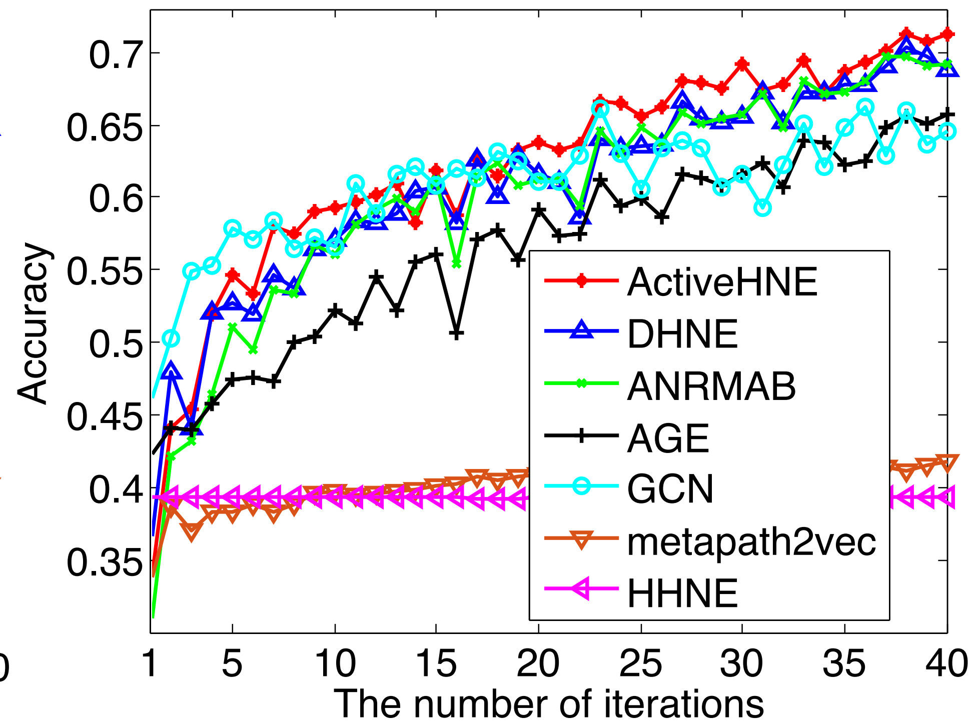
- Finally, select the top b nodes (from $\bigcup_{\lambda=1}^{\Lambda} \mathcal{C}_r^{\lambda}$) with the highest $\tilde{\mu}_r^*(v_i)$

$$\tilde{\mu}_r^*(v_i) = \sum_{\lambda=1}^{\Lambda} \tilde{\mu}_r(\lambda)(b - rank_r^{\lambda}(v_i))$$

Results, Active HNE vs baselines



MovieLens dataset: 9.7K movies, 12K writers, 4.9K directors, 0.6K users, and 1.5K tags, with a total of 140K links



Cora with 25K authors, 19K papers, and 12K terms, with 146K links

Outline

- Introduction of Network Embedding
- Co-embedding of Attributed Network for User Profiling
- Dynamic Embedding for User Profiling
- Walking with Reinforcement for Semi-supervised Embedding learning from Attributed Network
- Active Heterogenous Network Embedding
- Future work Discussion

Challenges and future directions

- Incompleteness and noise in graph
- Scalability on big graphs
- Attacks and Robustness of graph embedding
- Other applications, in biology, chemistry, etc

More of our network embedding and network mining work can be found at
https://mine.kaust.edu.sa/Pages/Category_Network.aspx

Our other papers at IJCAI 2019

1. **ActiveHNE: Active Heterogeneous Network Embedding**
   Xia Chen, Guoxian Yu, Jun Wang, Carlotta Domeniconi, Zhao Li, Xiangliang Zhang
   Presentation: **Tuesday Aug 13, 10:50 - 12:20ML**|AL - Active Learning 1
2. **Improving Cross-lingual Entity Alignment via Optimal Transport**
   Shichao Pei, Lu Yu, Xiangliang Zhang
   Presentation: **Tuesday Aug 13 16:30 - 18:00ML**|DM - Data Mining 3
3. **Multi-View Multiple Clustering**
   Shixin Yao, Guoxian Yu, Jun Wang, Carlotta Domeniconi, Xiangliang Zhang
   Presentation: **Friday Aug 16 09:30 - 10:30ML**|C2 - Clustering

Acknowledgement

My PhD students, visiting students and collaborators

- Basmah Altaf
- Uchenna Akujuob
- Shichao Pei
- Lu Yu
- Qianan Zhang
- Zhuo Yang
- Yufei Han
- Qiang Yang
- Hongyan Bao
- Peng Han
- Zaiqiao Meng
- Shangsong Liang
- Guoxian Yu
- Pinghui Wang

We are hiring!

My group website http://mine.kaust.edu.sa

Thank you for your attention!

Lab of Machine Intelligence and kNowledge Engineering (MINE): http://mine.kaust.edu.sa/

KAUST
King Abdullah University of Science and Technology