@KDD17

metapath2vec

Scalable Representation Learning for Heterogeneous Networks

Yuxiao Dong
Microsoft Research
& Notre Dame

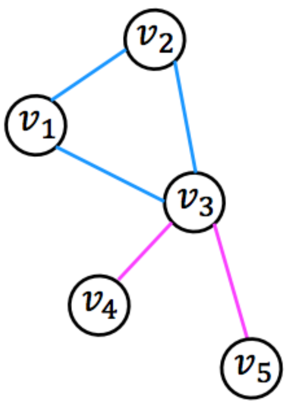
Nitesh V. Chawla
University of Notre Dame

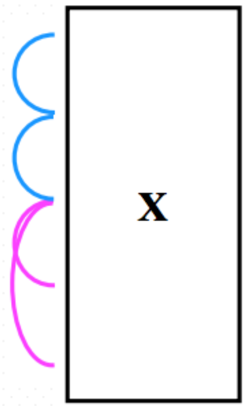
Ananthram Swami
Army Research Lab

Interdisciplinary Center for Network Science and Applications (*iCeNSA*)
University of Notre Dame

UNIVERSITY OF NOTRE DAME

Conventional Network Mining and Learning





hand-crafted feature matrix

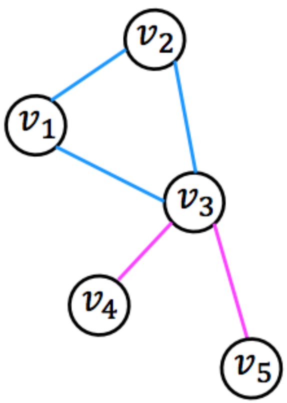
Network Mining Tasks

- ♣ node attribute inference
- ♣ community detection
- ♣ similarity search
- ♣ link prediction
- ♣ social recommendation
- ♣ …

feature engineering

machine learning models

Network Embedding for Mining and Learning



?

X

latent representation matrix

feature learning

machine learning models

Network Mining Tasks

- ♣ node attribute inference
- ♣ community detection
- ♣ similarity search
- ♣ link prediction
- ♣ social recommendation
- ♣ …

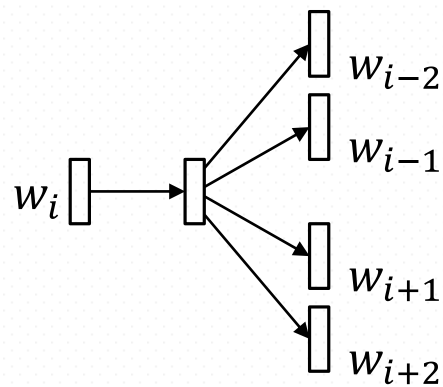
Y. Bengio, A. Courville, and P. Vincent. Representation learning: A review and new perspectives. **IEEE TPAMI**, 35(8):1798–1828, 2013.
Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. **Nature**, 521(7553):436–444, 2015.

Word Embedding in NLP

- Input: a text corpus $D = \{W\}$
- Output: $\boldsymbol{X} \in R^{|W| \times d}$, $d \ll |W|$, *d*-dim vector $\boldsymbol{X}_w$ for each word *w*.

- Computational lens on big social and information networks.
- The connections between individuals form the structural …
- In a network sense, individuals matters in the ways in which ...
- Accordingly, this thesis develops computational models to investigating the ways that ...
- We study two fundamental and interconnected directions: user demographics and network diversity
- ... ...

sentences



word2vec

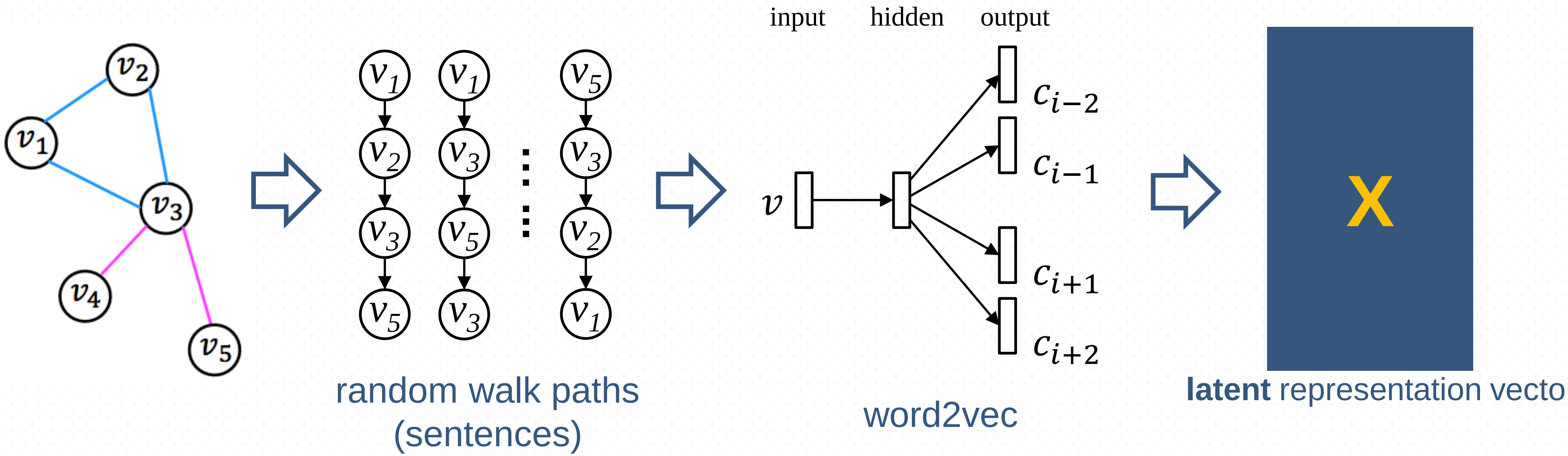
latent representation vector

- geographically close words---a word and its context words---in a sentence or document exhibit interrelations in human natural language.

1. T. Mikolov, I Sutskever, K Chen, GS Corrado, J Dean. Distributed representations of words and phrases and their compositionality. In ***NIPS*** *'13*, pp. 3111-31119.
2. T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv:1301.3781*, 2013.

Network Embedding

- Input: a network $G = (V, E)$
- Output: $\boldsymbol{X} \in R^{|V| \times d}, d \ll |V|$, d-dim vector $\boldsymbol{X}_v$ for each node *v.*



DeepWalk [Perozzi et al., KDD14]

1. B. Perozzi, R. Al-Rfou, and S. Skiena, "DeepWalk: Online learning of social representations," in ***KDD*** *' 14*, pp. 701–710.
2. A. Grover, J. Leskovec. node2vec: Scalable Feature Learning for Networks. in ***KDD*** *'16*, pp. 855—864.
3. T. Mikolov, I Sutskever, K Chen, GS Corrado, J Dean. Distributed representations of words and phrases and their compositionality. In ***NIPS*** *'13*, pp. 3111-31119.
4. T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv:1301.3781*, 2013.

Heterogeneous Network Embedding: Problem

- Input: **a heterogeneous information network** $G = (V, E, T)$
- Output: $\boldsymbol{X} \in R^{|V| \times d}$, $d \ll |V|$, d-dim vector $\boldsymbol{X}_v$ for each node *v*.

latent representation vector

Heterogeneous Network Embedding: Challenges

- How do we effectively preserve the concept of “node-context” among multiple types of nodes, e.g., authors, papers, & venues in academic heterogeneous networks?
- Can we directly apply homogeneous network embedding architectures to heterogeneous networks?
- It is also difficult for conventional meta-path based methods to model similarities between nodes without connected meta-paths.

Heterogeneous Network Embedding: Solutions

metapath2vec

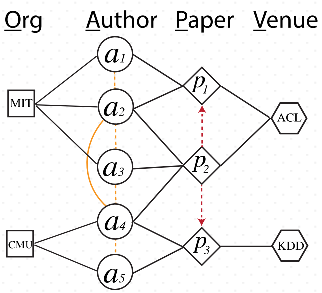
metapath2vec++

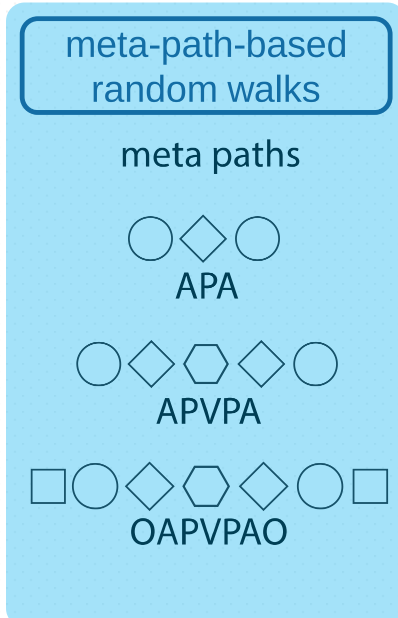
meta-path-based random walks

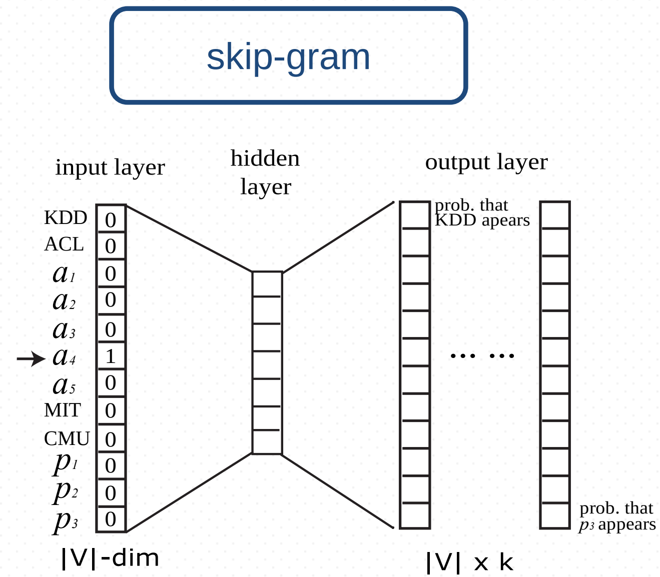
skip-gram

heterogeneous skip-gram

metapath2vec

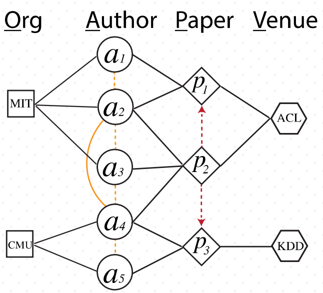






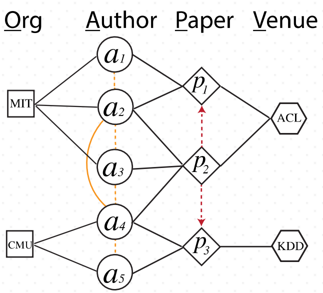
1. Y. Sun, J. Han. Mining heterogeneous information networks: Principles and Methodologies. Morgan & Claypool Publishers, 2012.
2. T. Mikolov, et al. Distributed representations of words and phrases and their compositionality. In ***NIPS*** *'13.*

metapath2vec: Meta-Path-Based Random Walks



Goal: to generate paths that are able to capture both the semantic and structural correlations between different types of nodes, facilitating the transformation of heterogeneous network structures into skip-gram.

metapath2vec: Meta-Path-Based Random Walks



- Given a meta-path scheme

$$\mathcal{P}: V_1 \xrightarrow{R_1} V_2 \xrightarrow{R_2} \cdots V_t \xrightarrow{R_t} V_{t+1} \cdots \xrightarrow{R_{l-1}} V_l$$

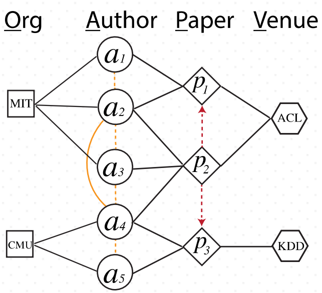
- The transition probability at step *i* is defined as

$$p(v^{i+1}|v_t^i, \mathcal{P}) = \begin{cases} \frac{1}{|N_{t+1}(v_t^i)|} & (v^{i+1}, v_t^i) \in E, \phi(v^{i+1}) = t+1 \\ 0 & (v^{i+1}, v_t^i) \in E, \phi(v^{i+1}) \neq t+1 \\ 0 & (v^{i+1}, v_t^i) \notin E \end{cases}$$

- Recursive guidance for random walkers, i.e.,

$$p(v^{i+1}|v_t^i) = p(v^{i+1}|v_1^i), \text{ if } t = l$$

metapath2vec: Meta-Path-Based Random Walks



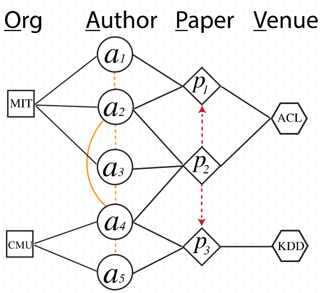
- Given a meta-path scheme (Example)

  OAPVPAO

- In a traditional random walk procedure, in the toy example, the next step of a walker on node a4 transitioned from node CMU can be all types of nodes surrounding it—a2, a3, a5, p2, p3, and CMU.

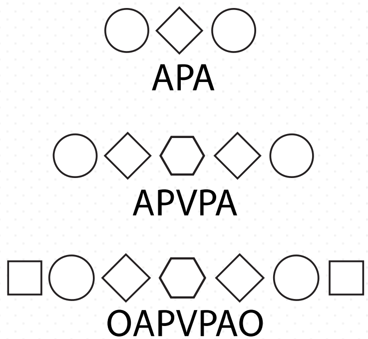
- Under the meta-path scheme 'OAPVPAO', for example, the walker is biased towards paper nodes (P) given its previous step on an organization node CMU (O), following the semantics of this meta-path.

metapath2vec

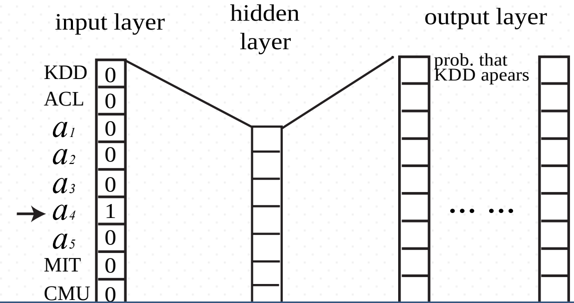


meta-path-based random walks

meta paths



skip-gram

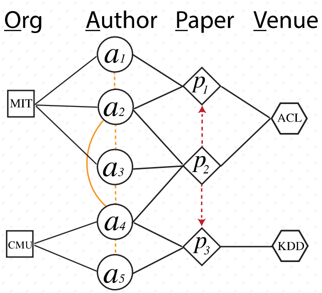


The potential issue of skip-gram for heterogeneous network embedding:

To predict the context node c_t (type t) given a node v, *metapath2vec* encourages all types of nodes to appear in this context position

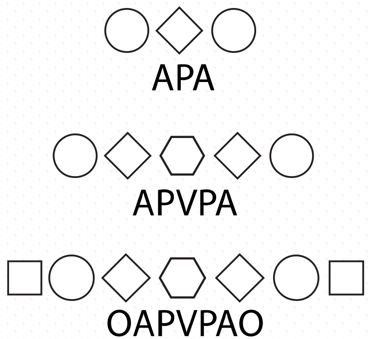
1. Y. Sun, J. Han. Mining heterogeneous information networks: Principles and Methodologies. Morgan & Claypool Publishers, 2012.
2. T. Mikolov, et al. Distributed representations of words and phrases and their compositionality. In ***NIPS*** *'13.*

metapath2vec++

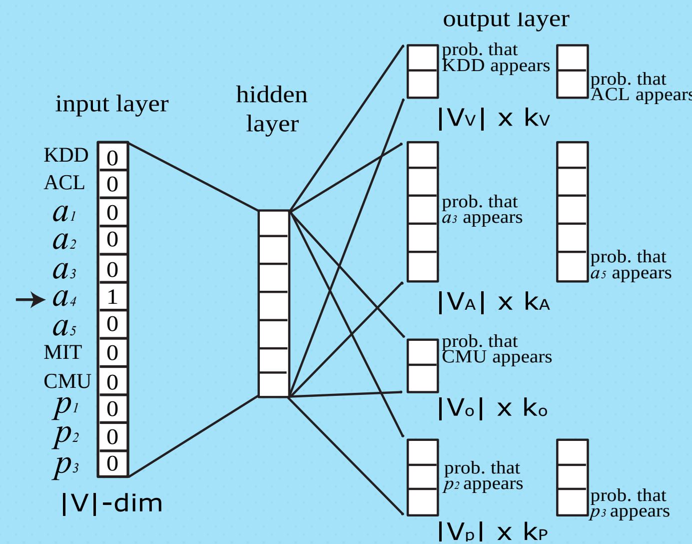


meta-path-based
random walks

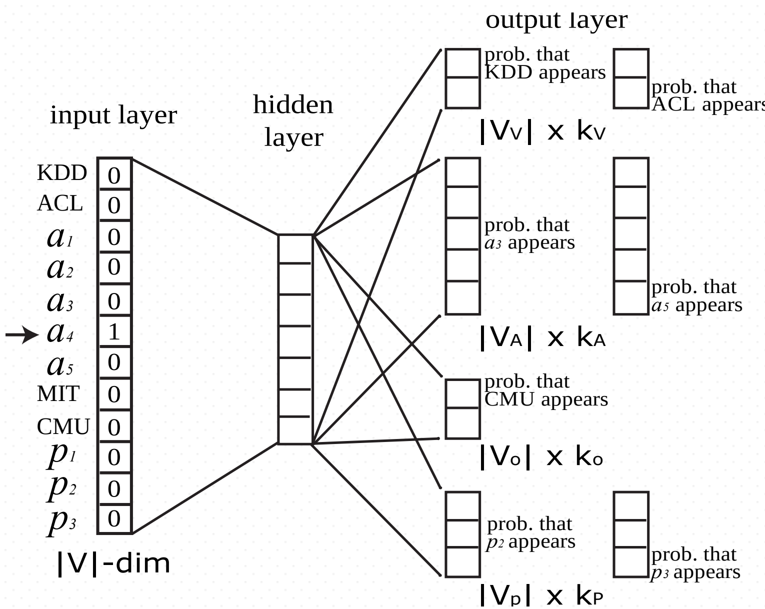
meta paths



heterogeneous
skip-gram



metapath2vec++: Heterogeneous Skip-Gram



- softmax in *metapath2vec*

$$p(c_t|v;\theta) = \frac{e^{X_{c_t}} \cdot e^{X_v}}{\sum_{u \in V} e^{X_u} \cdot e^{X_v}}$$

- softmax in *metapath2vec++*

$$p(c_t|v;\theta) = \frac{e^{X_{c_t}} \cdot e^{X_v}}{\sum_{u_t \in V_t} e^{X_{u_t}} \cdot e^{X_v}}$$

- objective function (heterogeneous negative sampling)

$$\mathcal{O}(\mathbf{X}) = \log \sigma(X_{c_t} \cdot X_v) + \sum_{k=1}^{K} \mathbb{E}_{u_t^k \sim P_t(u_t)}[\log \sigma(-X_{u_t^k} \cdot X_v)]$$

- stochastic gradient descent

$$\frac{\partial \mathcal{O}(\mathbf{X})}{\partial X_{u_t^k}} = (\sigma(X_{u_t^k} \cdot X_v - \mathbb{I}_{c_t}[u_t^k]))X_v$$

$$\frac{\partial \mathcal{O}(\mathbf{X})}{\partial X_v} = \sum_{k=0}^{K}(\sigma(X_{u_t^k} \cdot X_v - \mathbb{I}_{c_t}[u_t^k]))X_{u_t^k}$$

1. T. Mikolov, et al. Distributed representations of words and phrases and their compositionality. In ***NIPS*** *'13*.

metapath2vec++

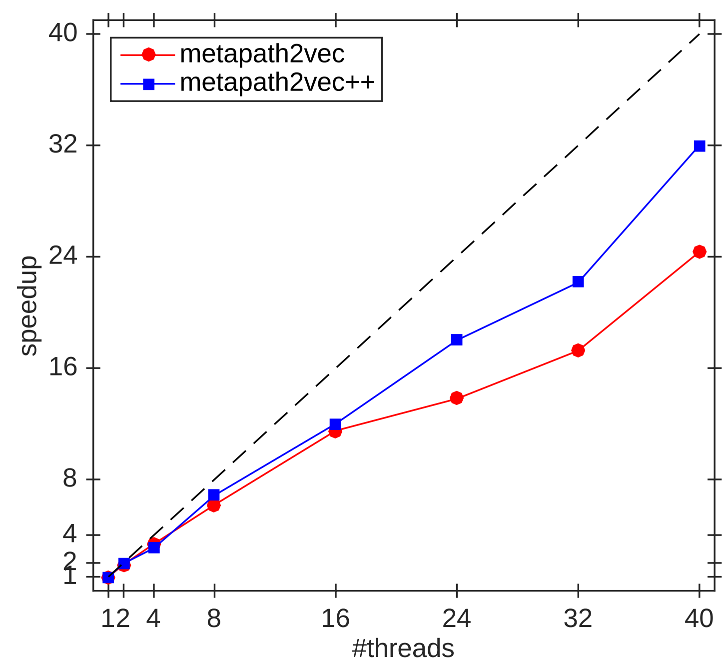
Input: The heterogeneous information network $G = (V, E, T)$, a meta-path scheme $\mathcal{P}$, #walks per node w, walk length l, embedding dimension d, neighborhood size k
Output: The latent node embeddings $\mathbf{X} \in \mathbb{R}^{|V| \times d}$

initialize $\mathbf{X}$;
for $i = 1 \rightarrow w$ **do**
    for $v \in V$ **do**
        MP = MetaPathRandomWalk(G, $\mathcal{P}$, v, l) ;
        $\mathbf{X}$ = HeterogeneousSkipGram($\mathbf{X}$, k, MP) ;
    end
end
return $\mathbf{X}$;

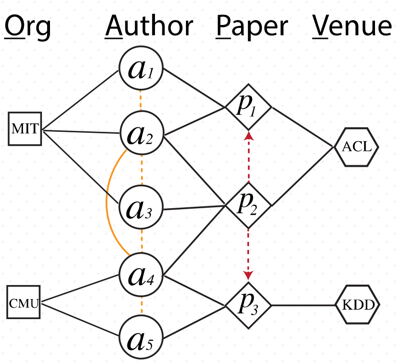
MetaPathRandomWalk(G, $\mathcal{P}$, v, l)
$MP[1] = v$;
for $i = 1 \rightarrow l{-}1$ **do**
    draw u according to Eq. 3 ;
    MP[i+1] = u ;
end
return MP ;

HeterogeneousSkipGram($\mathbf{X}$, k, MP)
for $i = 1 \rightarrow l$ **do**
    $v = MP$[i] ;
    for $j = max(0, i\text{-}k) \rightarrow min(i{+}k, l)$ & $j \neq i$ **do**
        $c_t = MP$[j] ;
        $X^{new} = X^{old} - \eta \cdot \frac{\partial O(\mathrm{X})}{\partial X}$ (Eq. 7) ;
    end
end

- every sub-procedure is easy to parallelize
- 24-32X speedup by using 40 cores



Network Mining and Learning Paradigm



metapath2vec

metapath2vec++

X

latent representation vector

Network Applications

- node attribute inference
- community detection
- similarity search
- link prediction
- social recommendation
- …

Experiments

Heterogeneous Data

- AMiner Academic Network
  - ~~9~~ 1.7 million authors
  - 3 million papers
  - 3800+ venues
  - 8 research areas

Baselines

- DeepWalk [KDD ’14]
- node2vec [KDD ’16]
- LINE [WWW ’15]
- PTE [KDD ’15]

Parameters

- #walks: 1000
- walk-length: 100
- #dimensions: 128
- neighborhood size: 7

Mining Tasks

- node classification
  - logistic regression
- node clustering
  - k-means
- similarity search
  - cosine similarity

J. Tang, et al. ArnetMiner: Extraction and Mining of Academic Social Networks. In *KDD 2008.*
https://aminer.org/aminernetwork

Application 1: Multi-Class Node Classification

Table 2: Multi-class venue node classification results in AMiner data.

Metric	Method	5%	10%	20%	30%	40%	50%	60%	70%	80%	90%
Macro-F1	DeepWalk/node2vec	0.0723	0.1396	0.1905	0.2795	0.3427	0.3911	0.4424	0.4774	0.4955	0.4457
	LINE (1st+2nd)	0.2245	0.4629	0.7011	0.8473	0.8953	0.9203	0.9308	0.9466	0.9410	0.9466
	PTE	0.1702	0.3388	0.6535	0.8304	0.8936	0.9210	0.9352	0.9505	0.9525	0.9489
	metapath2vec	0.3033	0.5247	0.8033	0.8971	0.9406	0.9532	0.9529	0.9701	0.9683	0.9670
	metapath2vec++	0.3090	0.5444	0.8049	0.8995	0.9468	0.9580	0.9561	0.9675	0.9533	0.9503
Micro-F1	DeepWalk/node2vec	0.1701	0.2142	0.2486	0.3266	0.3788	0.4090	0.4630	0.4975	0.5259	0.5286
	LINE (1st+2nd)	0.3000	0.5167	0.7159	0.8457	0.8950	0.9209	0.9333	0.9500	0.9556	0.9571
	PTE	0.2512	0.4267	0.6879	0.8372	0.8950	0.9239	0.9352	0.9550	0.9667	0.9571
	metapath2vec	0.4173	0.5975	0.8327	0.9011	0.9400	0.9522	0.9537	0.9725	0.9815	0.9857
	metapath2vec++	0.4331	0.6192	0.8336	0.9032	0.9463	0.9582	0.9574	0.9700	0.9741	0.9786

Application 1: Multi-Class Node Classification

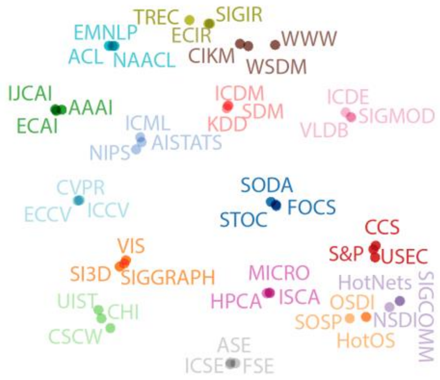
Table 3: Multi-class author node classification results in AMiner data.

Metric	Method	5%	10%	20%	30%	40%	50%	60%	70%	80%	90%
Macro-F1	DeepWalk/node2vec	0.7153	0.7222	0.7256	0.7270	0.7273	0.7274	0.7273	0.7271	0.7275	0.7275
	LINE (1st+2nd)	0.8849	0.8886	0.8911	0.8921	0.8926	0.8929	0.8934	0.8936	0.8938	0.8934
	PTE	0.8898	0.8940	0.897	0.8982	0.8987	0.8990	0.8997	0.8999	0.9002	0.9005
	metapath2vec	0.9216	0.9262	0.9292	0.9303	0.9309	0.9314	0.9315	0.9316	0.9319	0.9320
	metapath2vec++	0.9107	0.9156	0.9186	0.9199	0.9204	0.9207	0.9207	0.9208	0.9211	0.9212
Micro-F1	DeepWalk/node2vec	0.7312	0.7372	0.7402	0.7414	0.7418	0.7420	0.7419	0.7420	0.7425	0.7425
	LINE (1st+2nd)	0.8936	0.8969	0.8993	0.9002	0.9007	0.9010	0.9015	0.9016	0.9018	0.9017
	PTE	0.8986	0.9023	0.9051	0.9061	0.9066	0.9068	0.9075	0.9077	0.9079	0.9082
	metapath2vec	0.9279	0.9319	0.9346	0.9356	0.9361	0.9365	0.9365	0.9365	0.9367	0.9369
	metapath2vec++	0.9173	0.9217	0.9243	0.9254	0.9259	0.9261	0.9261	0.9262	0.9264	0.9266

Application 2: Node Clustering

Node clustering results (NMI) in AMiner

methods	venue	author
DeepWalk/node2vec	0.1952	0.2941
LINE (1st+2nd)	0.8967	0.6423
PTE	0.9060	0.6483
metapath2vec	0.9274	0.7470
metapath2vec++	0.9261	0.7354



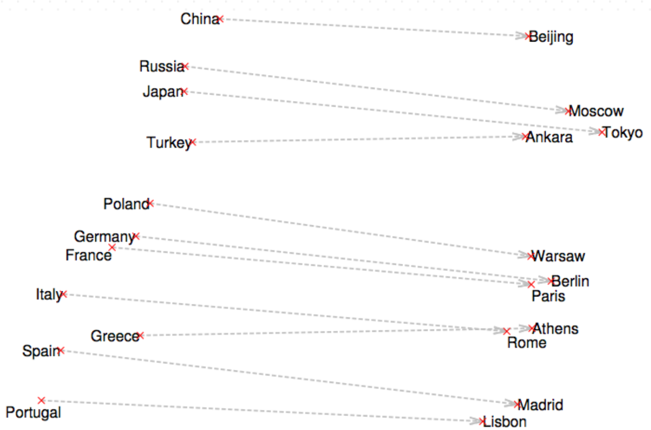
http://projector.tensorflow.org/

Application 3: Similarity Search

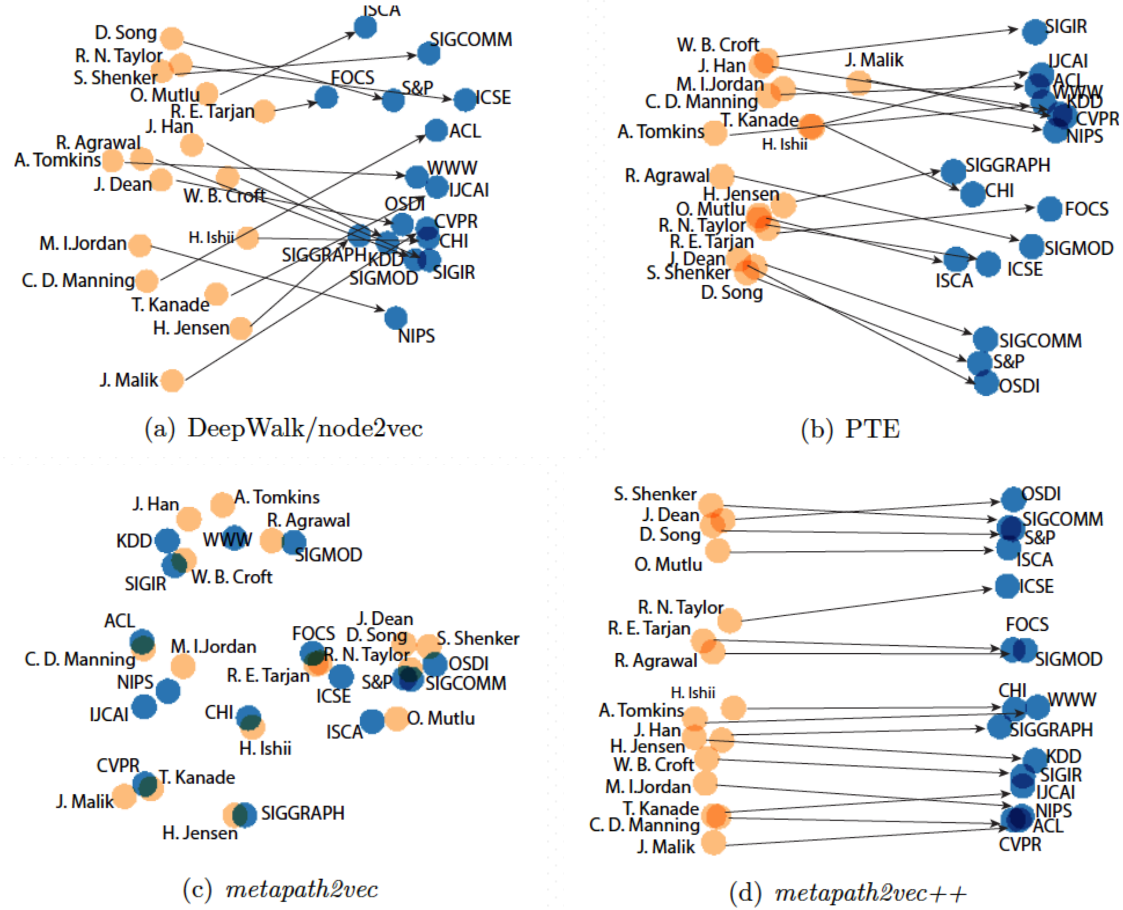
Table 5: Case study of similarity search in AMiner Data

Rank	ACL	NIPS	IJCAI	CVPR	FOCS	SOSP	ISCA	S&P	ICSE	SIGGRAPH	SIGCOMM	CHI	KDD	SIGMOD	SIGIR	WWW
0	ACL	NIPS	IJCAI	CVPR	FOCS	SOSP	ISCA	S&P	ICSE	SIGGRAPH	SIGCOMM	CHI	KDD	SIGMOD	SIGIR	WWW
1	EMNLP	ICML	AAAI	ECCV	STOC	TOCS	HPCA	CCS	TOSEM	TOG	CCR	CSCW	SDM	PVLDB	ECIR	WSDM
2	NAACL	AISTATS	AI	ICCV	SICOMP	OSDI	MICRO	NDSS	FSE	SI3D	HotNets	TOCHI	TKDD	ICDE	CIKM	CIKM
3	CL	JMLR	JAIR	IJCV	SODA	HotOS	ASPLOS	USENIX S	ASE	RT	NSDI	UIST	ICDM	DE Bull	IR J	TWEB
4	CoNLL	NC	ECAI	ACCV	A-R	SIGOPS E	PACT	ACSAC	ISSTA	CGF	CoNEXT	DIS	DMKD	VLDBJ	TREC	ICWSM
5	COLING	MLJ	KR	CVIU	TALG	ATC	ICS	JCS	E SE	NPAR	IMC	HCI	KDD E	EDBT	SIGIR F	HT
6	IJCNLP	COLT	AI Mag	BMVC	ICALP	NSDI	HiPEAC	ESORICS	MSR	Vis	TON	MobileHCI	WSDM	TODS	ICTIR	SIGIR
7	NLE	UAI	ICAPS	ICPR	ECCC	OSR	PPOPP	TISS	ESEM	JGT	INFOCOM	INTERACT	CIKM	CIDR	WSDM	KDD
8	ANLP	KDD	CI	EMMCVPR	TOC	ASPLOS	ICCD	ASIACCS	A SE	VisComp	PAM	GROUP	PKDD	SIGMOD R	TOIS	TIT
9	LREC	CVPR	AIPS	T on IP	JAlG	EuroSys	CGO	RAID	ICPC	GI	MobiCom	NordiCHI	ICML	WebDB	IPM	WISE
10	EACL	ECML	UAI	WACV	ITCS	SIGCOMM	ISLPED	CSFW	WICSA	CG	IPTPS	UbiComp	PAKDD	PODS	AIRS	WebSci

Visualization



word2vec [Mikolov, 2013]



(a) DeepWalk/node2vec

(b) PTE

(c) *metapath2vec*

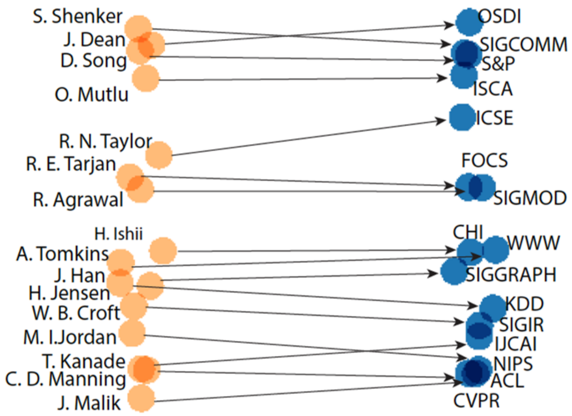
(d) *metapath2vec++*

http://projector.tensorflow.org/

- **Problem:** Heterogeneous Network Embedding

- **Models:** *metapath2vec* & *metapath2vec++*
  - The automatic discovery of internal semantic relationships between different types of nodes in heterogeneous networks

- **Applications:** classification, clustering, & similarity search



Thank you!

Data & Code

metapath2vec

https://ericdongyx.github.io/metapath2vec/m2v.html