Hyper Edge-Based Embedding in Heterogeneous Information Networks

JIAWEI HAN
COMPUTER SCIENCE
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

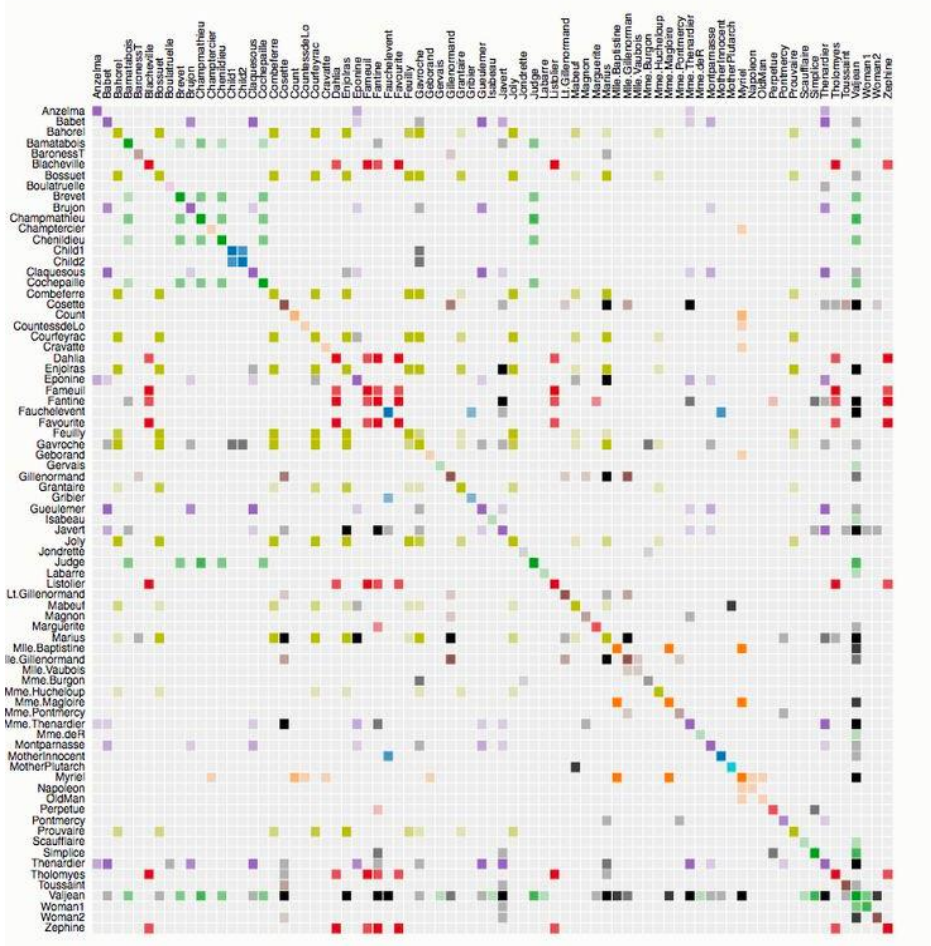
FEBRUARY 12, 2018

Outline

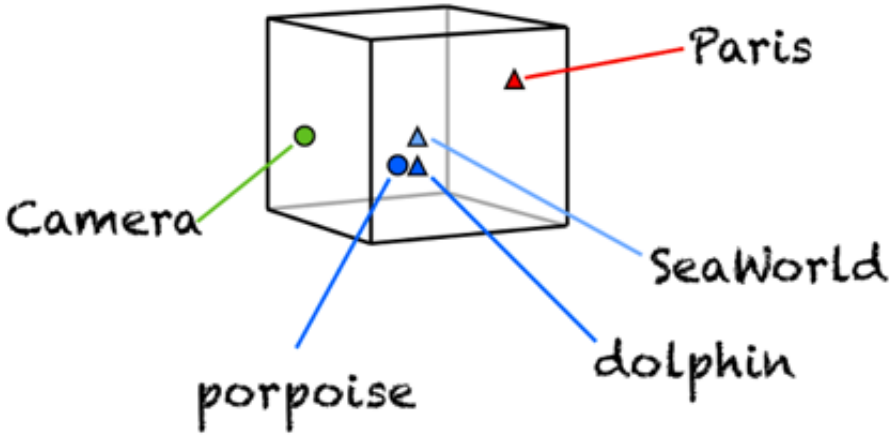
- Dimension Reduction: From Low-Rank Estimation vs. Embedding Learning
- Network Embedding for Homogeneous Networks
- Network Embedding for Heterogeneous Networks
- HEBE: Hyper-Edge Based Embedding in Heterogeneous Networks
- Aspect-Embedding in Heterogeneous Networks
- Locally-Trained Embedding for Expert-Finding in Heterogeneous Networks
- Summary and Discussions

Big Data Challenge: The Curse of High-Dimensionality

- Text: Word co-occurrence statistics matrix



- High-dimensionality:
  - There are over **171k** words in English language
- Redundancy:
  - Many words share similar semantic meanings
    - Sea, ocean, marine..

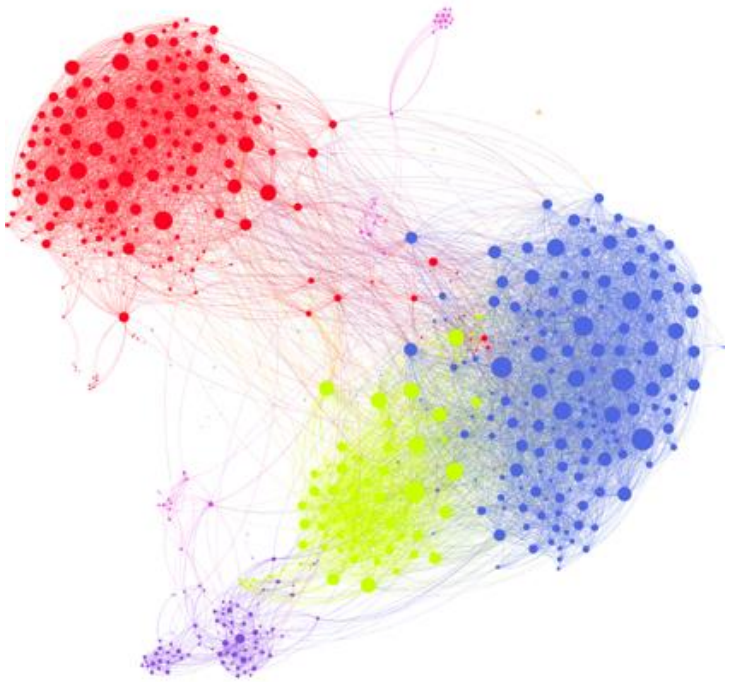


Multi-Genre Network Challenge: High-Dimensional Data too!

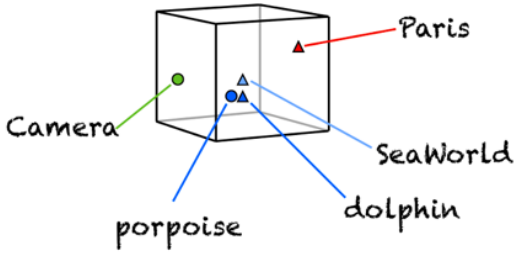
- Adjacency Matrix

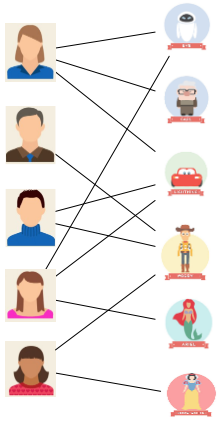
	1	2	3	4	5	6	7	8	9	10	…
1	0	1	1	1	1	0	0	1	0	0	…
2	1	0	1	1	0	0	1	0	0	0	…
3	1	1	0	1	0	0	0	0	1	0	…
4	1	1	1	0	0	0	0	0	0	0	…
5	1	0	0	0	0	0	0	0	0	0	…
6	1	0	0	0	0	0	0	0	0	0	…
7	1	0	0	0	1	0	0	0	0	0	…
8	1	1	1	1	0	0	0	0	0	0	…
9	0	0	1	0	0	0	0	0	0	1	…
10	0	0	1	0	0	0	0	1	0	1	…
11	0	0	0	0	0	0	0	0	1	1	…
12	0	1	0	0	0	0	0	0	1	1	…
13	1	0	0	0	0	0	0	0	1	1	…
14	0	0	1	0	0	1	1	1	0	1	…
15	0	0	0	0	0	1	1	1	1	0	…
…	…	…	…	…	…	…	…	…	…	…	…

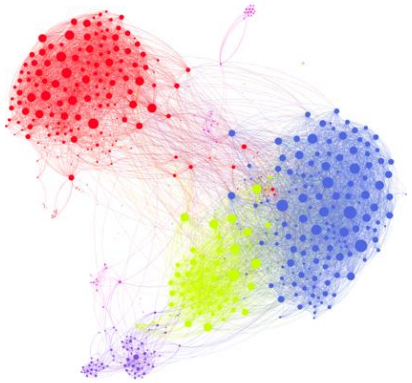
- High-dimension:
  - Facebook has 1860 Million monthly active users (Mar. 2017)
- Redundancy:
  - Users in the same cluster are likely to be connected



Solution to Data & Network Challenge: Dimension Reduction



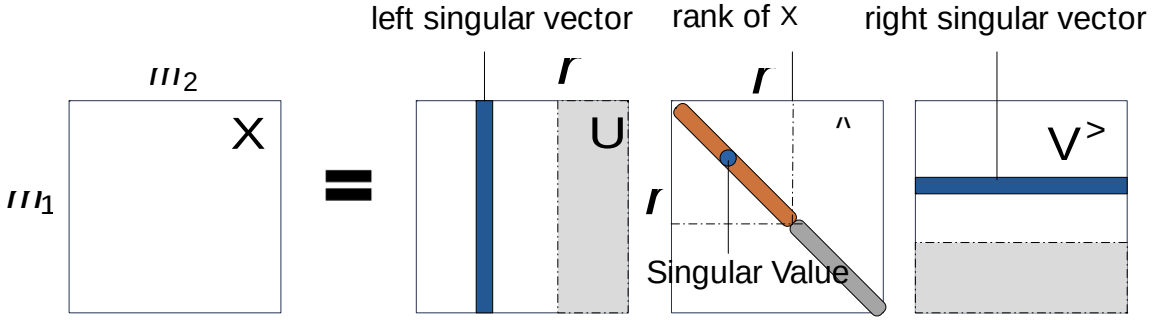




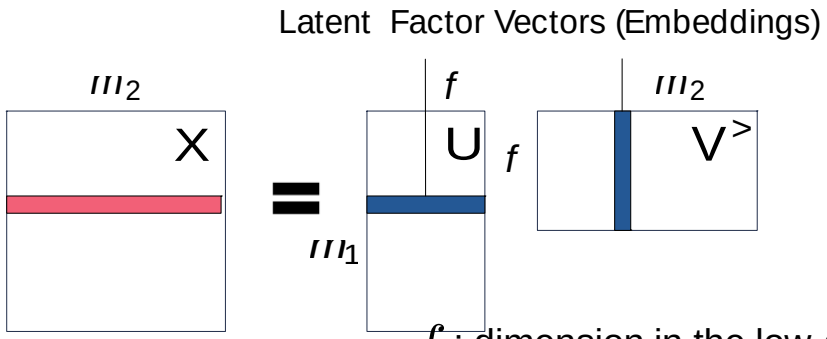
- Why Low-dimensional Space?
  - Visualization
  - Compression
  - Explanatory data analysis
  - Fill in (impute) missing entries (link/node prediction)
  - Classification and clustering
  - Identify / point
- How to automatically identify the lower-dimensional space that the high-dimensional data (approximately) lie in

Dimension Reduction Approaches: Low-Rank Estimation vs. Embedding Learning

$$\mathbf{X} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top$$



$$\mathbf{X} = \mathbf{U}\mathbf{V}^\top$$



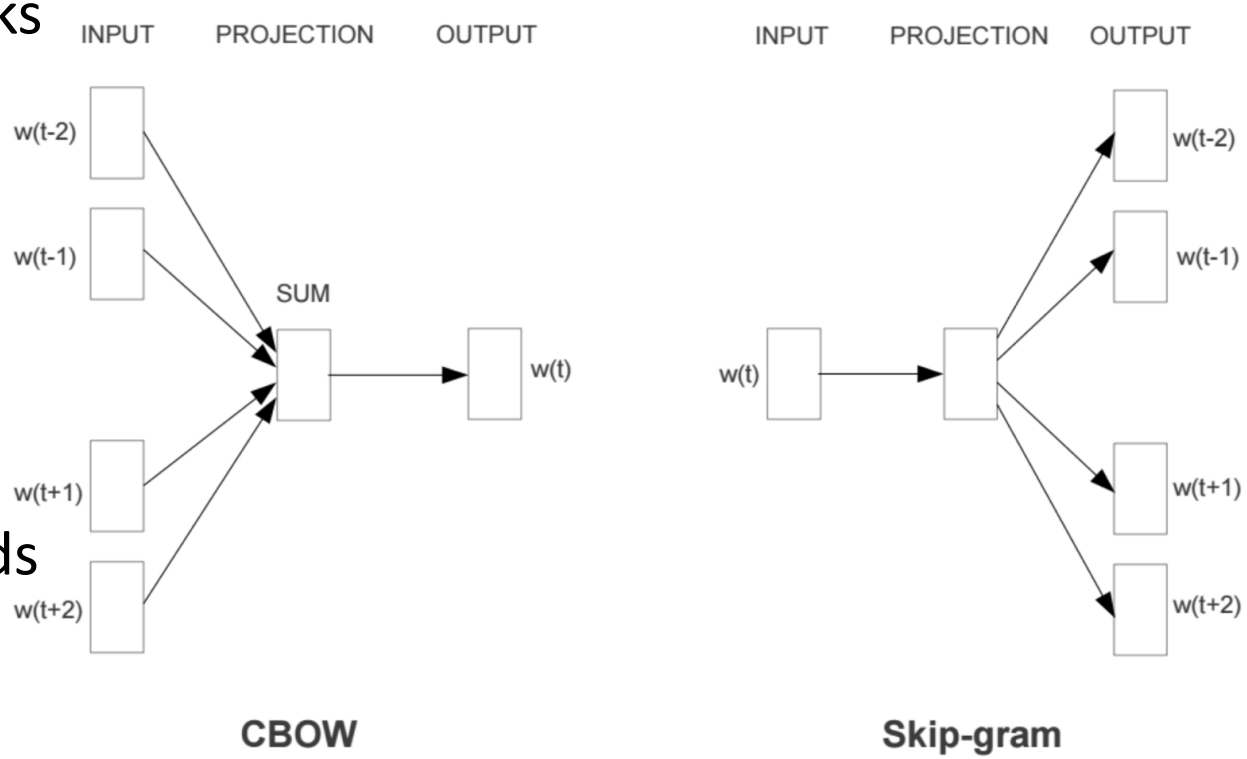
f : dimension in the low-dimensional space

- Low-rank estimation
  - Data recovery
  - Imposing low-rank assumption
    - Regularization
  - Low-dimension vector space
    - Singular vectors (U)
    - = r
  - Low-rank model

- Embedding Learning
  - Representation Learning
  - Project data into a low-dimensional space
  - Low-dimensional vector space
    - <u>Spanned</u> by columns of U
    - ≤ f
  - Generalized low-rank model

Word2Vec and Word Embedding

- Word2vec: created by T. Mikolov at Google (2013)
  - Input: a large corpus; output: a vector space, of 10^2 dimensions
  - Words sharing common contexts in close proximity in the vector space
- Embedding vectors created by Word2vec: better than LSA (Latent Semantic Analysis)
- Models: shallow, two-layer neural networks
- Two model architectures:
  - Continuous bag-of-words (CBOW)
    - Order does not matter, faster
  - Continuous skip-gram
    - Weigh nearby context words more heavily than more distant context words
    - Slower but better job for infrequent words



Outline

- Dimension Reduction: From Low-rank Estimation vs. Embedding Learning
- Network Embedding for Homogeneous Networks
- Network Embedding for Heterogeneous Networks
- HEBE: Hyper-Edge Based Embedding in Heterogeneous Networks
- Aspect-Embedding in Heterogeneous Networks
- Locally-Trained Embedding for Expert-Finding in Heterogeneous Networks
- Summary and Discussions

Embedding Networks into Low-Dimensional Vector Space

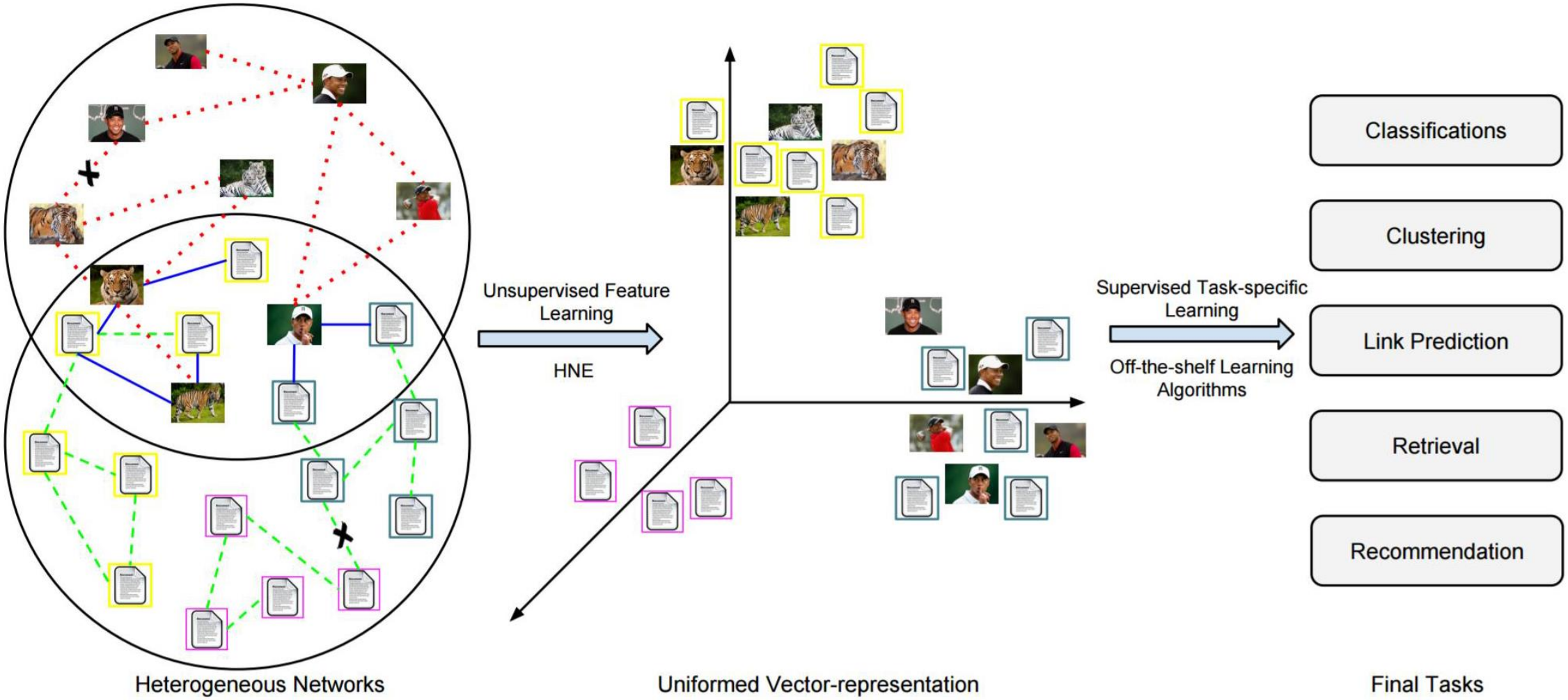


Figure 2: The flowchart of the proposed Heterogeneous Network Embedding (HNE) framework.

Recent Research Papers on Network Embedding (2013-2015)

Recent Research Papers on Network Embedding	Year
Distributed Large-scale Natural Graph Factorization	2013
Translating Embeddings for Modeling Multi-relational Data (TransE)	2013
DeepWalk: Online Learning of Social Representations	2014
Combining Two And Three-Way Embeddings Models for Link Prediction in Knowledge Bases (Tatec)	2015
Holographic Embeddings of Knowledge Graphs (HOLE) Diffusion Component Analysis: Unraveling	2015
Functional Topology in Biological Networks	2015
GraRep: Learning Graph Representations with Global Structural Information	2015
Deep Graph Kernels	2015
Heterogeneous Network Embedding via Deep Architectures	2015
PTE: Predictive Text Embedding through Large-scale Heterogeneous Text Networks	2015
LINE: Large-scale Information Network Embedding	**2015**

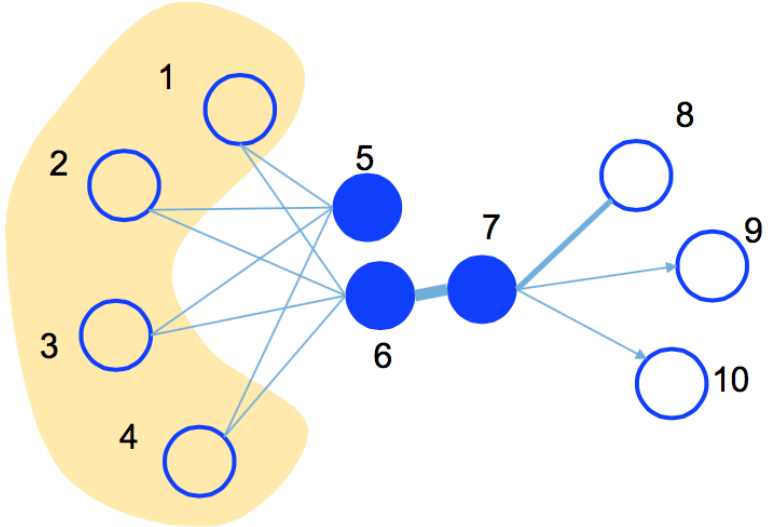
J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, and Q. Mei, “LINE: Large-scale information network embedding”, WWW'15 (cited 134 times)

Recent Research Papers on Network Embedding (2016)

Recent Research Papers on Network Embedding	Year
A General Framework for Content-enhanced Network Representation Learning (CENE)	2016
Variational Graph Auto-Encoders (VGAE)	2016
PROSNET: INTEGRATING HOMOLOGY WITH MOLECULAR NETWORKS FOR PROTEIN FUNCTION PREDICTION	2016
Large-Scale Embedding Learning in Heterogeneous Event Data (HEBE) Huan Gui, et al, ICDM 2016	**2016**
AFET: Automatic Fine-Grained Entity Typing by Hierarchical Partial-Label Embedding Xiang Ren, et al, EMNLP 2016	
Deep Neural Networks for Learning Graph Representations (DNGR)	2016
subgraph2vec: Learning Distributed Representations of Rooted Sub-graphs from Large Graphs 162	2016
Walklets: Multiscale Graph Embeddings for Interpretable Network Classification	2016
Asymmetric Transitivity Preserving Graph Embedding (HOPE)	2016
Label Noise Reduction in Entity Typing by Heterogeneous Partial-Label Embedding (PLE) Xiang Ren, et al, KDD 2016	
Semi-Supervised Classification with Graph Convolutional Networks (GCN)	2016
Revisiting Semi-Supervised Learning with Graph Embeddings (Planetoid)	2016
Structural Deep Network Embedding	2016
node2vec: Scalable Feature Learning for Networks	2016

LINE: Large-scale Information Network Embedding

- J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, and Q. Mei, "LINE: Large-scale information network embedding", WWW'15
- Nodes with strong ties turn to be similar
  - 1st order similarity
- Nodes share many neighbors turn to be similar
  - 2nd order similarity
- **Well-learnt embedding should preserve both 1st order and 2nd order similarity**



Nodes 6 & 7: high 1st order similarity

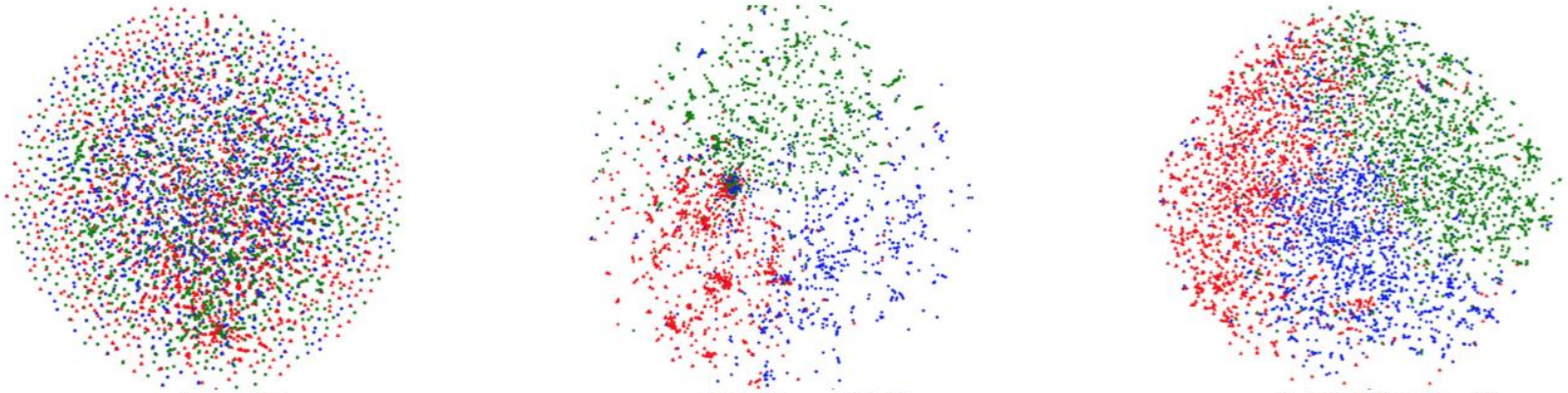
Nodes 5 & 6: high 2nd order similarity

Experiment Setup

- Dataset

	Language Network	Social Network		Citation Network	
Name	WIKIPEDIA	FLICKR	YOUTUBE	DBLP(AUTHORCITATION)	DBLP(PAPERCITATION)
Type	undirected,weighted	undirected,binary	undirected,binary	dircted,weighted	directed,binary
|V|	1,985,098	1,715,256	1,138,499	524,061	781,109
|E|	1,000,924,086	22,613,981	2,990,443	20,580,238	4,191,677
Avg. degree	504.22	26.37	5.25	78.54	10.73
#Labels	7	5	47	7	7
#train	70,000	75,958	31,703	20,684	10,398

- Task
  - Word analogy: Evaluated on Accuracy
  - Document classification: Evaluated on Macro-F1 Micro-F1
  - Vertex classification: Evaluated on Macro-F1 Micro-F1
- Result visualization



(a) GF (b) DeepWalk (c) LINE(2nd)

Results: Language Networks

- Word Analogy
  - GF (Graph Factorization) Ahmed et al., WWW2013)

Algorithm	Semantic (%)	Syntactic (%)	Overall (%)	Running time
GF	61.38	44.08	51.93	2.96h
DeepWalk	50.79	37.70	43.65	16.64h
SkipGram	69.14	57.94	63.02	2.82h
LINE-SGD(1st)	9.72	7.48	8.50	3.83h
LINE-SGD(2nd)	20.42	9.56	14.49	3.94h
LINE(1st)	58.08	49.42	53.35	2.44h
LINE(2nd)	**73.79**	**59.72**	**66.10**	2.55h

- Document Classification

Metric	Algorithm	10%	20%	30%	40%	50%	60%	70%	80%	90%
Micro-F1	GF	79.63	80.51	80.94	81.18	81.38	81.54	81.63	81.71	81.78
	DeepWalk	78.89	79.92	80.41	80.69	80.92	81.08	81.21	81.35	81.42
	SkipGram	79.84	80.82	81.28	81.57	81.71	81.87	81.98	82.05	82.09
	LINE-SGD(1st)	76.03	77.05	77.57	77.85	78.08	78.25	78.39	78.44	78.49
	LINE-SGD(2nd)	74.68	76.53	77.54	78.18	78.63	78.96	79.19	79.40	79.57
	LINE(1st)	79.67	80.55	80.94	81.24	81.40	81.52	81.61	81.69	81.67
	LINE(2nd)	79.93	80.90	81.31	81.63	81.80	81.91	82.00	82.11	82.17
	LINE(1st+2nd)	**81.04****	**82.08****	**82.58****	**82.93****	**83.16****	**83.37****	**83.52****	**83.63****	**83.74****
Macro-F1	GF	79.49	80.39	80.82	81.08	81.26	81.40	81.52	81.61	81.68
	DeepWalk	78.78	79.78	80.30	80.56	80.82	80.97	81.11	81.24	81.32
	SkipGram	79.74	80.71	81.15	81.46	81.63	81.78	81.88	81.98	82.01
	LINE-SGD(1st)	75.85	76.90	77.40	77.71	77.94	78.12	78.24	78.29	78.36
	LINE-SGD(2nd)	74.70	76.45	77.43	78.09	78.53	78.83	79.08	79.29	79.46
	LINE(1st)	79.54	80.44	80.82	81.13	81.29	81.43	81.51	81.60	81.59
	LINE(2nd)	79.82	80.81	81.22	81.52	81.71	81.82	81.92	82.00	82.07
	LINE(1st+2nd)	**80.94****	**81.99****	**82.49****	**82.83****	**83.07****	**83.29****	**83.42****	**83.55****	**83.66****

Significantly outperforms GF at the: ** 0.01 and * 0.05 level, paired t-test.

Results: Social Networks

- Flickr dataset

Metric	Algorithm	10%	20%	30%	40%	50%	60%	70%	80%	90%
Micro-F1	GF	53.23	53.68	53.98	54.14	54.32	54.38	54.43	54.50	54.48
	DeepWalk	60.38	60.77	60.90	61.05	61.13	61.18	61.19	61.29	61.22
	DeepWalk(256dim)	60.41	61.09	61.35	61.52	61.69	61.76	61.80	61.91	61.83
	LINE(1st)	63.27	63.69	63.82	63.92	63.96	64.03	64.06	64.17	64.10
	LINE(2nd)	62.83	63.24	63.34	63.44	63.55	63.55	63.59	63.66	63.69
	LINE(1st+2nd)	**63.20****	**63.97****	**64.25****	**64.39****	**64.53****	**64.55****	**64.61****	**64.75****	**64.74****
Macro-F1	GF	48.66	48.73	48.84	48.91	49.03	49.03	49.07	49.08	49.02
	DeepWalk	58.60	58.93	59.04	59.18	59.26	59.29	59.28	59.39	59.30
	DeepWalk(256dim)	59.00	59.59	59.80	59.94	60.09	60.17	60.18	60.27	60.18
	LINE(1st)	62.14	62.53	62.64	62.74	62.78	62.82	62.86	62.96	62.89
	LINE(2nd)	61.46	61.82	61.92	62.02	62.13	62.12	62.17	62.23	62.25
	LINE(1st+2nd)	**62.23****	**62.95****	**63.20****	**63.35****	**63.48****	**63.48****	**63.55****	**63.69****	**63.68****

Significantly outperforms DeepWalk at the: ** 0.01 and * 0.05 level, paired t-test.

- Youtube dataset

Metric	Algorithm	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%
Micro-F1	GF	25.43 (24.97)	26.16 (26.48)	26.60 (27.25)	26.91 (27.87)	27.32 (28.31)	27.61 (28.68)	27.88 (29.01)	28.13 (29.21)	28.30 (29.36)	28.51 (29.63)
	DeepWalk	39.68	41.78	42.78	43.55	43.96	44.31	44.61	44.89	45.06	45.23
	DeepWalk(256dim)	39.94	42.17	43.19	44.05	44.47	44.84	45.17	45.43	45.65	45.81
	LINE(1st)	35.43 (36.47)	38.08 (38.87)	39.33 (40.01)	40.21 (40.85)	40.77 (41.33)	41.24 (41.73)	41.53 (42.05)	41.89 (42.34)	42.07 (42.57)	42.21 (42.73)
	LINE(2nd)	32.98 (36.78)	36.70 (40.37)	38.93 (42.10)	40.26 (43.25)	41.08 (43.90)	41.79 (44.44)	42.28 (44.83)	42.70 (45.18)	43.04 (45.50)	43.34 (45.67)
	LINE(1st+2nd)	39.01* (**40.20**)	41.89 (**42.70**)	43.14 (**43.94****)	44.04 (**44.71****)	44.62 (**45.19****)	45.06 (**45.55****)	45.34 (**45.87****)	45.69** (**46.15****)	45.91** (**46.33****)	46.08** (**46.43****)
Macro-F1	GF	7.38 (11.01)	8.44 (13.55)	9.35 (14.93)	9.80 (15.90)	10.38 (16.45)	10.79 (16.93)	11.21 (17.38)	11.55 (17.64)	11.81 (17.80)	12.08 (18.09)
	DeepWalk	28.39	30.96	32.28	33.43	33.92	34.32	34.83	35.27	35.54	35.86
	DeepWalk (256dim)	28.95	31.79	33.16	34.42	34.93	35.44	35.99	36.41	36.78	37.11
	LINE(1st)	28.74 (29.40)	31.24 (31.75)	32.26 (32.74)	33.05 (33.41)	33.30 (33.70)	33.60 (33.99)	33.86 (34.26)	34.18 (34.52)	34.33 (34.77)	34.44 (34.92)
	LINE(2nd)	17.06 (22.18)	21.73 (27.25)	25.28 (29.87)	27.36 (31.88)	28.50 (32.86)	29.59 (33.73)	30.43 (34.50)	31.14 (35.15)	31.81 (35.76)	32.32 (36.19)
	LINE(1st+2nd)	**29.85** (29.24)	31.93 (**33.16****)	33.96 (**35.08****)	35.46** (**36.45****)	36.25** (**37.14****)	36.90** (**37.69****)	37.48** (**38.30****)	38.10** (**38.80****)	38.46** (**39.15****)	38.82** (**39.40****)

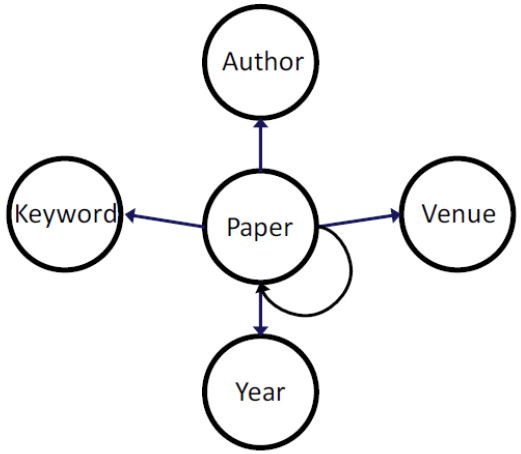
Significantly outperforms DeepWalk at the: ** 0.01 and * 0.05 level, paired t-test.

Outline

- Dimension Reduction: From Low-rank Estimation vs. Embedding Learning
- Network Embedding for Homogeneous Networks
- Network Embedding for Heterogeneous Networks
- HEBE: Hyper-Edge Based Embedding in Heterogeneous Networks
- Aspect-Embedding in Heterogeneous Networks
- Locally-Trained Embedding for Expert-Finding in Heterogeneous Networks
- Summary and Discussions

Task-Guided and Path-Augmented Heterogeneous Network Embedding

- T. Chen and Y. Sun, Task-guided and Path-augmented Heterogeneous Network Embedding for Author Identification, WSDM'17
- Given an anonymized paper (often: double-blind review), with
  - Venue (e.g., WSDM)
  - Year (e.g., 2017)
  - Keywords (e.g., "heterogeneous network embedding")
  - References (e.g., [Chen et al., IJCAI'16])
  - Can we predict its authors?
- Previous work on author identification: Feature engineering
- New approach: Heterogeneous Network Embedding
  - Embedding: automatically represent nodes into lower dimensional feature vectors
  - Heterogeneous network embedding: Key challenge—select the best type of info due to the heterogeneity of the network



Task-Guided Embedding

- Consider the ego-network of p:
  - $X_p = (X_p^1, X_p^2, \ldots, X_p^T)$,
  - T: # types of nodes associated with paper type
  - X_p^t: the set of nodes with type *t* associated with paper p
- u_a: embedding of author *a*
- u_n: embedding of node *n*
- V_p: embedding of paper *p*
  - *Weighted average of all the neighbors*
- The score function between *p* and *a* is:

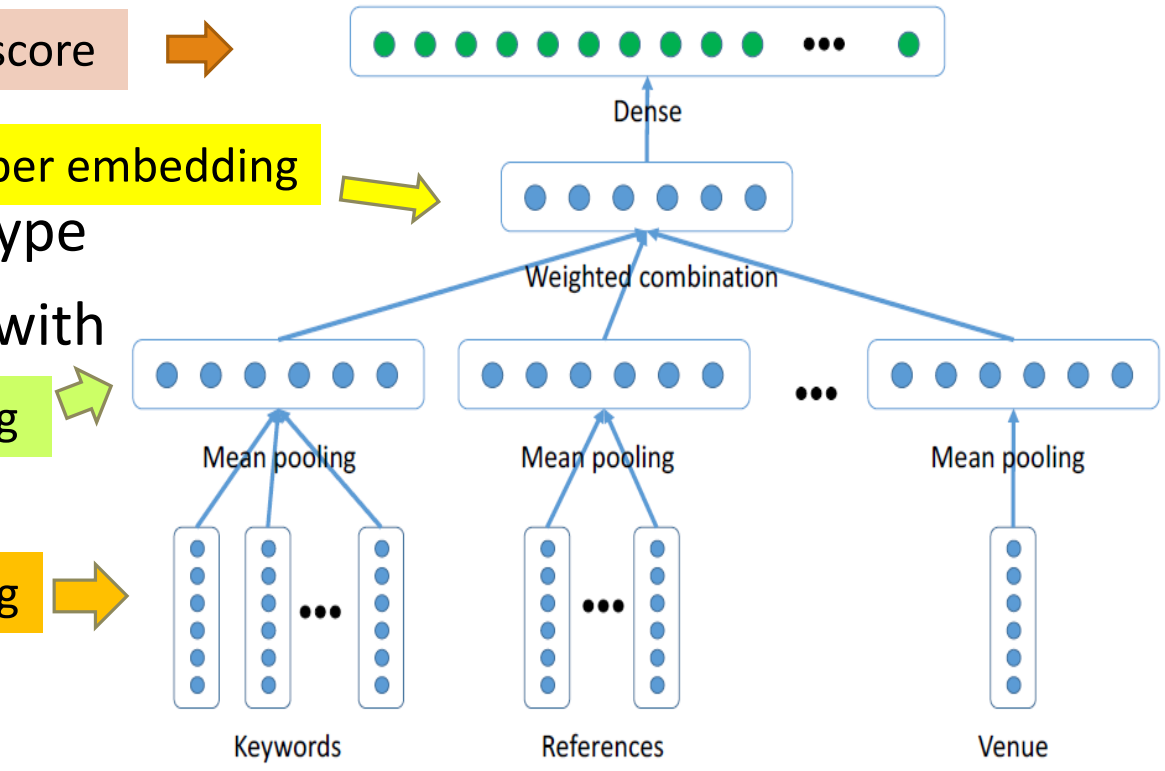
$$f(p,a) = u_a^T V_p = u_a^T \left(\sum_t w_t V_p^{(t)} \right)$$

$$= u_a^T \left(\sum_t w_t \sum_{n \in X_p^{(t)}} u_n / |X_p^{(t)}| \right)$$

- Ranking-based objective: maximize the difference between authors b and a:

$$\max\left(0, f(p,b) - f(p,a) + \xi\right)$$

Soft hinge loss



The embedding architecture for author identification

Identification of Anonymous Authors: Experiments

- Dataset:
  - AMiner Citation data set
  - Papers before 2012 are used in training, and papers on and after 2012 are used as test

Table 1 : Node statistics

	Paper	Author	keyword	Venue	Year
Train	1.6M	1M	4M	7K	60
Test	34K	62K	42K	1K	2

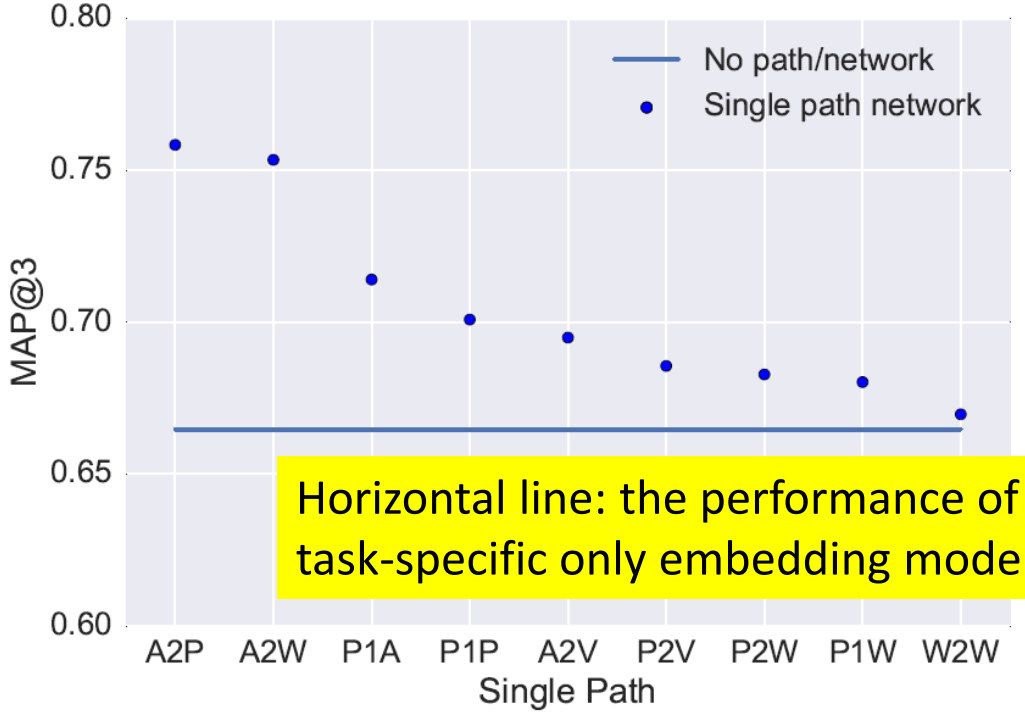
Table 3 : Length-2 link statistics

A-P-A	A-P-P	A-P-V	A-P-W	A-P-Y	P-P-V	P-P-W	V-P-W	W-P-W	Y-P-W
17M	18M	4M	38M	4M	3M	27M	12M	118M	12M

- Baselines
  - Supervised feature-based baselines (i.e. LR, SVM, RF, LambdaMart)
    - Manually crafted features
  - Task-specific embedding
  - Network-general embedding
  - Pre-training + Task-specific embedding
    - Take general embedding as initialization of task-specific embedding

Which Meta-Paths Are Selected?

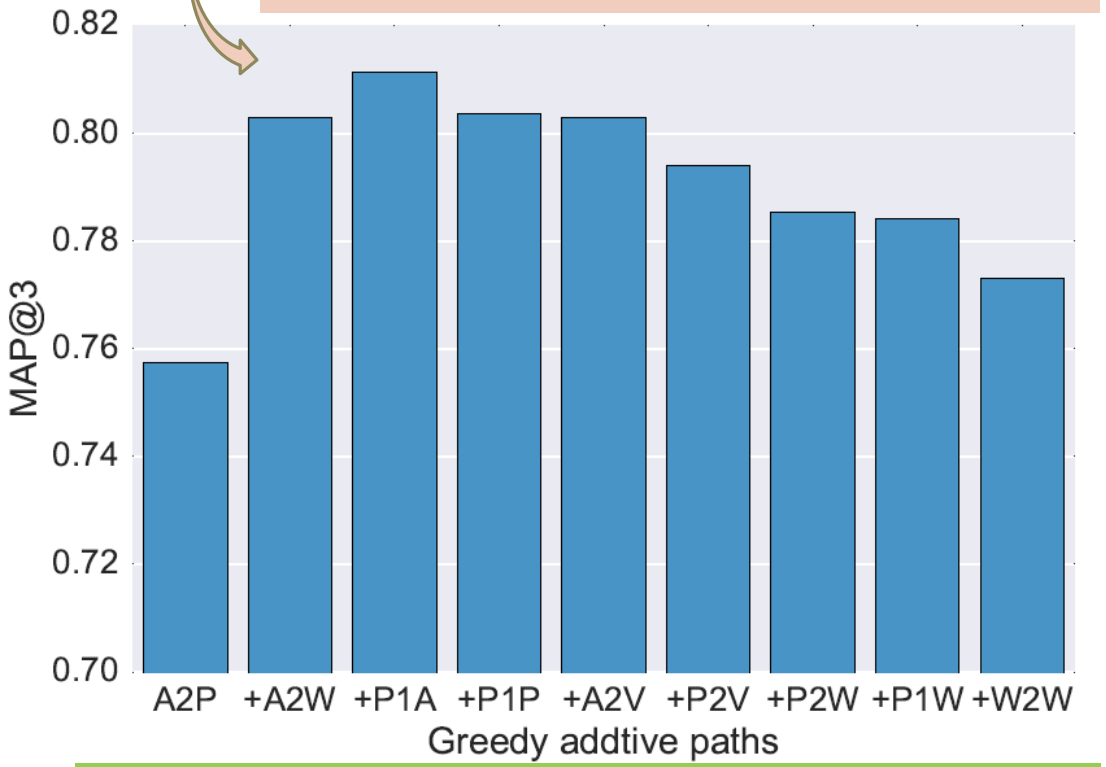
- A-P-P: author *write* paper *cite* paper
- A-P-W: author *write* paper *contain* keyword
- P-A: paper *written-by* author



Horizontal line: the performance of task-specific only embedding model

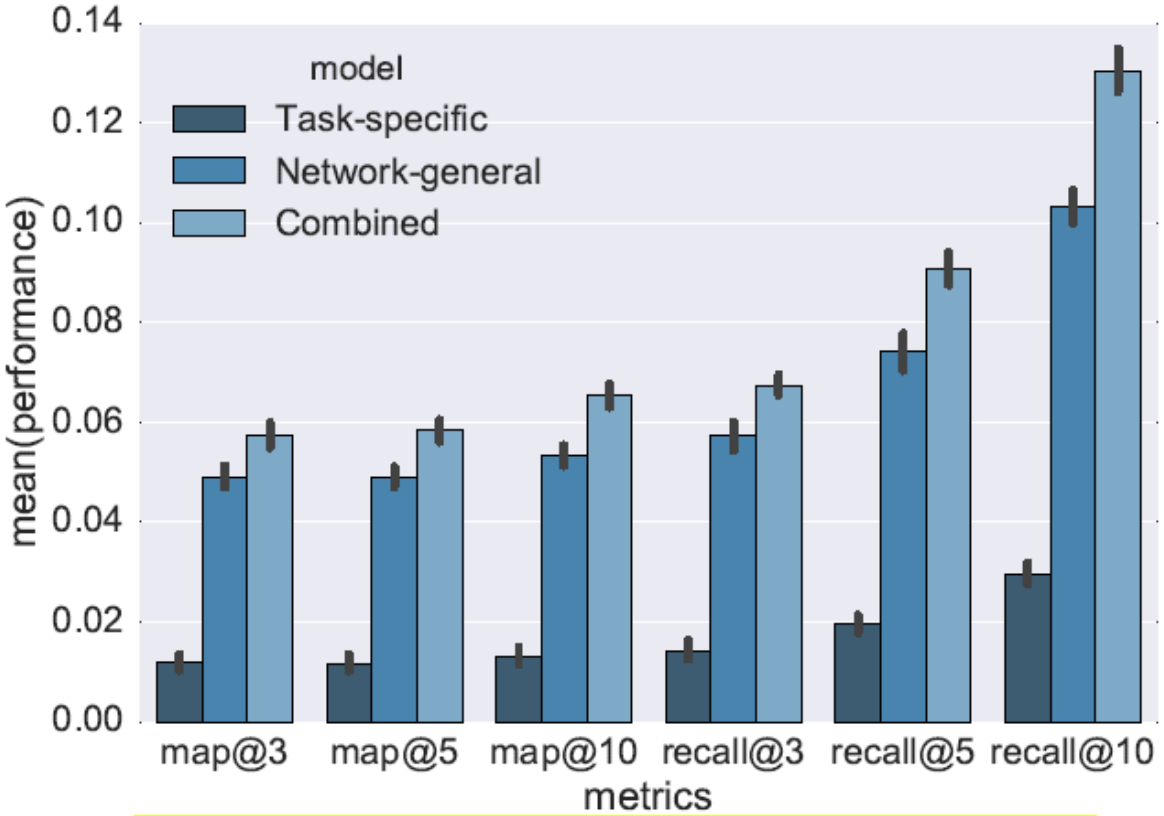
- Paths are sorted according to their performance
- Only paths that can help improve the author identification task are shown

- The first several paths are most relevant and helpful
- Latter ones can be harmful to use in network-general embedding



The performance of the combined model when meta-paths are added gradually

The Real Game and Case Study



Treat all the authors as candidates

Top ranked authors for queried paper

(a) "Active learning for networked data based on non-progressive diffusion model"

Ground-truth	Task-specific	Network-general	Combined
Z. Yang	L. Yu	J. Leskovec	**J. Tang**
J. Tang	Y. Gao	A. Ahmed	H. Liu
B. Xu	J. Wang	L. Getoor	Y. Guo
C. Xing	H. Liu	S.-D. Lin	X. Shi
	Y. Gao	D. Chakrabarti	W. Fan
	Z. Wang	P. Melville	B. Zhang
	Z. Zhang	T. Eliassi-Rad	S.-D. Lin
	J. Zhu	G. Lebanon	H. Zha
	Y. Ye	Y. Sun	L. H. Ungar
	R. Pan	L. H. Ungar	C. Wang

(b) "CatchSync: catching sync. behavior in large directed graphs"

Ground-truth	Task-specific	Network-general	Combined
M. Jiang	H. Wang	L. Akoglu	**C. Faloutsos**
P. Cui	H. Tong	T. Eliassi-Rad	A. Gionis
A. Beutel	**C. Faloutsos**	U. Kang	L. Akoglu
C. Faloutsos	D. Chakrabarti	H. Tong	J. Kleinberg
S. Yang	H. Yang	D. Chakrabarti	J. Leskovec
	G. Konidaris	A. Gionis	D. Chakrabarti
	I. Stanton	X. Yan	A. X. Zheng
	C. Wang	**C. Faloutsos**	T. Eliassi-Rad
	Y. Yang	J. Leskovec	U. Kang
	S. Kale	C. Tsourakakis	H. Tong

Outline

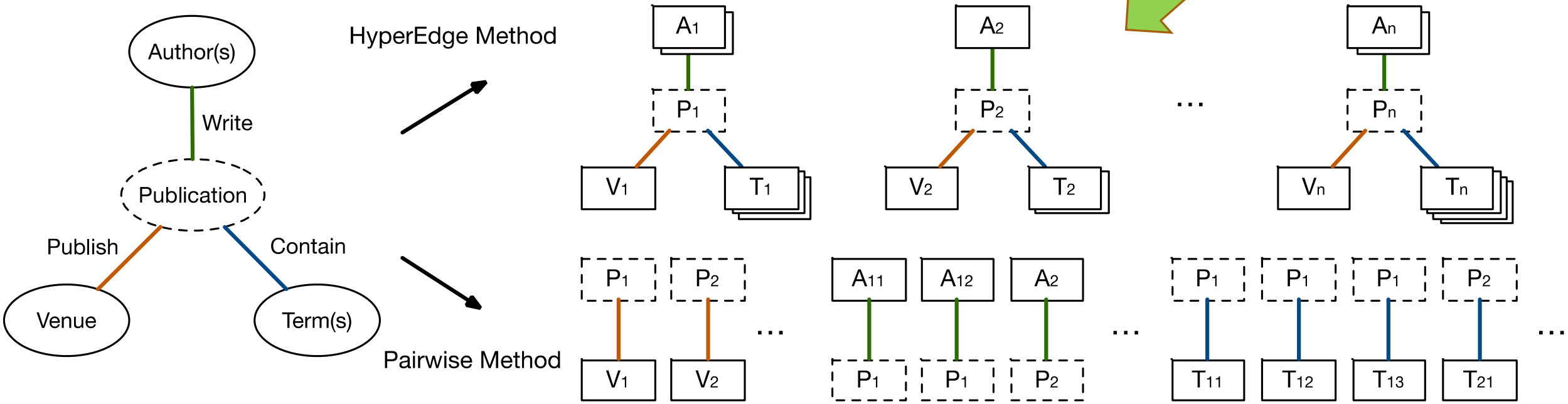
- Dimension Reduction: From Low-rank Estimation vs. Embedding Learning
- Network Embedding for Homogeneous Networks
- Network Embedding for Heterogeneous Networks
- HEBE: Hyper-Edge Based Embedding in Heterogeneous Networks
- Aspect-Embedding in Heterogeneous Networks
- Locally-Trained Embedding for Expert-Finding in Heterogeneous Networks
- Summary and Discussions

Large-Scale Embedding Learning in Heterogeneous Events (HEBE)

- Embedding in Heterogeneous Information Networks
  - Multiple types of **Objects**
  - Multiple types of **Interactions**
  - How to preserve information among objects?
    - Event: Interactions that happen simultaneously

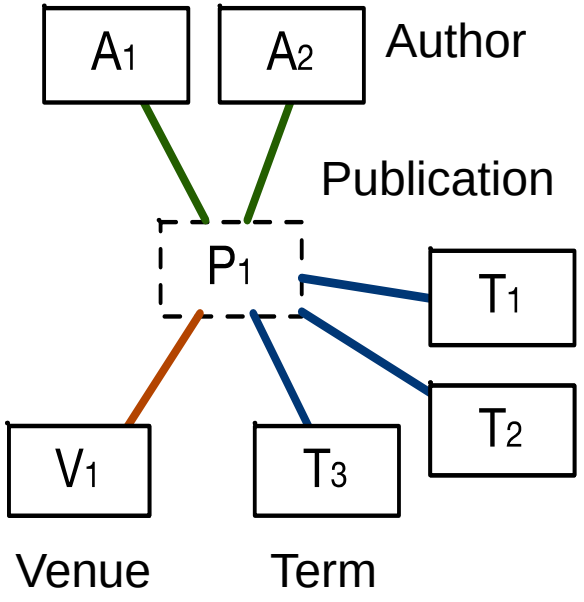
H. Gui, J. Liu, F. Tao, M. Jiang, B. Norick, L. Kaplan, J. Han, "Large-Scale Embedding Learning in Heterogeneous Event Data", ICDM'16 + IEEE TKDE'17

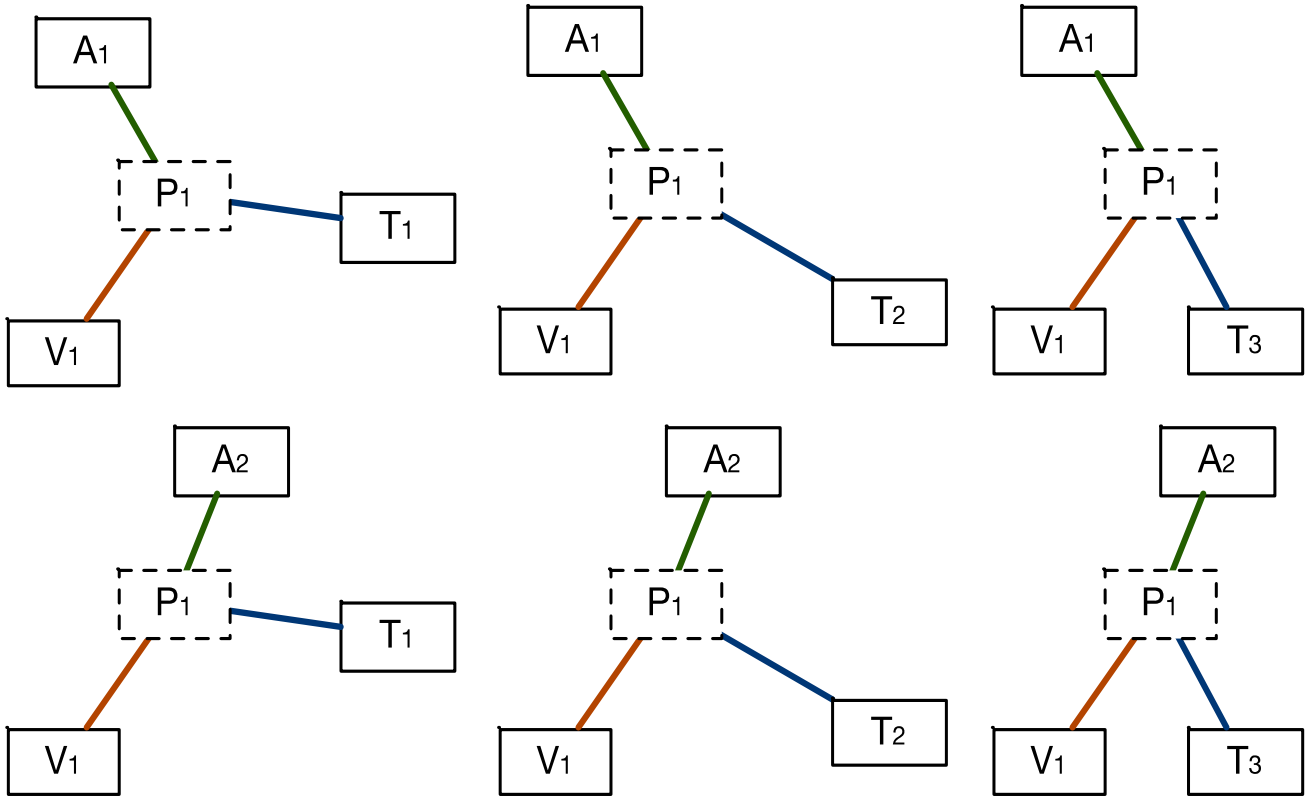
Hyper-edge embedding is better than pairwise embedding



SubEvent Sampling

- More than one object for each object type
  - Sample object

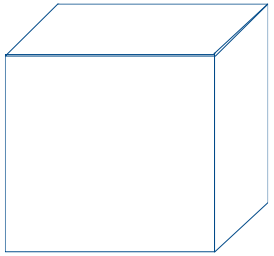




Hyper-Edge Based Embedding Framework (I)

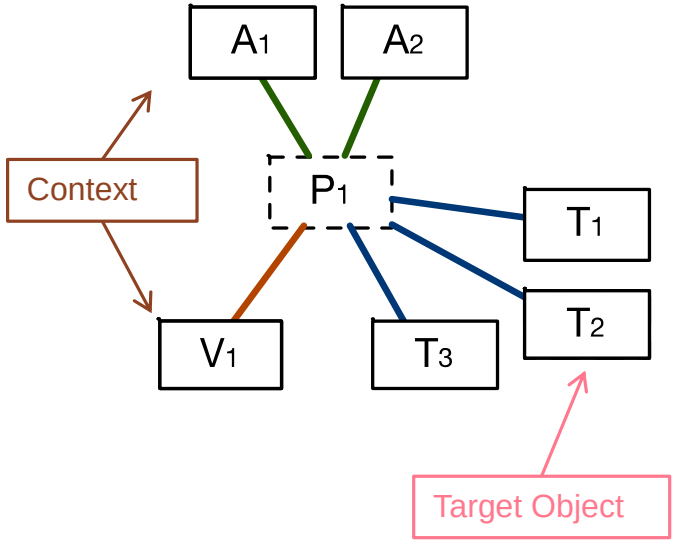
- For object prediction
  - Embedding Learning Model

$$\{\hat{\mathbf{U}}_t\}_{t=1}^{T} = \underset{\mathbf{U}_t \in \mathbb{R}^{m_t \times k}, t=1,\ldots,T}{\operatorname{argmin}} \mathcal{D}\big(\mathcal{Z}_\Omega(\mathbf{M}), \mathcal{Z}_\Omega(f(\mathbf{U}_1, \ldots, \mathbf{U}_t))\big)$$



- Object Driven
  - Empirical conditional probability

$$\mathcal{Z}_\Omega(\mathbf{M}_i) = \hat{\mathbb{P}}(T_2 | A_1, V_1)$$



  - Model conditional probability via Softmax

Scoring Function

Context

$$\mathbb{P}_o(u|C) = \frac{\exp\left(S(u, C)\right)}{\sum_{v \in X_1} \exp\left(S(v, C)\right)}$$

Target Object

Object Set

Alternative Object

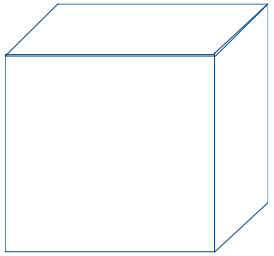
$$\mathcal{Z}_\Omega(f(\mathbf{U}_1, \ldots, \mathbf{U}_t))$$

  - Distance Measure: KL-divergence
    - Measure distance between conditional probability distribution

Hyper-Edge Based Embedding Framework (II)

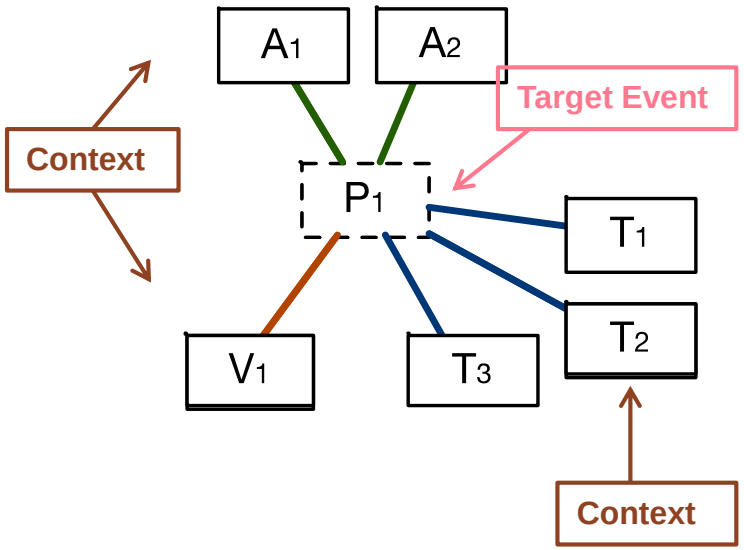
- For event prediction
  - Embedding Learning Model

$$\{\hat{\mathbf{U}}_t\}_{t=1}^{T} = \underset{\mathbf{U}_t \in \mathbb{R}^{m_t \times k}, t=1,\ldots,T}{\operatorname{argmin}} \mathcal{D}\big(\mathcal{Z}_\Omega(\mathbf{M}), \mathcal{Z}_\Omega(f(\mathbf{U}_1, \ldots, \mathbf{U}_t))\big)$$



- Event Driven
  - Empirical conditional probability

$$\mathcal{Z}_\Omega(\mathbf{M}_i) = \hat{\mathbb{P}}(P_1 | A_1, V_1, T_2)$$



  - Model conditional probability via Softmax

$$\mathbb{P}_e(q_i | V_i) = \frac{\exp(S(q_i, V_i))}{\sum_{q_j \in \mathcal{Q}} \exp(S(q_j, V_i))}$$

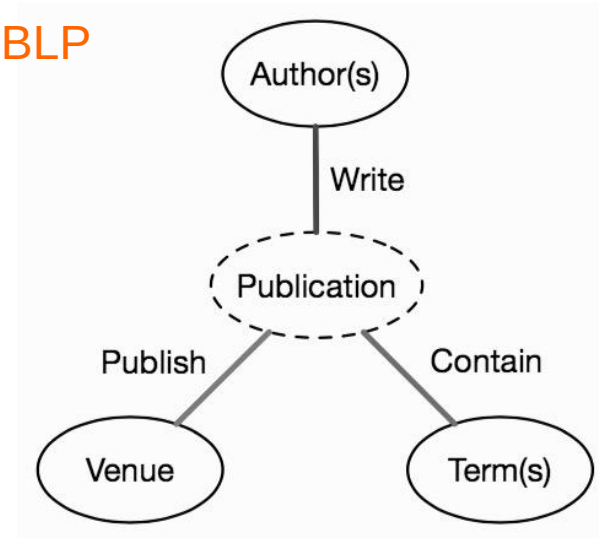
Scoring Function; SubEvent object set; Target Event; Event Set; Alternative Event

$$\mathcal{Z}_\Omega(f(\mathbf{U}_1, \ldots, \mathbf{U}_t))$$

  - Distance Measure: KL-divergence

Experiments: Dataset Statistics

DBLP



Yelp

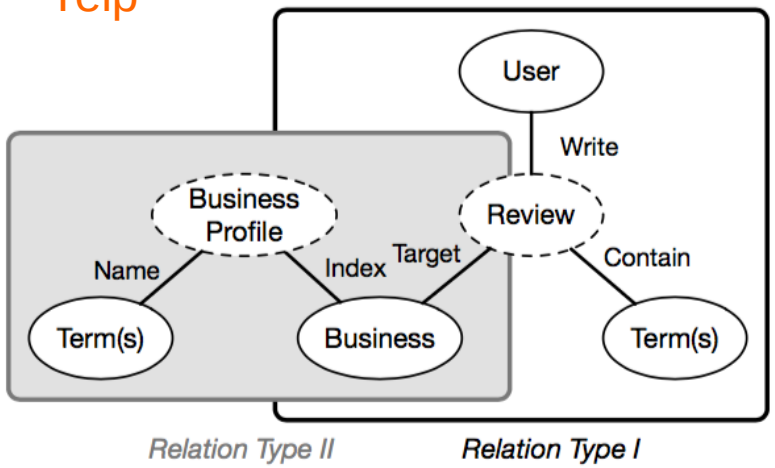


Table 1: Number of entities for DBLP and Yelp.

DBLP	Author	Term	Venue	Paper
	209,679	165,657	7953	1,938,912
Yelp	Business	Term_r	Term_p	Review
	12,241	130,259	6,709	905,658

Experiments: Classification Results

- DBLP: Author in four research groups/areas
- Yelp: Restaurants in eleven cuisine categories
- HEBE: Hyper-Edge Based Embedding
  - Gives better classification accuracy, more robust to data sparsity

	Research Group		Research Area		Restaurant Type	
Method	Acc.	AUC	Acc.	AUC	Acc.	AUC
SVD	81.03	0.7137	83.27	0.5720	74.09	0.7147
NSVD	72.41	0.6958	89.75	0.6271	66.45	0.6244
PPMI	70.69	0.7513	90.22	0.7450	82.82	0.6504
NMF	73.28	0.6210	75.69	0.5798	79.64	0.7955
NNMF	72.41	0.7223	88.31	0.7665	72.00	0.7328
LINE	78.45	0.5607	79.48	0.5565	79.82	0.6378
PTE	**87.93**	0.7235	90.27	0.6646	81.91	0.7195
HEBE-PO	84.48	0.7957	**92.18**	0.7905	**88.00**	**0.8961**
HEBE-PE	87.07	**0.8207**	91.66	**0.8417**	87.27	0.8826

Experiments: Data Sparsity (DBLP)

- HEBE is more robust to data sparsity
- Density Measure: Averaged number of publications each author has

Sampling %.	1%		5%		10%		20%		30%		50%	
Density Measure	1.264		2.028		2.882		4.595		6.400		10.315	
Method	Acc.	AUC	Acc.	AUC	Acc.	AUC	Acc.	AUC	Acc.	AUC	Acc.	AUC
Research Group												
SVD	38.46	0.5602	66.67	0.6169	65.59	0.6481	72.55	0.6494	72.86	0.6720	77.28	0.6924
NSVD	43.59	0.5504	58.73	0.5919	68.82	0.6330	70.59	0.6345	72.64	0.6517	74.55	0.6790
PPMI	46.15	0.5502	60.32	0.5993	76.34	0.6557	71.57	0.6703	72.97	0.6792	74.55	0.7192
NMF	41.03	0.5583	57.14	0.5989	56.99	0.5874	54.90	0.6009	66.96	0.5950	70.91	0.6120
NNMF	46.15	0.5462	55.56	0.6601	60.22	0.6806	75.49	0.7167	70.55	0.7197	71.82	0.7294
LINE	**56.41**	0.6004	66.67	0.6254	72.04	0.5877	77.45	0.5619	77.86	0.5669	85.45	0.5871
PTE	**56.41**	0.6190	69.84	0.6727	76.34	0.6434	84.31	0.6778	**85.94**	0.7034	**88.18**	0.6783
HEBE-PO	53.85	0.6034	66.67	0.7082	72.04	0.7151	74.51	0.7515	75.55	0.7640	82.73	0.7841
HEBE-PE	**56.41**	**0.6547**	**73.02**	**0.7434**	**83.87**	**0.7749**	**85.29**	**0.8221**	84.13	**0.8220**	**88.18**	**0.8316**
Research Area												
SVD	47.88	0.5162	62.47	0.5337	66.27	0.5411	71.66	0.5516	75.47	0.5551	79.15	0.5644
NSVD	52.39	0.5076	66.21	0.5004	72.15	0.5021	77.91	0.5157	80.13	0.5299	85.23	0.5600
PPMI	51.67	0.5063	68.00	0.5092	72.66	0.5180	78.15	0.5395	80.59	0.5669	85.91	0.6203
NMF	43.37	0.5143	53.54	0.5329	59.30	0.5391	63.63	0.5493	68.01	0.5560	70.72	0.5637
NNMF	50.50	0.5303	62.50	0.5773	67.73	0.6206	72.37	0.6486	76.51	0.6807	82.91	0.7594
LINE	57.17	0.5552	69.83	0.5764	72.15	0.5716	74.89	0.5501	74.53	0.5339	80.82	0.5822
PTE	53.29	0.5291	**71.54**	0.5858	73.95	0.5782	79.03	0.6015	82.68	0.6356	86.80	0.6340
HEBE-PO	**57.53**	**0.5635**	69.71	0.6108	74.91	0.6798	80.26	0.7199	81.66	0.7293	86.17	0.7817
HEBE-PE	54.64	0.5500	71.09	**0.6282**	**75.90**	**0.6834**	**81.64**	**0.7405**	**83.94**	**0.7645**	**87.84**	**0.8075**

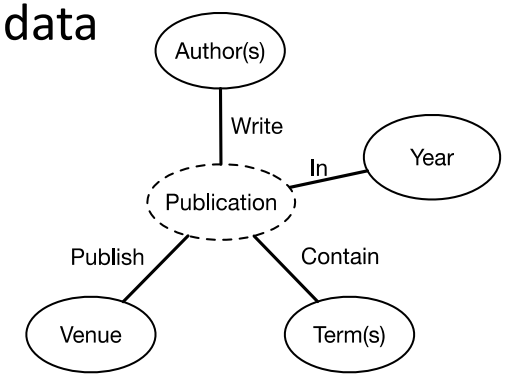
Experiments: Data Sparsity (Yelp)

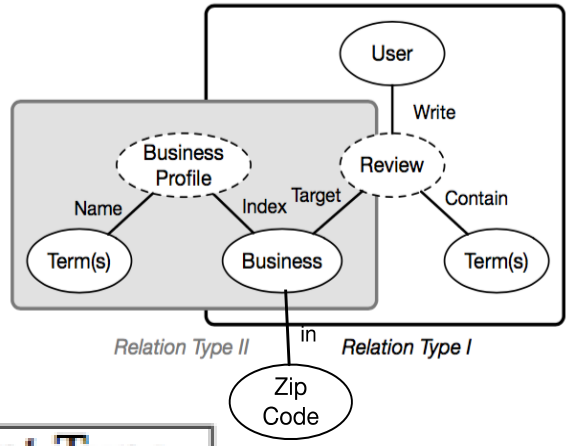
- HBHE is more robust to data sparsity
- **Density Measure**: Averaged number of reviews each business has

Sampling %.	1%		5%		10%		20%		30%		50%	
Density Measure	1.963		4.791		8.155		15.09		22.32		37.01	
Method	Acc.	AUC	Acc.	AUC	Acc.	AUC	Acc.	AUC	Acc.	AUC	Acc.	AUC
SVD	64.12	0.6133	70.85	0.6786	73.44	0.7001	73.98	0.7100	73.82	0.7121	74.82	0.7134
NSVD	62.07	0.6081	63.36	0.6236	65.17	0.6308	66.97	0.6275	67.00	0.6280	67.36	0.6259
PPMI	59.35	0.5561	65.01	0.5484	69.94	0.5626	75.43	0.5824	78.55	0.6089	80.55	0.6253
NMF	63.61	0.6790	71.23	0.7381	75.09	0.7594	76.34	0.7877	78.09	0.7907	78.18	0.7991
NNMF	60.71	0.6710	66.76	0.7022	68.47	0.7082	70.79	0.7213	70.73	0.7297	70.73	0.7312
LINE	60.88	0.5337	71.72	0.5367	77.32	0.5689	80.71	0.6665	80.91	0.6789	81.27	0.6833
PTE	64.29	0.6315	72.89	0.6758	76.28	0.6993	79.25	0.7163	81.00	0.7043	80.91	0.7266
HEBE-PO	71.09	0.7576	79.01	0.8316	82.63	0.8621	85.08	**0.8825**	**86.36**	**0.8845**	**86.82**	**0.8938**
HEBE-PE	**73.30**	**0.7747**	**79.69**	**0.8434**	**83.06**	**0.8746**	**85.44**	0.8779	85.82	0.8765	86.36	0.8862

Experiments: Noise Objects

- HBBE is more robust to noise in the data





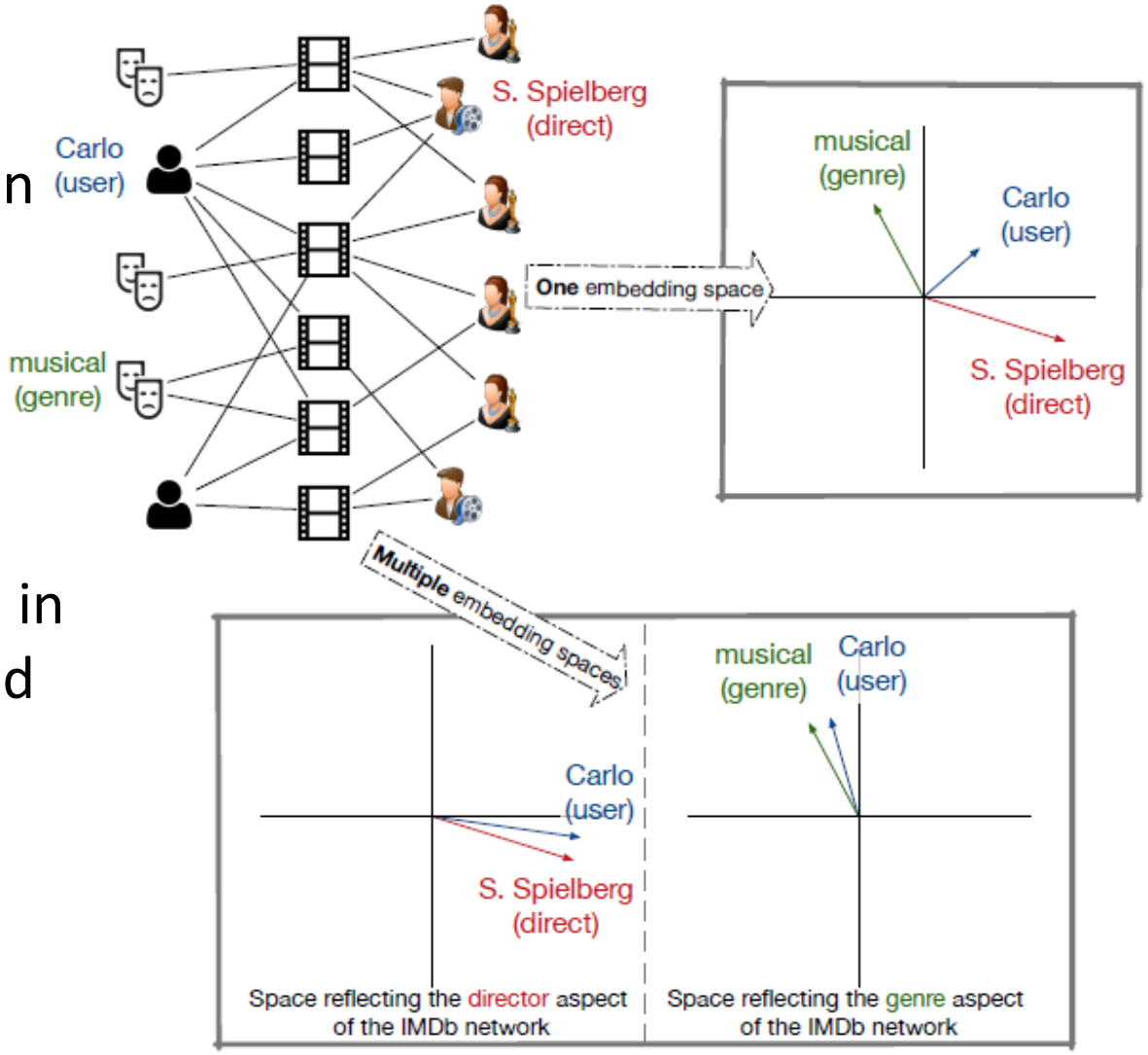
	Research Group		Research Area		Restaurant Type	
Method	Acc.	AUC	Acc.	AUC	Acc.	AUC
SVD	78.03	0.6846	80.10	0.5374	67.73	0.6902
NSVD	70.69	0.6668	87.48	0.6112	48.81	0.6138
PPMI	68.09	0.7175	88.99	0.7162	81.09	0.6892
NMF	72.73	0.6121	71.96	0.5635	67.00	0.7469
NNMF	71.38	0.6823	86.12	0.7411	43.45	0.6142
LINE	80.17	0.5465	78.94	0.5425	76.09	0.6035
PTE	85.34	0.6297	89.83	0.5873	75.18	0.6702
HEBE-PO	76.72	0.7582	89.11	0.7614	85.91	0.8296
HEBE-PE	**85.34**	**0.8214**	**91.26**	**0.8425**	**86.73**	**0.8834**

Outline

- Dimension Reduction: From Low-rank Estimation vs. Embedding Learning
- Network Embedding for Homogeneous Networks
- Network Embedding for Heterogeneous Networks
- HEBE: Hyper-Edge Based Embedding in Heterogeneous Networks
- Aspect-Embedding in Heterogeneous Networks
- Locally-Trained Embedding for Expert-Finding in Heterogeneous Networks
- Summary and Discussions

AspEm: Aspect Embedding in Heterogeneous Networks

- Y. Shi, Huan Gui, Qi Zhu, L. Kaplan, J. Han, "AspEm: Large-Scale Embedding Learning from Aspects in Heterogeneous Information Networks", SDM 2018
- Typed edges may not fully align with each other
  - Like movie, why? Director vs. genre
- AspEm: Preserve the semantic information in heterogeneous Information networks based on multiple aspects
  - Embedding on each aspect individually
- AspEm outperforms existing network embedding learning methods



AspEm Captures More Semantic Info. in Heter. Info. Nets

Classification accuracy on DBLP-group, DBLP-area, and IMDb using LR and SVM as classifiers

Dataset/task	DBLP-group		DBLP-area		IMDb	
Classifier	LR	SVM	LR	SVM	LR	SVM
SVD	0.7566	0.7550	0.8158	0.8008	0.6978	0.7013
DeepWalk	0.6629	0.7077	0.8308	0.8390	0.6407	0.6834
LINE	0.7037	0.7314	0.8526	0.8540	0.2949	0.2954
OneSpace	0.7685	0.8333	0.8758	0.8731	0.6888	0.6919
AspEm	**0.8425**	**0.8889**	**0.8786**	**0.8813**	**0.7091**	**0.7139**

Dataset	DBLP						IMDb					
Metrics	*P*@1	*P*@3	*P*@10	*R*@1	*R*@3	*R*@10	*P*@1	*P*@3	*P*@10	*R*@1	*R*@3	*R*@10
ComNeigh	0.6358	0.4620	0.2177	0.2785	0.5539	0.8212	0.1855	0.1590	0.1535	0.0110	0.0279	0.0901
JacCoef	0.6403	0.4692	0.2228	0.2813	0.5596	0.8435	0.3626	0.3453	0.2968	0.0231	0.0669	0.1910
PrefAttach	0.3743	0.2855	0.1618	0.1479	0.3280	0.6062	0.1664	0.1233	0.1106	0.0088	0.0178	0.0512
AdAdar	0.6322	0.4564	0.2131	0.2772	0.5476	0.8053	0.1792	0.1562	0.1528	0.0106	0.0270	0.0904
SVD	0.6648	0.5164	0.2274	0.2939	0.6178	0.8512	0.2470	0.2474	0.2249	0.0152	0.0445	0.1343
DeepWalk	0.7395	0.5297	0.2303	0.3268	0.6329	0.8622	0.3499	0.3605	0.3416	0.0253	0.0774	0.2236
LINE	0.7404	0.5367	0.2299	0.3267	0.6375	0.8596	0.4782	0.4701	0.4130	0.0379	0.1133	0.3137
OneSpace	0.7440	0.5381	0.2279	0.3301	0.6401	0.8519	0.4665	0.4386	0.3852	0.0435	0.1146	0.3038
AspEm	**0.7724**	**0.5645**	**0.2356**	**0.3479**	**0.6749**	**0.8810**	**0.5090**	**0.4853**	**0.4219**	**0.0464**	**0.1296**	**0.3420**

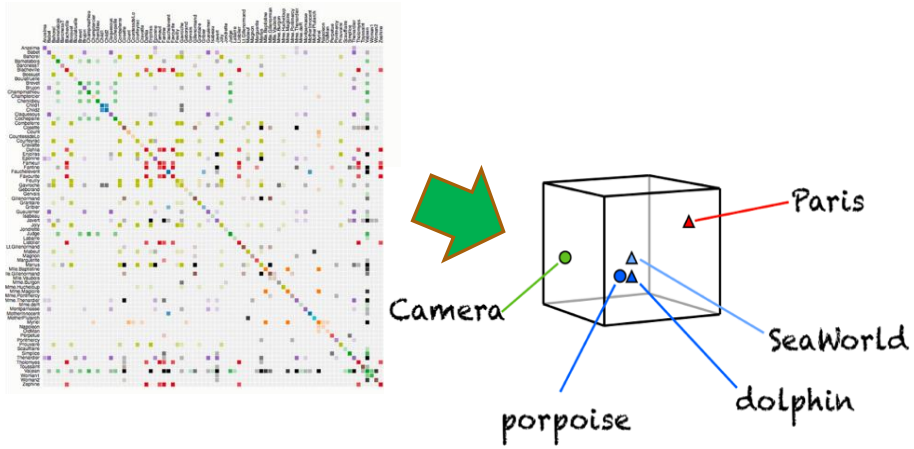
Link Prediction Results on DBLP and IMDb

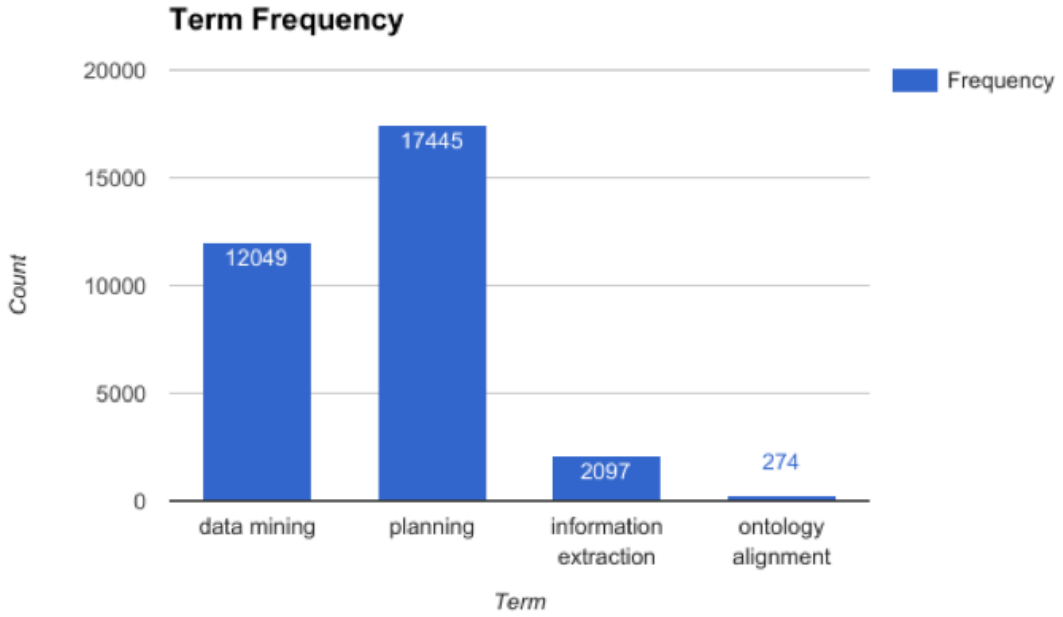
Outline

- Dimension Reduction: From Low-rank Estimation vs. Embedding Learning
- Network Embedding for Homogeneous Networks
- Network Embedding for Heterogeneous Networks
- HEBE: Hyper-Edge Based Embedding in Heterogeneous Networks
- Aspect-Embedding in Heterogeneous Networks
- Locally-Trained Embedding for Expert-Finding in Heterogeneous Networks
- Summary and Discussions

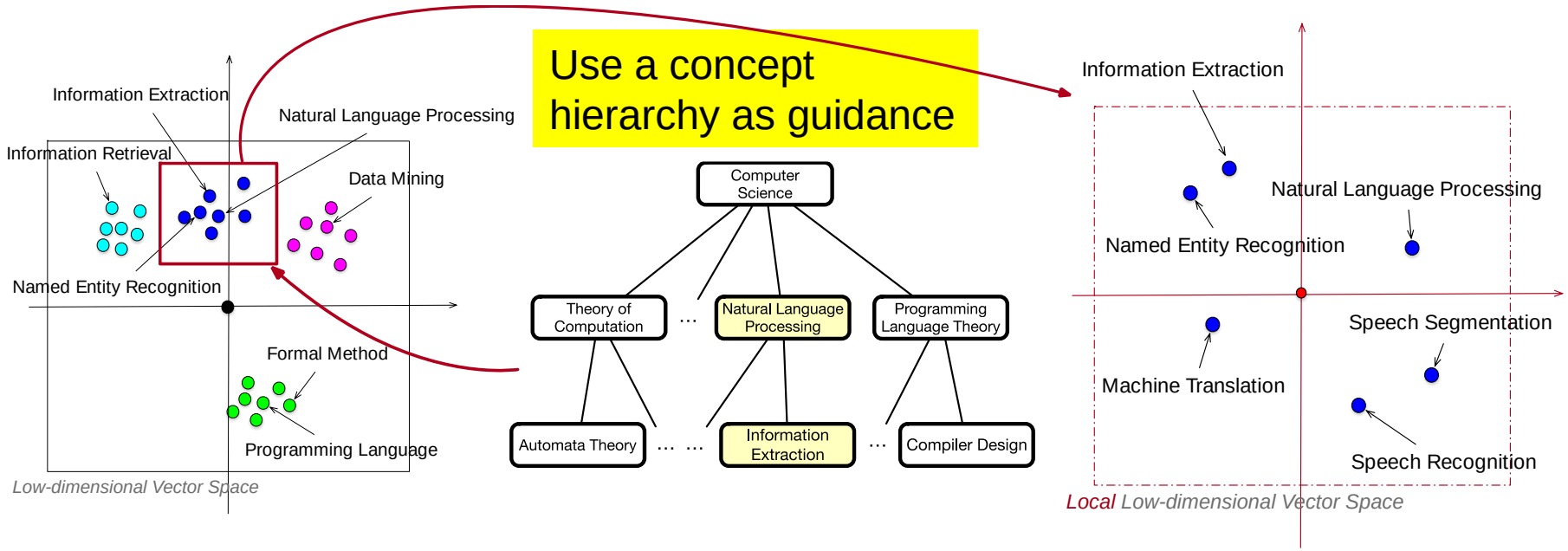
Problem: Expert Finding in Bibliographic Networks

- Given a set of keywords, find related experts
  - Ex. Find expert on "information extraction"
- Challenges: Vocabulary gap
  - *"relation extraction", "named entity recognition", …*
- The power of word embedding
  - Use word embedding to close the vocabulary gap
- Difficulty: Discrepancy in queries
  - Specific queries: Narrow semantic meanings
    - "Information Extraction"
    - "Ontology Alignment"
  - General queries: Broad semantic meanings
    - "Data Mining"
    - "Planning"





Local Embedding Training with Concept Hierarchy



- For an arbitrary query, local embedding can be learned with the sub-corpus constrained on the parent topic — The parent topic becomes background
- Recursive Local Embedding Training
- The idea was proposed and developed by Huan Gui, et al. 2017 "Expert Finding in Heterogeneous Bibliographic Networks with Locally-trained Embeddings"(submitted to ECMLPKDD 2017)

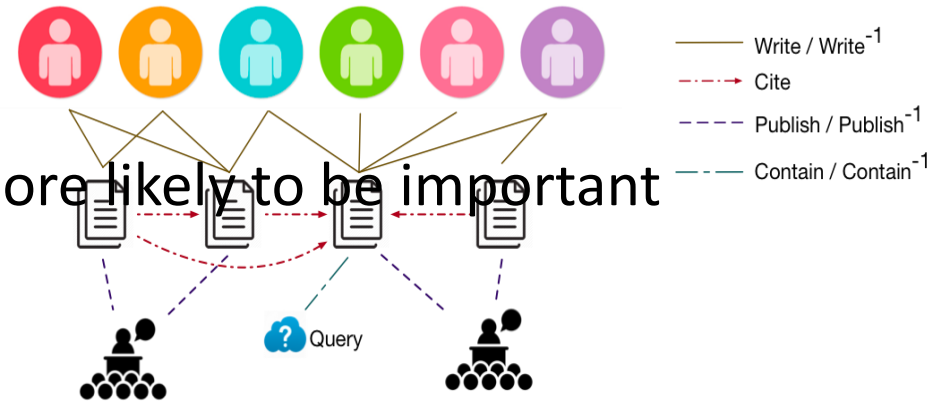
Ranking Experts in Heterogeneous Information Networks

- Expert Finding: Based on both **relevance** and **importance**
- Ranking in networks
  - Relevance Network
    - A candidate may have expertise on multiple topics
    - Only papers relevant to the query can serve as evidence
- Heterogeneous Information Networks
  - Citation may have time-delay factor
  - Papers published in a higher-ranked venue are more likely to be important
  - Venues play an important role for ranking
- Ranking Philosophy
  - Important & relevant papers will be cited by many important & relevant papers
  - Relevant experts will publish many important & relevant papers
  - Relevant conferences will publish many important & relevant papers

Andrew McCallum

Information Extraction

Social Network analysis



Experiments: LE-expert vs. Other Methods

Dataset (DBLP):

Documents: 2,244,018

Authors: 1,274,360

Labels (20 queries):

General: machine-learning, natural-language-processing, planning

Specific: face-recognition, information-extraction, kernel-methods, ontology-alignment…

measure	P@5	P@10	P@20	NDCG@5	NDCG@10	NDCG@20	MAP	bpref
BALOG	0.4941	0.3824	0.2853	0.5068	0.4248	0.3416	0.1608	0.8536
NMF	0.3176	0.2706	0.2118	0.3525	0.3075	0.253	0.1151	0.7303
SVD	0.4353	0.3471	0.2912	0.4553	0.3871	0.3336	0.1548	0.7590
CORANK	0.6941	0.5741	0.4235	0.7181	0.6386	0.5024	0.291	0.8843
EMBED	0.0353	0.0294	0.0265	0.0354	0.0317	0.0289	0.005	0.6331
JOINTHYP	0.6235	0.4176	0.2882	0.6447	0.4913	0.3725	0.1579	**0.9704**
EXACT	0.7059	0.5882	0.4529	0.7548	0.6549	0.5361	0.311	0.8676
RankClass	0.7529	0.6647	0.5176	0.7666	0.7026	0.5867	0.3598	0.8981
LE-expert	**0.8118**	**0.7118**	**0.5559**	**0.8027**	**0.7361**	**0.618**	**0.3826**	0.9451
↑ vs BALOG	64.30%	86.14%	94.84%	58.38%	73.28%	80.91%	137.93%	10.73 %

Significant improvement compared with document-based model (BALOG)

Case Study

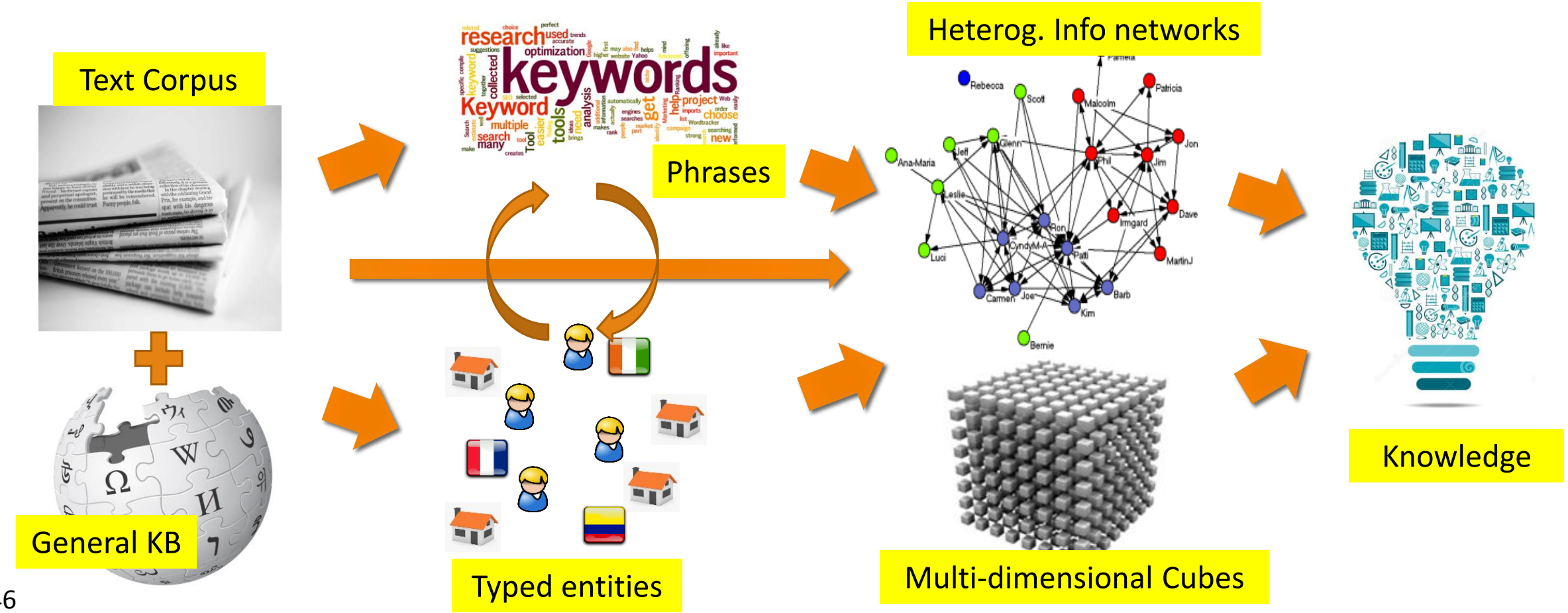
boosting		support vector machine	
Co-ranking	LE-expert	Co-ranking	LE-expert
Robert E. Schapire	**Robert E. Schapire**	Qi Wu	**Bernhard Scholkopf**
Yoav Freund	**Yoav Freund**	Isabelle Guyon	**Vladimir Vapnik**
Ron Kohavi	**Leo Breiman**	**Jason Weston**	**Christopher J. C. Burges**
Thomas G. Dietterich	**Yoram Singer**	**Vladimir Vapnik**	**Thorsten Joachims**
Yoram Singer	**David P. Helmbold**	Bao-Liang Lu	**Chih-Jen Lin**
information extraction		ontology alignment	
Co-ranking	LE-expert	Co-ranking	LE-expert
Ralph Grishman	**Dayne Freitag**	**Jerome euzenat**	**W. Marco Schorlemmer**
Andrew McCallum	**Ralph Grishman**	Patrick Lambrix	**Yannis Kalfoglou**
Ellen Riloff	**Andrew McCallum**	Jason J. Jung	**Anhai Doan**
Oren Etzioni	**Nicholas Kushmerick**	He Tan	**Jerome Euzenat**
Dayne Freitag	**Stephen Soderland**	Marc Ehrig	Alon Y. Halevy

Outline

- Dimension Reduction: From Low-rank Estimation vs. Embedding Learning
- Network Embedding for Homogeneous Networks
- Network Embedding for Heterogeneous Networks
- HEBE: Hyper-Edge Based Embedding in Heterogeneous Networks
- Aspect-Embedding in Heterogeneous Networks
- Locally-Trained Embedding for Expert-Finding in Heterogeneous Networks
- Summary and Discussions

Summary

- Embedding will play an important role in the whole game of data to network to knowledge
- Lots can be explored for network embedding in heterogeneous info. networks!



Acknowledgements

- Thanks for the research support from: ARL/NSCTA, NIH, NSF, DARPA, DTRA, ……