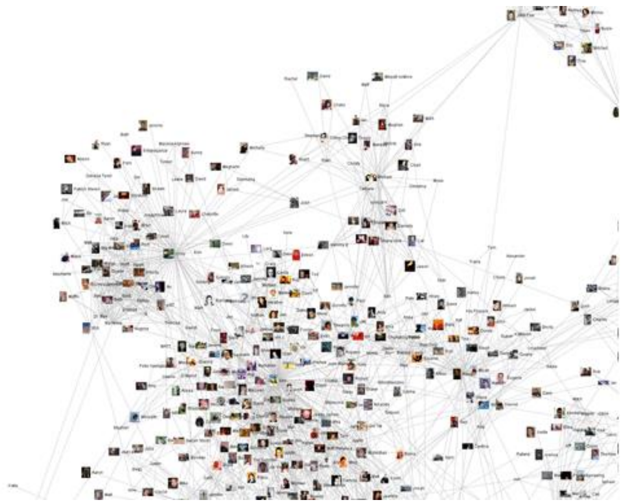
Representation Learning on Networks

Yuxiao Dong

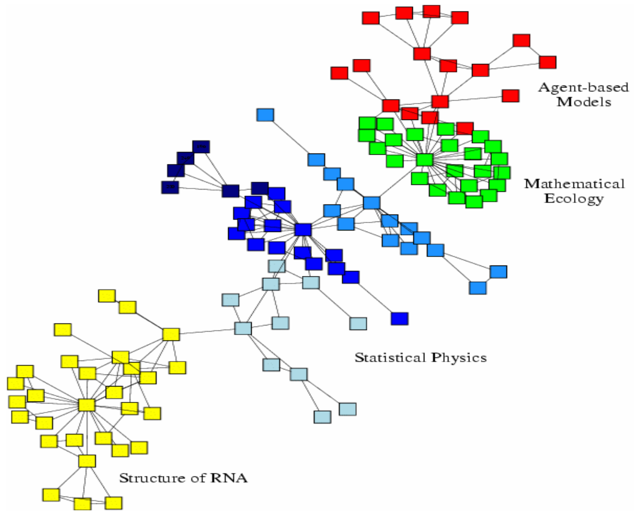
Microsoft Research, Redmond

Joint work with Jiezhong Qiu, Jie Zhang, Jie Tang (Tsinghua University)
Hao Ma (MSR & Facebook AI) and Kuansan Wang (MSR)

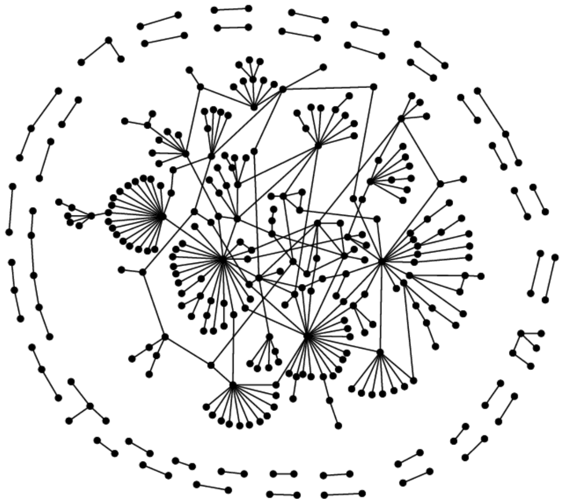
Networks



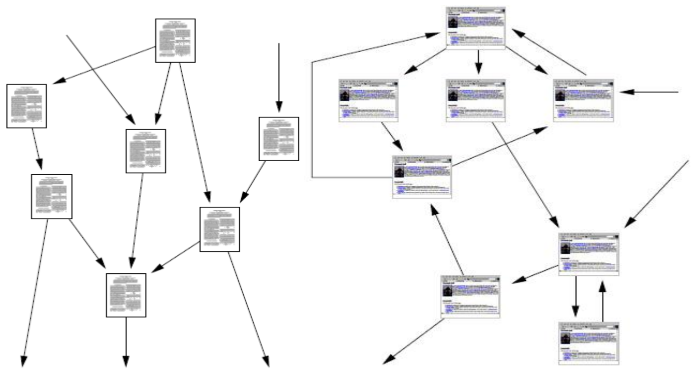
Social networks



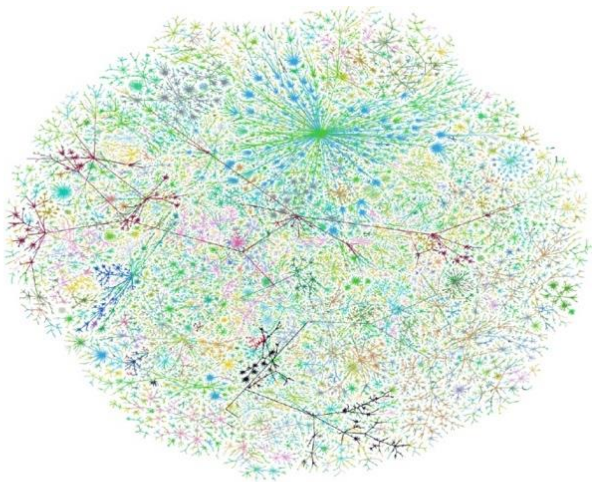
Economic networks



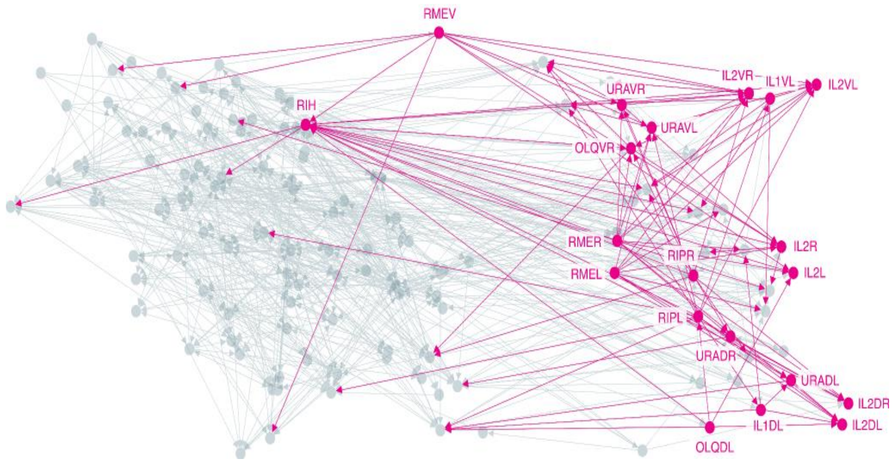
Biomedical networks



Information networks



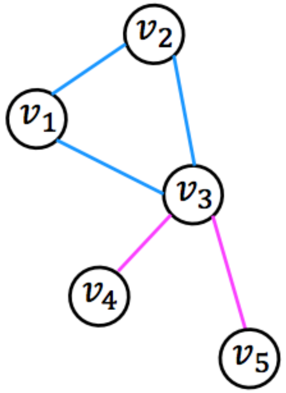
Internet



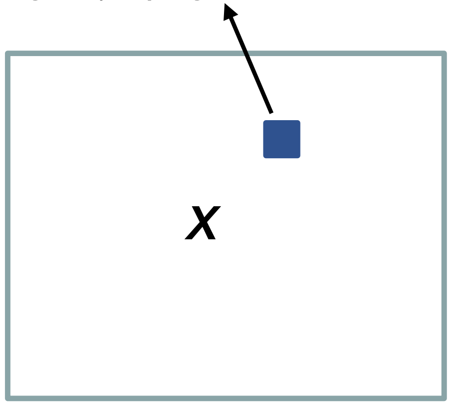
Networks of neurons

Slides credit: Jure Leskovec

The Network & Graph Mining Paradigm



x_{ij}: node v_i's j^{th} feature,
e.g., v_i's pagerank value



hand-crafted feature matrix

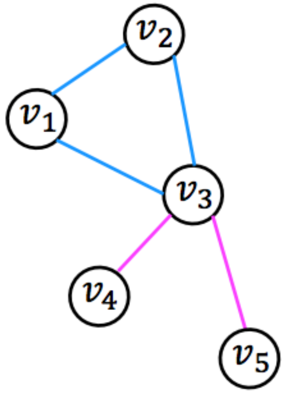
Graph & network applications

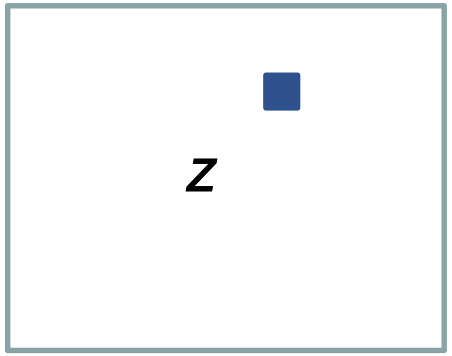
- Node label inference;
- Link prediction;
- User behavior… …

feature **engineering**

machine learning models

Representation Learning for Networks





~~hand-crafted~~ **latent** feature matrix

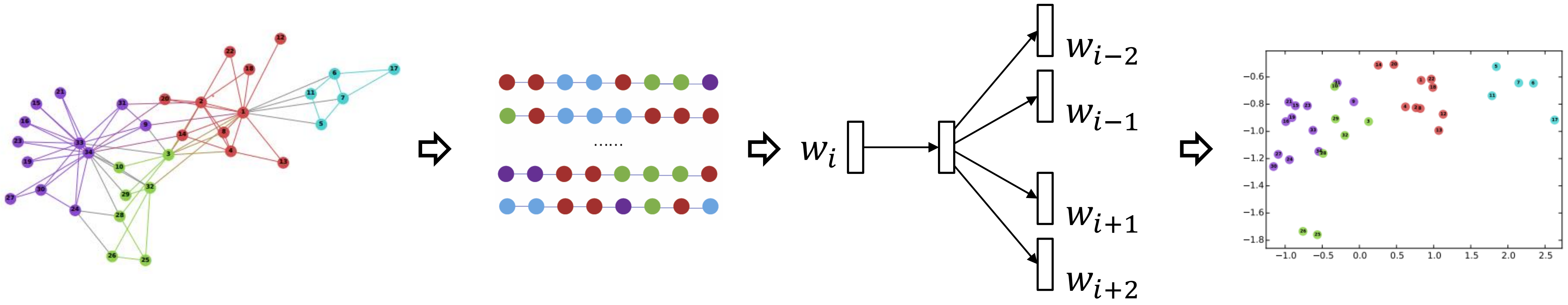
Graph & network applications
- Node label inference;
- Node clustering;
- Link prediction;
- … …

Feature ~~engineering~~ **learning**

machine learning models

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

Network Embedding: Random Walk + Skip-Gram



- sentences in NLP
- vertex-paths in Networks

skip-gram
(word2vec)

Perozzi et al. DeepWalk: Online learning of social representations. In *KDD'14*, pp. 701–710.

Random Walk Strategies

- Random Walk
  - DeepWalk (walk length > 1)
  - LINE (walk length = 1)
- Biased Random Walk
  - 2nd order Random Walk
    - **node2vec**
  - Metapath guided Random Walk
    - **metapath2vec**

Application: Embedding Heterogeneous Academic Graph

225,572,477
Papers

244,499,947
Authors

664,891
Topics

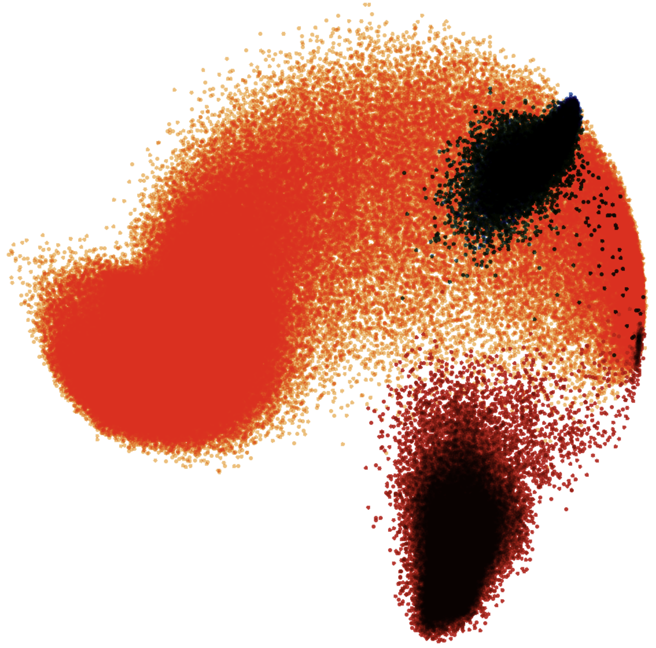
4,397
Conferences

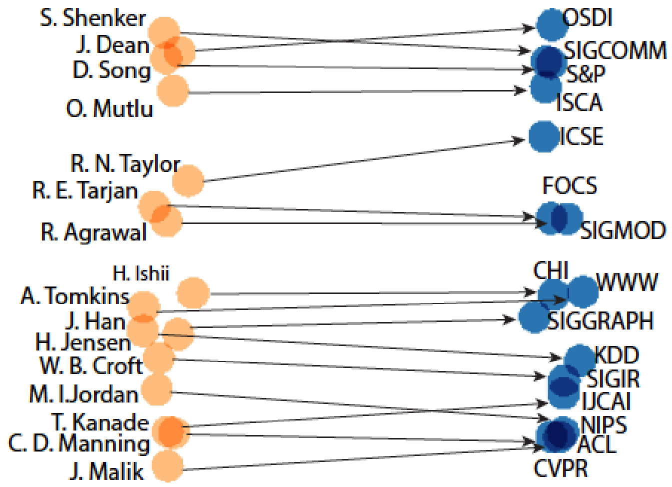
48,758
Journals

25,554
Institutions

Microsoft Academic Graph

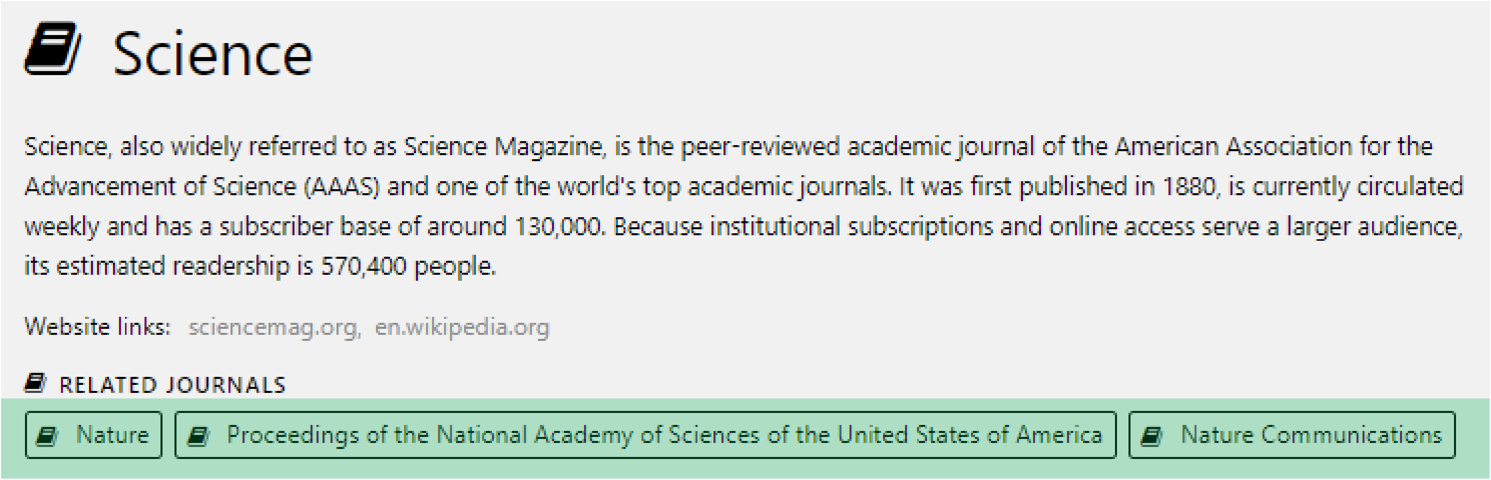
⇨ metapath2vec ⇨





- https://academic.microsoft.com/
- https://www.openacademic.ai/oag/
- metapath2vec: scalable representation learning for heterogeneous networks. In *KDD* 2017.

Application 1: Related Venues



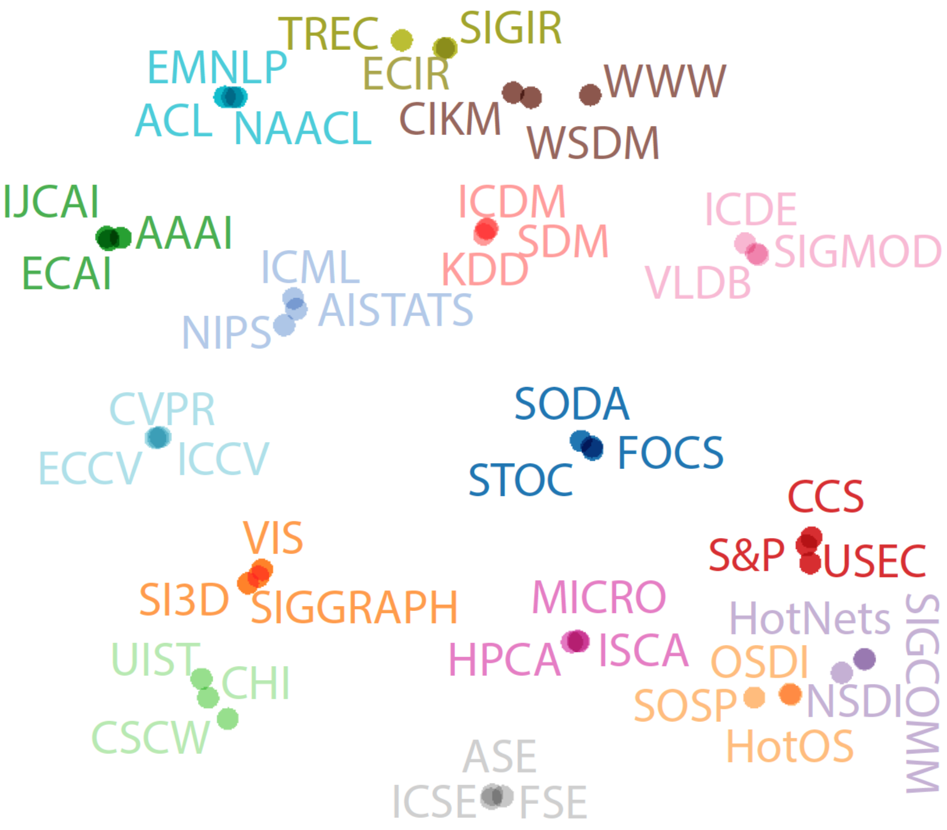
Science

Science, also widely referred to as Science Magazine, is the peer-reviewed academic journal of the American Association for the Advancement of Science (AAAS) and one of the world's top academic journals. It was first published in 1880, is currently circulated weekly and has a subscriber base of around 130,000. Because institutional subscriptions and online access serve a larger audience, its estimated readership is 570,400 people.

Website links: sciencemag.org, en.wikipedia.org

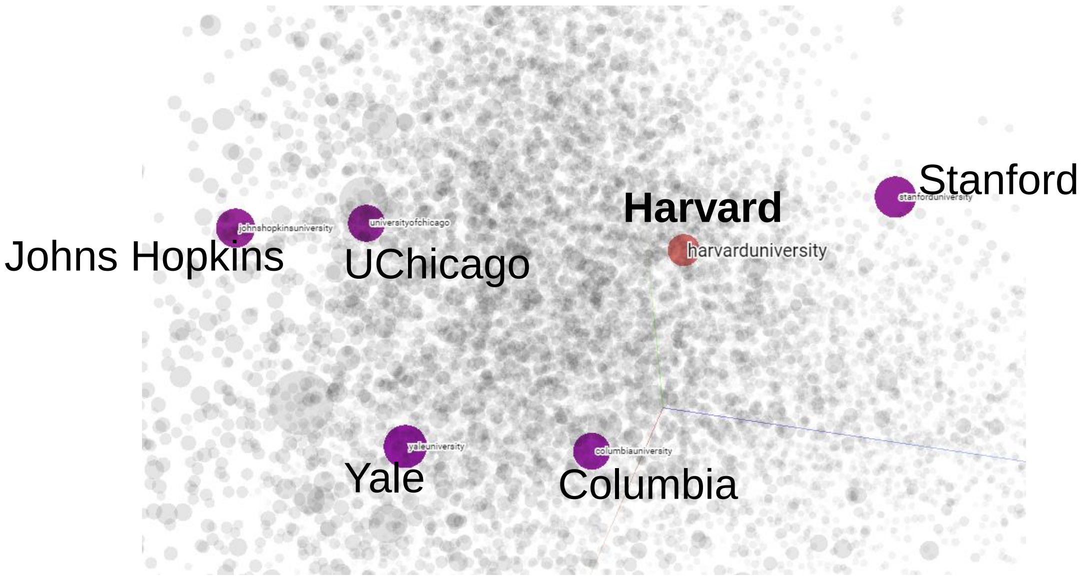
RELATED JOURNALS

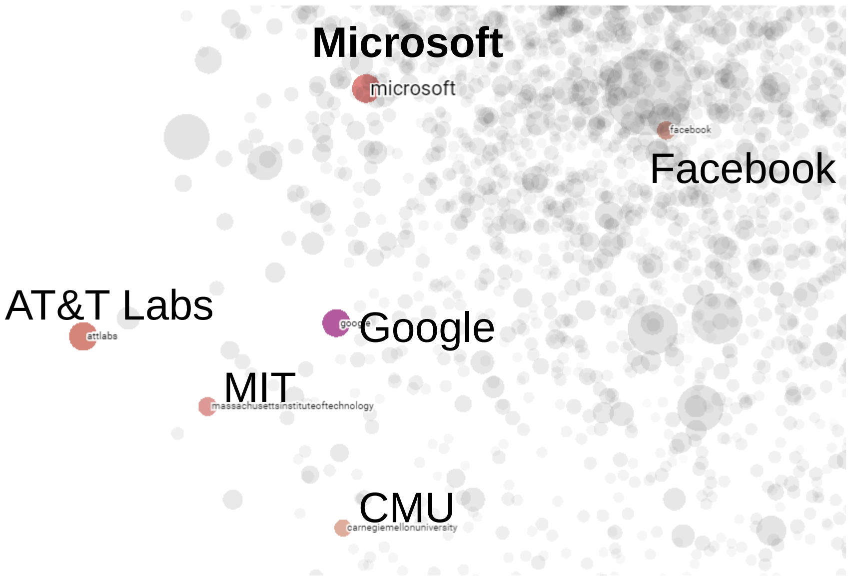
Nature | Proceedings of the National Academy of Sciences of the United States of America | Nature Communications



- https://academic.microsoft.com/
- https://www.openacademic.ai/oag/
- metapath2vec: scalable representation learning for heterogeneous networks. In *KDD* 2017.

Application 2: Similarity Search (Institution)





- https://academic.microsoft.com/
- https://www.openacademic.ai/oag/
- metapath2vec: scalable representation learning for heterogeneous networks. In *KDD* 2017.

Network Embedding

Random Walk ⇨ Skip Gram

Input:

Adjacency Matrix

$\boldsymbol{A}$

⇨

- Random Walk
  - DeepWalk (walk length > 1)
  - LINE (walk length = 1)
- Biased Random Walk
  - 2nd order Random Walk
    - **node2vec**
  - Metapath guided Random Walk
    - **metapath2vec**

⇨

Output:

Vectors

$\boldsymbol{Z}$

Unifying DeepWalk, LINE, PTE, & node2vec as Matrix Factorization

- DeepWalk $\log\left(\frac{\mathrm{vol}(G)}{b}\left(\frac{1}{T}\sum_{r=1}^{T}\left(\boldsymbol{D}^{-1}\boldsymbol{A}\right)^{r}\right)\boldsymbol{D}^{-1}\right)$
- LINE $\log\left(\frac{\mathrm{vol}(G)}{b}\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{D}^{-1}\right)$
- PTE $\log\left(\begin{bmatrix}\alpha\,\mathrm{vol}(G_{\mathrm{ww}})(\boldsymbol{D}_{\mathrm{row}}^{\mathrm{ww}})^{-1}\boldsymbol{A}_{\mathrm{ww}}(\boldsymbol{D}_{\mathrm{col}}^{\mathrm{ww}})^{-1}\\ \beta\,\mathrm{vol}(G_{\mathrm{dw}})(\boldsymbol{D}_{\mathrm{row}}^{\mathrm{dw}})^{-1}\boldsymbol{A}_{\mathrm{dw}}(\boldsymbol{D}_{\mathrm{col}}^{\mathrm{dw}})^{-1}\\ \gamma\,\mathrm{vol}(G_{\mathrm{lw}})(\boldsymbol{D}_{\mathrm{row}}^{\mathrm{lw}})^{-1}\boldsymbol{A}_{\mathrm{lw}}(\boldsymbol{D}_{\mathrm{col}}^{\mathrm{lw}})^{-1}\end{bmatrix}\right)-\log b$
- node2vec $\log\left(\frac{\frac{1}{2T}\sum_{r=1}^{T}\left(\sum_{u}\boldsymbol{X}_{w,u}\underline{\boldsymbol{P}}_{c,w,u}^{r}+\sum_{u}\boldsymbol{X}_{c,u}\underline{\boldsymbol{P}}_{w,c,u}^{r}\right)}{b\left(\sum_{u}\boldsymbol{X}_{w,u}\right)\left(\sum_{u}\boldsymbol{X}_{c,u}\right)}\right)$

$\boldsymbol{A}$ Adjacency matrix

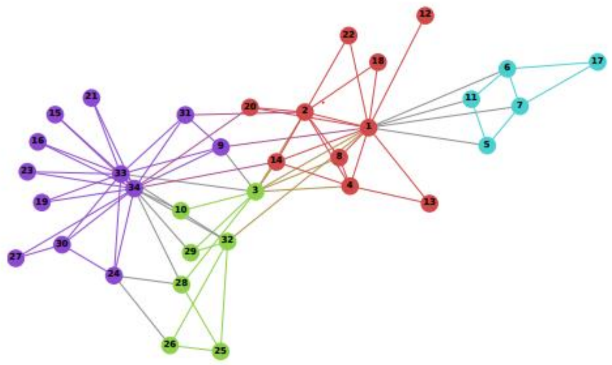
$\boldsymbol{D}$ Degree matrix

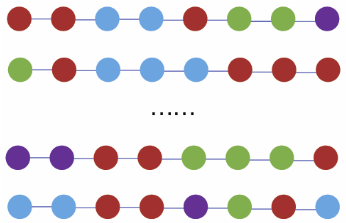
$$vol(G)=\sum_i\sum_j A_{ij}$$

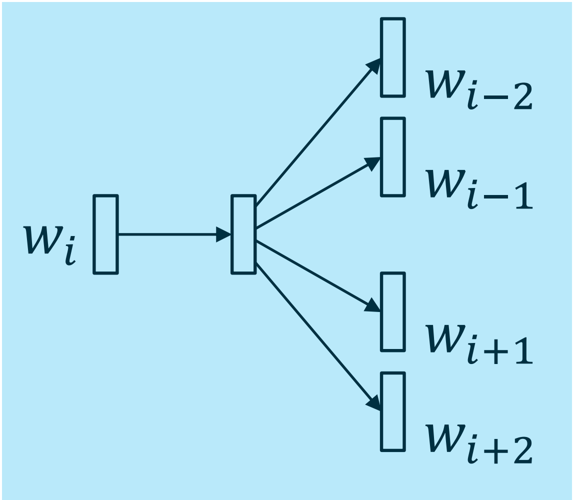
b: #negative samples
T: context window size

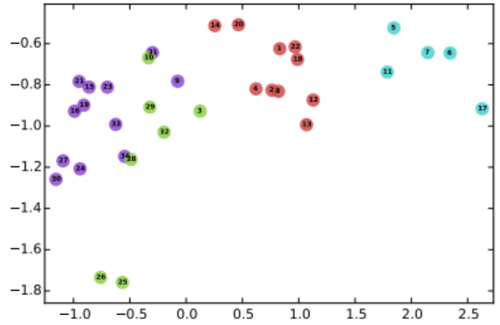
1. Qiu et al. Network embedding as matrix factorization: unifying deepwalk, line, pte, and node2vec. In *WSDM'18.*

Understanding Random Walk + Skip Gram









$G = (V, E)$

- Adjacency matrix $\boldsymbol{A}$
- Degree matrix $\boldsymbol{D}$
- Volume of G: $vol(G)$

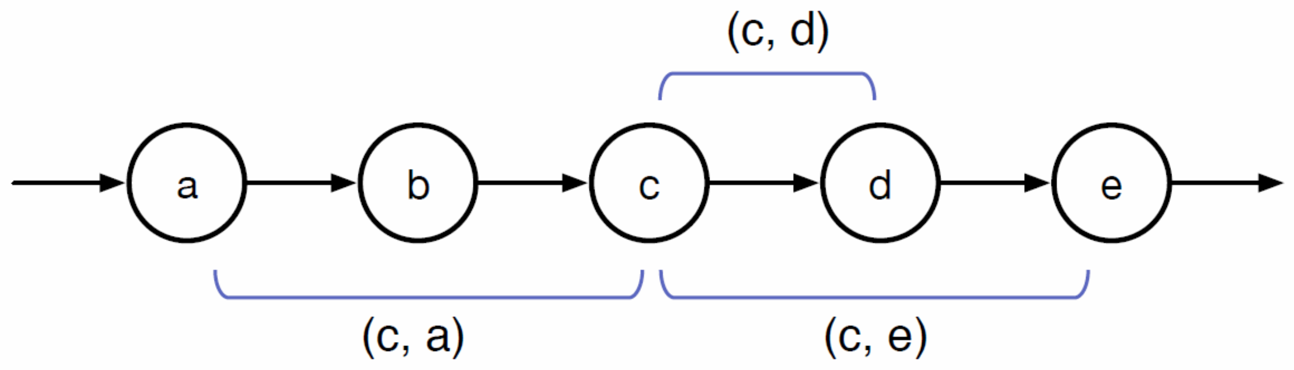
?

$$\log\left(\frac{\#(\boldsymbol{w},\boldsymbol{c})|\mathcal{D}|}{b\#(w)\#(c)}\right)$$

- (w, c): co-occurrence of w & c
- (w): occurrence of node w
- (c): occurrence of context c
- $\mathcal{D}$: node−context pair (w, c) multi−set
- $|\mathcal{D}|$: number of node-context pairs

Levy and Goldberg. Neural word embeddings as implicit matrix factorization. In *NIPS 2014*

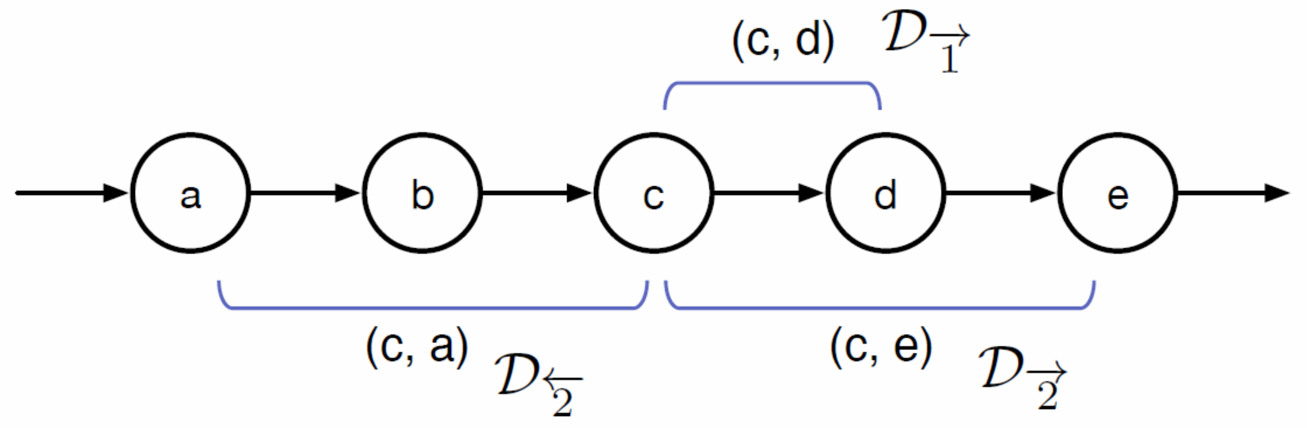
Understanding Random Walk + Skip Gram



$$\log(\frac{\#(\boldsymbol{w},\boldsymbol{c})|\mathcal{D}|}{b\#(w)\#(c)})$$

- (w, c): co-occurrence of w & c
- (w): occurrence of node w
- (c): occurrence of context c
- $\mathcal{D}$: node−context pair (w, c) multi−set
- $|\mathcal{D}|$: number of node-context pairs

Understanding Random Walk + Skip Gram



$$\log\left(\frac{\#(\boldsymbol{w},\boldsymbol{c})|\mathcal{D}|}{b\#(w)\#(c)}\right)$$

- (w, c): co-occurrence of w & c
- (w): occurrence of node w
- (c): occurrence of context c
- $\mathcal{D}$: node−context pair (w, c) multi−set
- $|\mathcal{D}|$: number of node-context pairs

- Partition the multiset $\mathcal{D}$ into several sub-multisets according to the way in which each node and its context appear in a random walk node sequence.
- More formally, for $r = 1, 2, \cdots, T$, we define

$$\mathcal{D}_{\overrightarrow{r}} = \left\{(w,c) : (w,c) \in \mathcal{D}, w = w_j^n, c = w_{j+r}^n\right\}$$
$$\mathcal{D}_{\overleftarrow{r}} = \left\{(w,c) : (w,c) \in \mathcal{D}, w = w_{j+r}^n, c = w_j^n\right\}$$

Distinguish direction and distance

Understanding Random Walk + Skip Gram

$$\log\left(\frac{\#(w,c)\,|\mathcal{D}|}{b\#(w)\cdot\#(c)}\right) = \log\left(\frac{\frac{\#(w,c)}{|\mathcal{D}|}}{b\frac{\#(w)}{|\mathcal{D}|}\frac{\#(c)}{|\mathcal{D}|}}\right)$$

- (w,c): co-occurrence of w & c
- $\mathcal{D}$: *(w, c)* multi−set

the length of random walk $L \to \infty$

$$\frac{\#(w,c)}{|\mathcal{D}|} = \frac{1}{2T}\sum_{r=1}^{T}\left(\frac{\#(w,c)_{\overrightarrow{r}}}{|\mathcal{D}_{\overrightarrow{r}}|} + \frac{\#(w,c)_{\overleftarrow{r}}}{|\mathcal{D}_{\overleftarrow{r}}|}\right)$$

$$\frac{\#(w,c)_{\overrightarrow{r}}}{|\mathcal{D}_{\overrightarrow{r}}|} \xrightarrow{p} \frac{d_w}{\mathrm{vol}(G)}\left(\boldsymbol{P}^r\right)_{w,c}$$

$$\frac{\#(w,c)_{\overleftarrow{r}}}{|\mathcal{D}_{\overleftarrow{r}}|} \xrightarrow{p} \frac{d_c}{\mathrm{vol}(G)}\left(\boldsymbol{P}^r\right)_{c,w}$$

$$\frac{\#(w,c)}{|\mathcal{D}|} \xrightarrow{p} \frac{1}{2T}\sum_{r=1}^{T}\left(\frac{d_w}{\mathrm{vol}(G)}\left(\boldsymbol{P}^r\right)_{w,c} + \frac{d_c}{\mathrm{vol}(G)}\left(\boldsymbol{P}^r\right)_{c,w}\right)$$

$$\boldsymbol{P} = \boldsymbol{D}^{-1}\boldsymbol{A}$$

$$\frac{\#(w)}{|\mathcal{D}|} \xrightarrow{p} \frac{d_w}{\mathrm{vol}(G)} \qquad \frac{\#(c)}{|\mathcal{D}|} \xrightarrow{p} \frac{d_c}{\mathrm{vol}(G)}$$

Understanding Random Walk + Skip Gram

$$\log\left(\frac{\#(w,c)\,|\mathcal{D}|}{b\#(w)\cdot\#(c)}\right) = \log\left(\frac{\frac{\#(w,c)}{|\mathcal{D}|}}{b\frac{\#(w)}{|\mathcal{D}|}\frac{\#(c)}{|\mathcal{D}|}}\right)$$

the length of random walk $L \rightarrow \infty$

$$\frac{\#(w,c)}{|\mathcal{D}|} \xrightarrow{p} \frac{1}{2T}\sum_{r=1}^{T}\left(\frac{d_w}{\mathrm{vol}(G)}(\boldsymbol{P}^r)_{w,c} + \frac{d_c}{\mathrm{vol}(G)}(\boldsymbol{P}^r)_{c,w}\right)$$

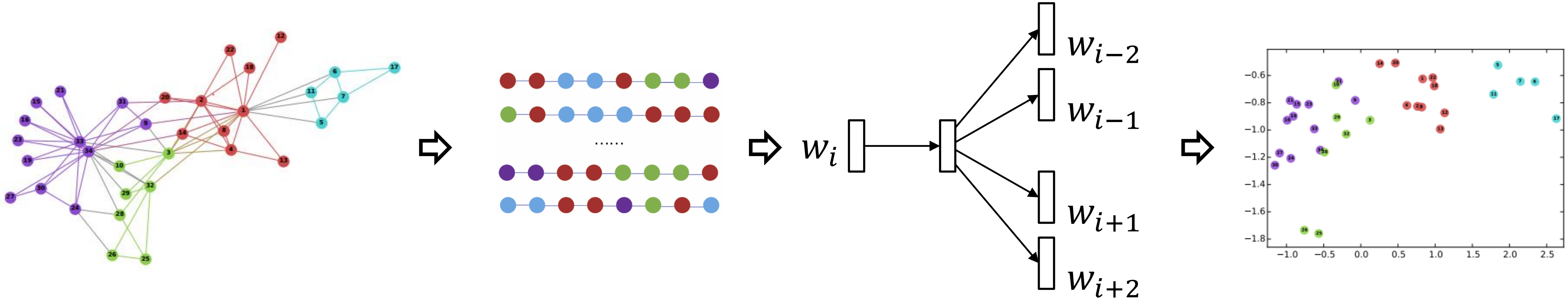
$$\boldsymbol{P} = \boldsymbol{D}^{-1}\boldsymbol{A}$$

$$\frac{\#(w)}{|\mathcal{D}|} \xrightarrow{p} \frac{d_w}{\mathrm{vol}(G)} \qquad \frac{\#(c)}{|\mathcal{D}|} \xrightarrow{p} \frac{d_c}{\mathrm{vol}(G)}$$

$$\frac{\#(w,c)\,|\mathcal{D}|}{\#(w)\cdot\#(c)} = \frac{\frac{\#(w,c)}{|\mathcal{D}|}}{\frac{\#(w)}{|\mathcal{D}|}\cdot\frac{\#(c)}{|\mathcal{D}|}} \xrightarrow{p} \frac{\frac{1}{2T}\sum_{r=1}^{T}\left(\frac{d_w}{\mathrm{vol}(G)}(\boldsymbol{P}^r)_{w,c} + \frac{d_c}{\mathrm{vol}(G)}(\boldsymbol{P}^r)_{c,w}\right)}{\frac{d_w}{\mathrm{vol}(G)}\cdot\frac{d_c}{\mathrm{vol}(G)}}$$

$$= \mathrm{vol}(G)\left(\frac{1}{T}\sum_{r=1}^{T}\boldsymbol{P}^r\right)\boldsymbol{D}^{-1}.$$

Understanding Random Walk + Skip Gram



DeepWalk is asymptotically and implicitly factorizing

$$\log\left(\frac{\mathrm{vol}(G)}{b}\left(\frac{1}{T}\sum_{r=1}^{T}\left(\boldsymbol{D}^{-1}\boldsymbol{A}\right)^{r}\right)\boldsymbol{D}^{-1}\right)$$

$\boldsymbol{A}$ Adjacency matrix

$\boldsymbol{D}$ Degree matrix

$$vol(G) = \sum_i \sum_j A_{ij}$$

b: #negative samples

T: context window size

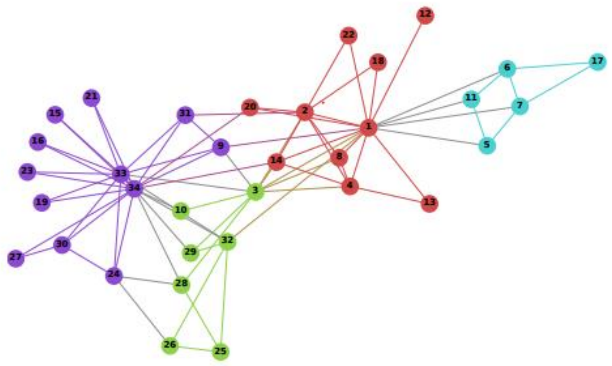
1. Qiu et al. Network embedding as matrix factorization: unifying deepwalk, line, pte, and node2vec. In *WSDM'18.*

Unifying DeepWalk, LINE, PTE, & node2vec as Matrix Factorization

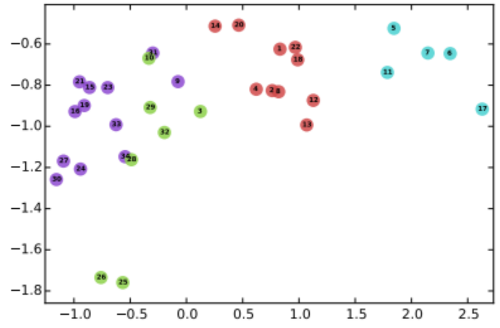
- DeepWalk $$\log\left(\frac{\operatorname{vol}(G)}{b}\left(\frac{1}{T}\sum_{r=1}^{T}\left(\boldsymbol{D}^{-1}\boldsymbol{A}\right)^{r}\right)\boldsymbol{D}^{-1}\right)$$
- LINE $$\log\left(\frac{\operatorname{vol}(G)}{b}\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{D}^{-1}\right)$$
- PTE $$\log\left(\begin{bmatrix}\alpha\operatorname{vol}(G_{\mathsf{ww}})(\boldsymbol{D}_{\mathsf{row}}^{\mathsf{ww}})^{-1}\boldsymbol{A}_{\mathsf{ww}}(\boldsymbol{D}_{\mathsf{col}}^{\mathsf{ww}})^{-1}\\ \beta\operatorname{vol}(G_{\mathsf{dw}})(\boldsymbol{D}_{\mathsf{row}}^{\mathsf{dw}})^{-1}\boldsymbol{A}_{\mathsf{dw}}(\boldsymbol{D}_{\mathsf{col}}^{\mathsf{dw}})^{-1}\\ \gamma\operatorname{vol}(G_{\mathsf{lw}})(\boldsymbol{D}_{\mathsf{row}}^{\mathsf{lw}})^{-1}\boldsymbol{A}_{\mathsf{lw}}(\boldsymbol{D}_{\mathsf{col}}^{\mathsf{lw}})^{-1}\end{bmatrix}\right)-\log b$$
- node2vec $$\log\left(\frac{\frac{1}{2T}\sum_{r=1}^{T}\left(\sum_u \boldsymbol{X}_{w,u}\underline{\boldsymbol{P}}^{r}_{c,w,u}+\sum_u \boldsymbol{X}_{c,u}\underline{\boldsymbol{P}}^{r}_{w,c,u}\right)}{b\left(\sum_u \boldsymbol{X}_{w,u}\right)\left(\sum_u \boldsymbol{X}_{c,u}\right)}\right)$$

Qiu et al. Network embedding as matrix factorization: unifying deepwalk, line, pte, and node2vec. In *WSDM'18.* **The most cited paper in WSDM'18 as of May 2019**

NetMF: explicitly factorizing the DeepWalk matrix



Matrix Factorization



DeepWalk is asymptotically and implicitly factorizing

$$\log\left(\frac{\operatorname{vol}(G)}{b}\left(\frac{1}{T}\sum_{r=1}^{T}\left(\boldsymbol{D}^{-1}\boldsymbol{A}\right)^{r}\right)\boldsymbol{D}^{-1}\right)$$

1. Qiu et al. Network embedding as matrix factorization: unifying deepwalk, line, pte, and node2vec. In *WSDM'18.*

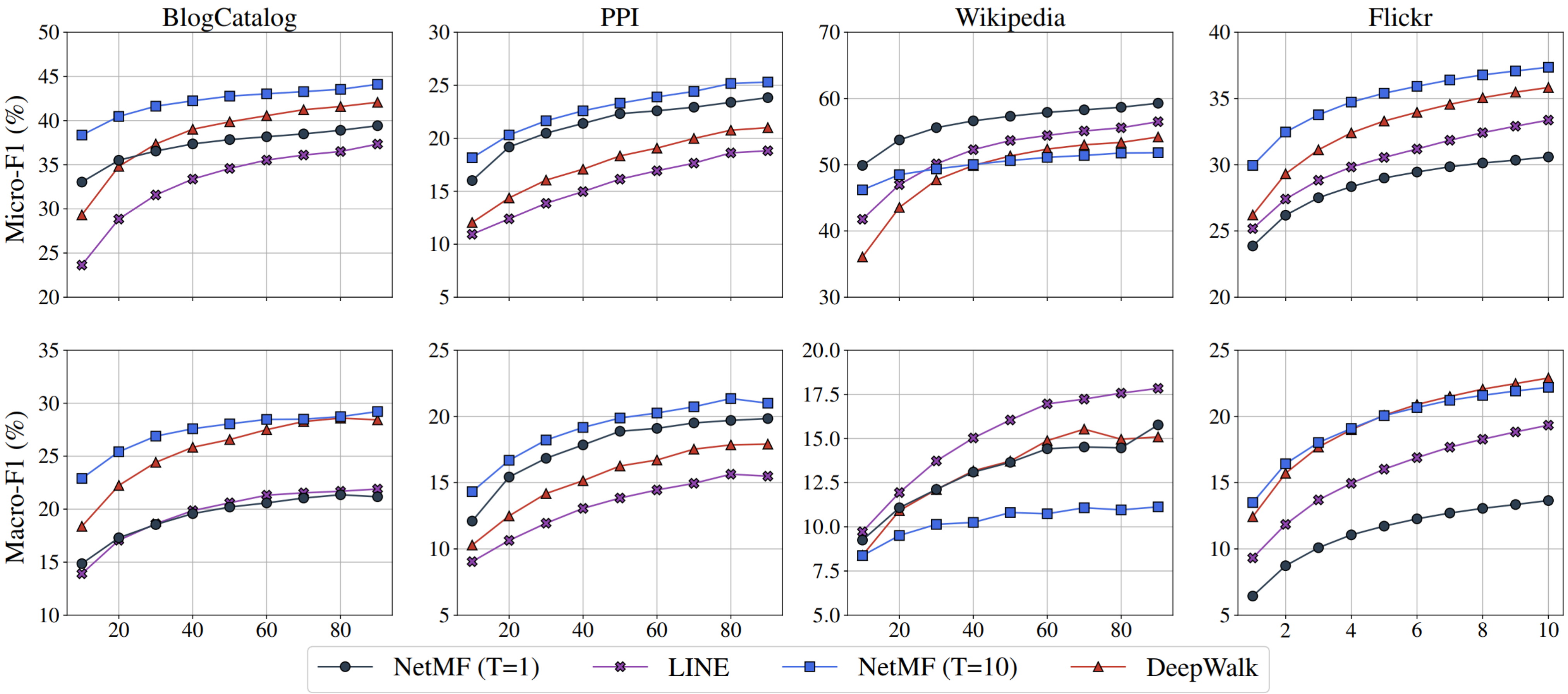
the NetMF algorithm

1. Construction
2. Factorization

$$\boldsymbol{S} = \log\left(\frac{\mathrm{vol}(G)}{b}\left(\frac{1}{T}\sum_{r=1}^{T}\left(\boldsymbol{D}^{-1}\boldsymbol{A}\right)^{r}\right)\boldsymbol{D}^{-1}\right)$$

1. Qiu et al. Network embedding as matrix factorization: unifying deepwalk, line, pte, and node2vec. In *WSDM'18.*

Results



•Predictive performance on varying the ratio of training data;

•The *x*-axis represents the ratio of labeled data (%)

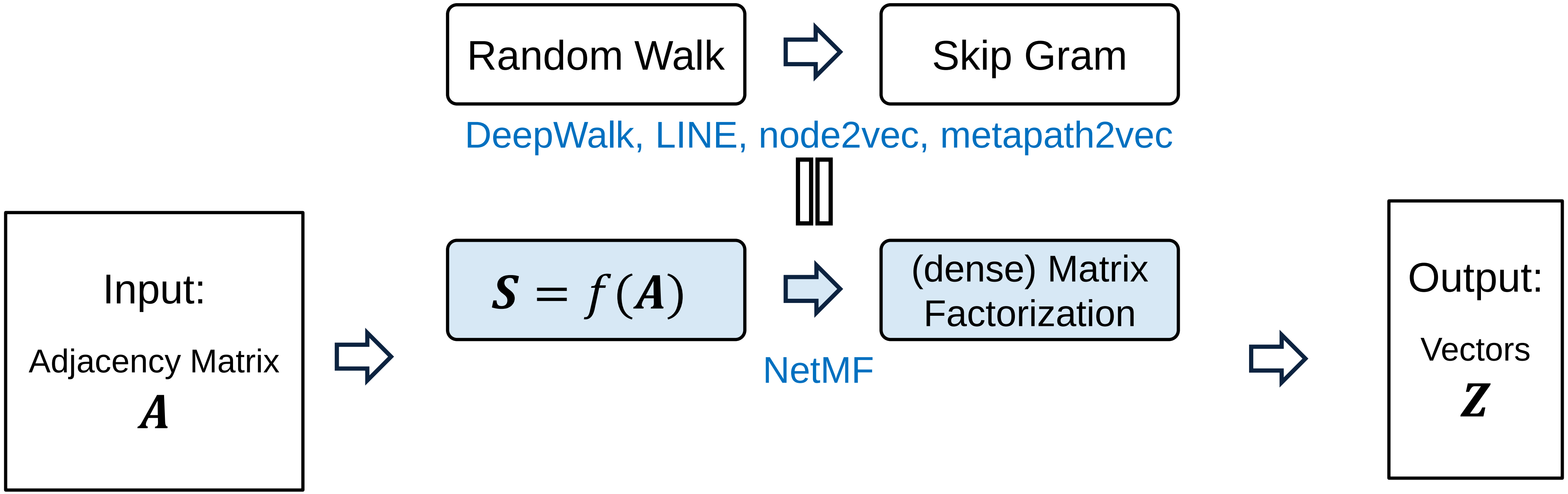
1. Qiu et al. Network embedding as matrix factorization: unifying deepwalk, line, pte, and node2vec. In *WSDM'18.*

Results

Explicit matrix factorization (NetMF) offers performance gains over implicit matrix factorization (DeepWalk & LINE)

1. Qiu et al. Network embedding as matrix factorization: unifying deepwalk, line, pte, and node2vec. In *WSDM'18.*

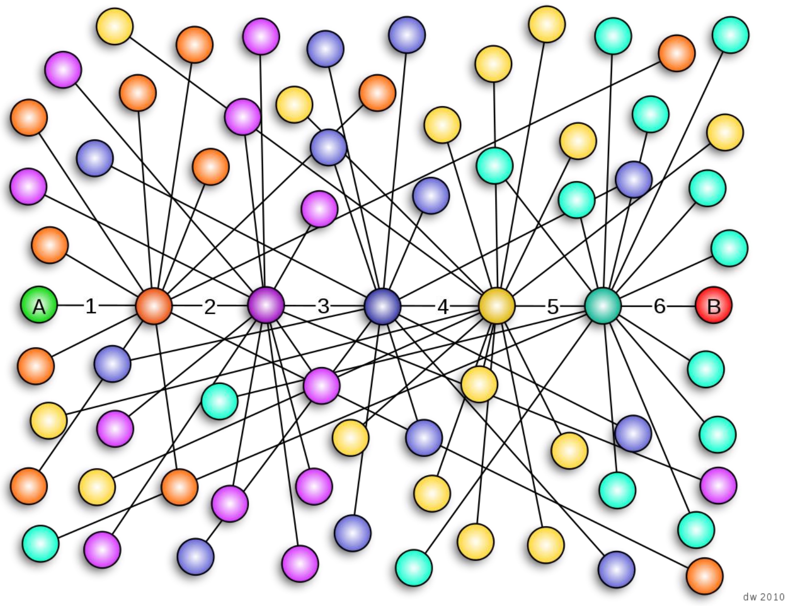
Network Embedding



Incorporate network structures $\boldsymbol{A}$ into the similarity matrix $\boldsymbol{S}$, and then factorize $\boldsymbol{S}$

$$f(\boldsymbol{A}) = \log\left(\frac{\mathrm{vol}(G)}{b}\left(\frac{1}{T}\sum_{r=1}^{T}\left(\boldsymbol{D}^{-1}\boldsymbol{A}\right)^{r}\right)\boldsymbol{D}^{-1}\right)$$

Challenges



⇨ $$\boldsymbol{S} = \log\left(\frac{\mathrm{vol}(G)}{b}\left(\frac{1}{T}\sum_{r=1}^{T}\left(\boldsymbol{D}^{-1}\boldsymbol{A}\right)^{r}\right)\boldsymbol{D}^{-1}\right)$$

dense

NetMF is not practical for very large networks

NetMF

How can we solve this issue?

1. Construction
2. Factorization

$$\boldsymbol{S} = \log\left(\frac{\mathrm{vol}(G)}{b}\left(\frac{1}{T}\sum_{r=1}^{T}\left(\boldsymbol{D}^{-1}\boldsymbol{A}\right)^{r}\right)\boldsymbol{D}^{-1}\right)$$

1. Qiu et al. NetSMF: Network embedding as sparse matrix factorization. In **WWW 2019**

NetSMF--Sparse

How can we solve this issue?

1. **Sparse** Construction
2. **Sparse** Factorization

$$\boldsymbol{S} = \log\left(\frac{\mathrm{vol}(G)}{b}\left(\frac{1}{T}\sum_{r=1}^{T}\left(\boldsymbol{D}^{-1}\boldsymbol{A}\right)^{r}\right)\boldsymbol{D}^{-1}\right)$$

1. Qiu et al. NetSMF: Network embedding as sparse matrix factorization. In **WWW 2019**

Sparsify $\boldsymbol{S}$

For random-walk matrix polynomial $\boldsymbol{L} = \boldsymbol{D} - \sum_{r=1}^{T} \alpha_r \boldsymbol{D} \left(\boldsymbol{D}^{-1}\boldsymbol{A}\right)^r$

where $\sum_{r=1}^{T} \alpha_r = 1$ and α_r non-negative

One can construct a $(1+\epsilon)$-spectral sparsifier $\tilde{\boldsymbol{L}}$ with $O(n \log n \epsilon^{-2})$ non-zeros

in time $O(T^2 m \epsilon^{-2} \log^2 n)$

$O(T^2 m \epsilon^{-2} \log n)$ for undirected graphs

Suppose $G = (V, E, \boldsymbol{A})$ and $\widetilde{G} = (V, \widetilde{E}, \widetilde{\boldsymbol{A}})$ are two weighted undirected networks. Let $\boldsymbol{L} = \boldsymbol{D}_G - \boldsymbol{A}$ and $\widetilde{\boldsymbol{L}} = \boldsymbol{D}_{\widetilde{G}} - \widetilde{\boldsymbol{A}}$ be their Laplacian matrices, respectively. We define G and $\widetilde{G}$ are $(1+\epsilon)$-spectrally similar if

$$\forall \boldsymbol{x} \in \mathbb{R}^n, (1-\epsilon) \cdot \boldsymbol{x}^\top \widetilde{\boldsymbol{L}} \boldsymbol{x} \leq \boldsymbol{x}^\top \boldsymbol{L} \boldsymbol{x} \leq (1+\epsilon) \cdot \boldsymbol{x}^\top \widetilde{\boldsymbol{L}} \boldsymbol{x}.$$

- Dehua Cheng, Yu Cheng, Yan Liu, Richard Peng, and Shang-Hua Teng, Efficient Sampling for Gaussian Graphical Models via Spectral Sparsification, COLT 2015.
- Dehua Cheng, Yu Cheng, Yan Liu, Richard Peng, and Shang-Hua Teng. Spectral sparsification of random-walk matrix polynomials. arXiv:1502.03496.

Sparsify $\boldsymbol{S}$

For random-walk matrix polynomial $\boldsymbol{L} = \boldsymbol{D} - \sum_{r=1}^{T} \alpha_r \boldsymbol{D} \left(\boldsymbol{D}^{-1}\boldsymbol{A}\right)^r$

where $\sum_{r=1}^{T} \alpha_r = 1$ and α_r non-negative

One can construct a $(1+\epsilon)$-spectral sparsifier $\tilde{\boldsymbol{L}}$ with $O(n \log n \epsilon^{-2})$ non-zeros

in time $O(T^2 m \epsilon^{-2} \log^2 n)$

$$\boldsymbol{S} = \log^{\circ}\left(\frac{\text{vol}(G)}{b}\left(\frac{1}{T}\sum_{r=1}^{T}\left(\boldsymbol{D}^{-1}\boldsymbol{A}\right)^r\right)\boldsymbol{D}^{-1}\right)$$

$\alpha_1 = \cdots = \alpha_T = \frac{1}{T}$ ⇨

$$= \log^{\circ}\left(\frac{\text{vol}(G)}{b}\boldsymbol{D}^{-1}(\boldsymbol{D}-\boldsymbol{L})\boldsymbol{D}^{-1}\right)$$

$$\approx \log^{\circ}\left(\frac{\text{vol}(G)}{b}\boldsymbol{D}^{-1}(\boldsymbol{D}-\widetilde{\boldsymbol{L}})\boldsymbol{D}^{-1}\right)$$

1. Qiu et al. NetSMF: Network embedding as sparse matrix factorization. In **WWW 2019**

NetSMF --- Sparse

- Construct a random walk matrix polynomial sparsifier, $\widetilde{\boldsymbol{L}}$
- Construct a NetMF matrix sparsifier.

$$\operatorname{trunc\_log}^{\circ}\left(\frac{\operatorname{vol}(G)}{b}\boldsymbol{D}^{-1}(\boldsymbol{D}-\widetilde{\boldsymbol{L}})\boldsymbol{D}^{-1}\right)$$

- Factorize the constructed *sparse* matrix

	Time	Space
Step 1	$O(MT \log n)$ for weighted networks $O(MT)$ for unweighted networks	$O(M+n+m)$
Step 2	$O(M)$	$O(M+n)$
Step 3	$O(Md+nd^2+d^3)$	$O(M+nd)$

1. Qiu et al. NetSMF: Network embedding as sparse matrix factorization. In **WWW 2019**

NetSMF---bounded approximation error

$$\log^\circ\left(\frac{\operatorname{vol}(G)}{b}\left(\frac{1}{T}\sum_{r=1}^{T}\left(\boldsymbol{D}^{-1}\boldsymbol{A}\right)^r\right)\boldsymbol{D}^{-1}\right)$$

$$=\log^\circ\left(\frac{\operatorname{vol}(G)}{b}\boldsymbol{D}^{-1}(\boldsymbol{D}-\boldsymbol{L})\boldsymbol{D}^{-1}\right) \rightarrow \boldsymbol{M}$$

$$\approx\log^\circ\left(\frac{\operatorname{vol}(G)}{b}\boldsymbol{D}^{-1}(\boldsymbol{D}-\widetilde{\boldsymbol{L}})\boldsymbol{D}^{-1}\right) \rightarrow \widetilde{\boldsymbol{M}}$$

Theorem

The singular value of $\widetilde{\boldsymbol{M}}-\boldsymbol{M}$ satisfies

$$\sigma_i(\widetilde{\boldsymbol{M}}-\boldsymbol{M}) \leq \frac{4\epsilon}{\sqrt{d_i d_{\min}}}, \forall i \in [n].$$

Theorem

Let $\|\cdot\|_F$ be the matrix Frobenius norm. Then

$$\left\|\operatorname{trunc\_log}^\circ\left(\frac{\operatorname{vol}(G)}{b}\widetilde{\boldsymbol{M}}\right)-\operatorname{trunc\_log}^\circ\left(\frac{\operatorname{vol}(G)}{b}\boldsymbol{M}\right)\right\|_F \leq \frac{4\epsilon\operatorname{vol}(G)}{b\sqrt{d_{\min}}}\sqrt{\sum_{i=1}^{n}\frac{1}{d_i}}.$$

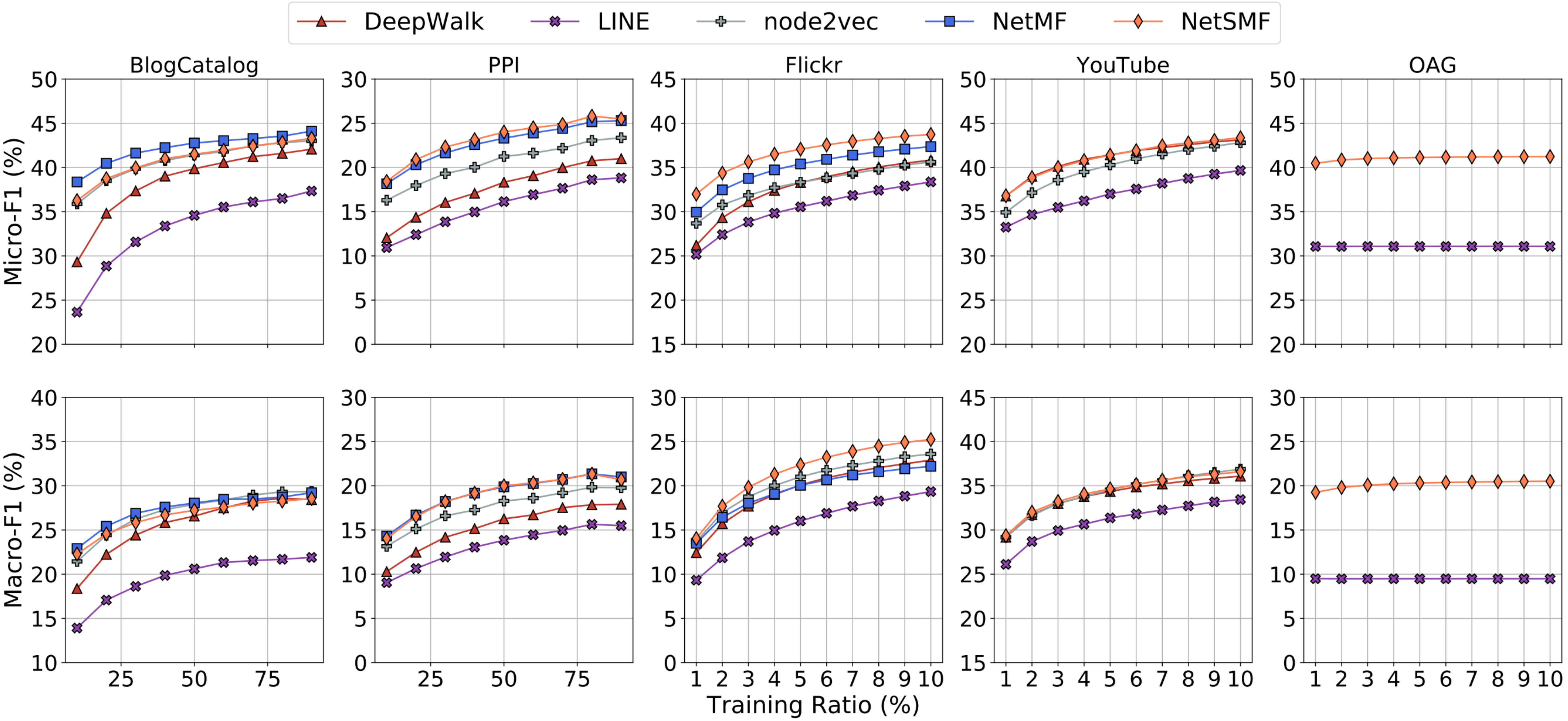
1. Qiu et al. NetSMF: Network embedding as sparse matrix factorization. In **WWW 2019**

Dataset	BlogCatalog	PPI	Flickr	YouTube	OAG
$\|V\|$	10,312	3,890	80,513	1,138,499	67,768,244
$\|E\|$	333,983	76,584	5,899,882	2,990,443	895,368,962
#labels	39	50	195	47	19

#non-zeros

~4.5 Quadrillion → 45 Billion

1. Qiu et al. NetSMF: Network embedding as sparse matrix factorization. In **WWW 2019**



1. Qiu et al. NetSMF: Network embedding as sparse matrix factorization. In **WWW 2019**

Effectiveness:

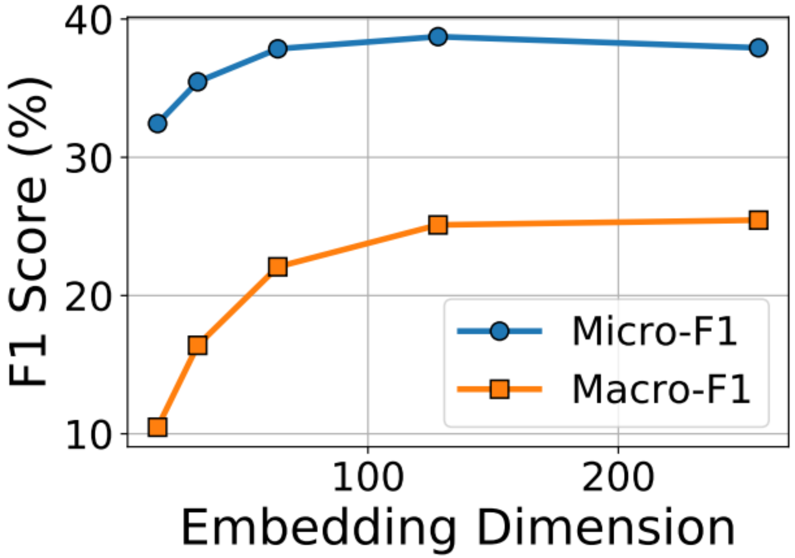
- (sparse MF)NetSMF **≈** (explicit MF)NetMF **>** (implicit MF) DeepWalk/LINE

Efficiency:

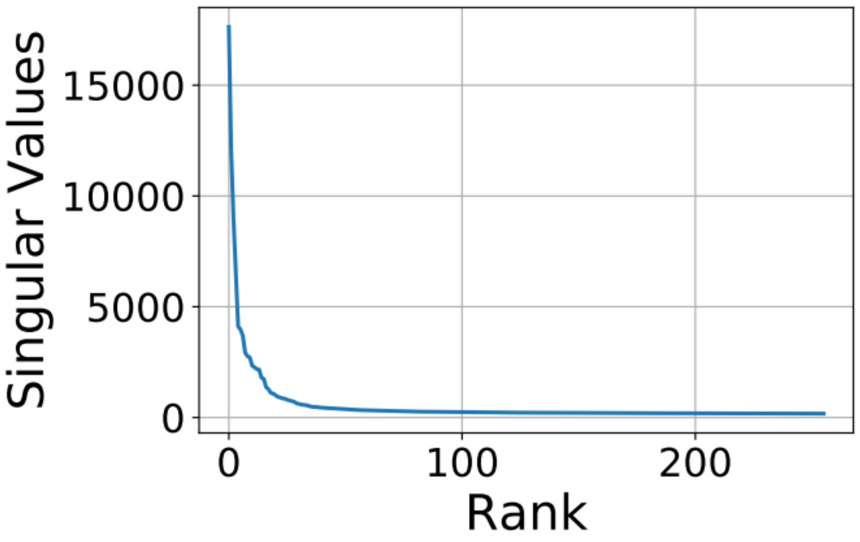
- Sparse MF can handle billion-scale network embedding

1. Qiu et al. NetSMF: Network embedding as sparse matrix factorization. In **WWW 2019**

Embedding Dimension?



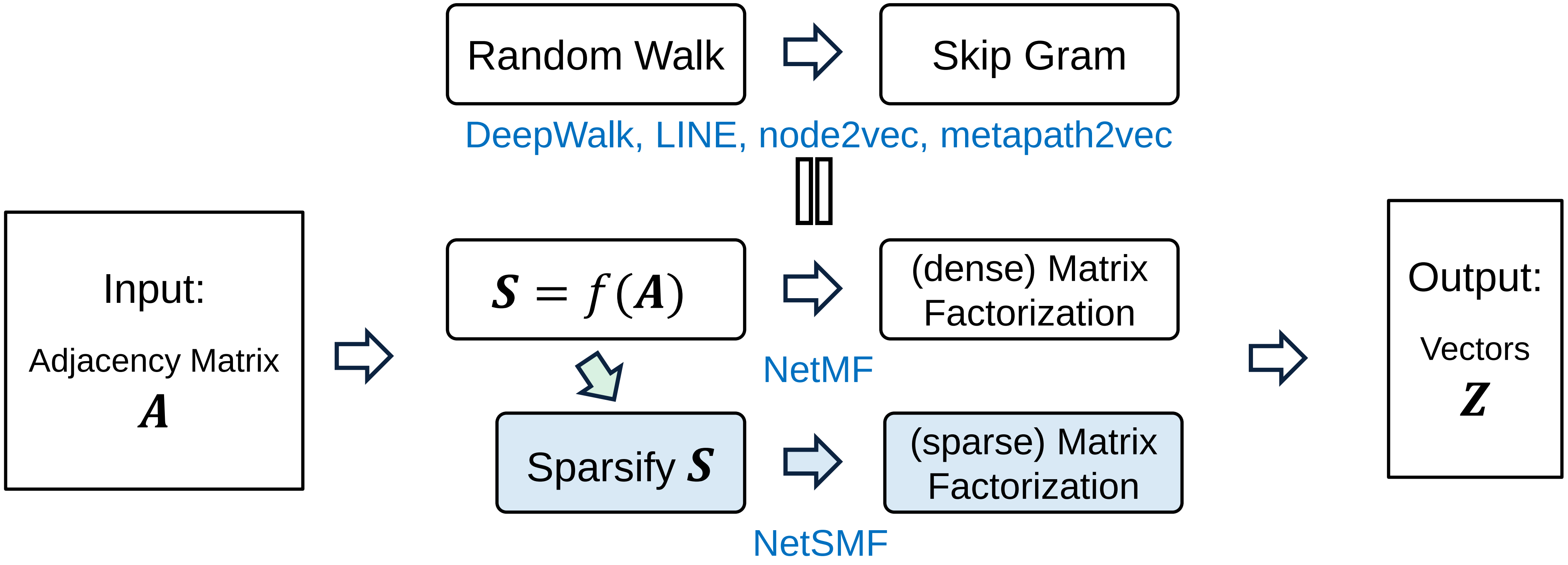
(a)



(b)

1. Qiu et al. NetSMF: Network embedding as sparse matrix factorization. In **WWW 2019**

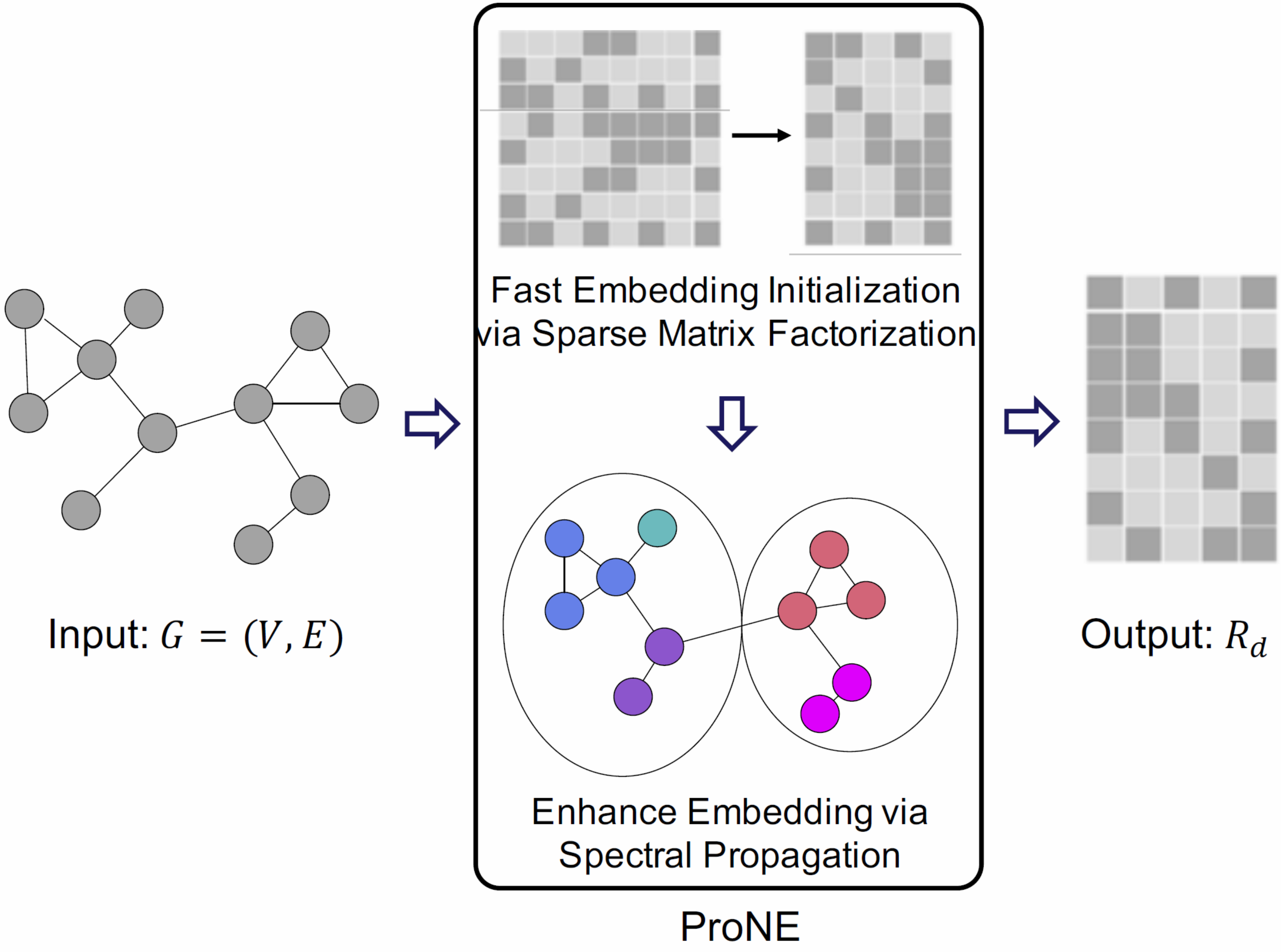
Network Embedding



Incorporate network structures $\boldsymbol{A}$ into the similarity matrix $\boldsymbol{S}$, and then factorize $\boldsymbol{S}$

$$f(\boldsymbol{A}) = \log\left(\frac{\mathrm{vol}(G)}{b}\left(\frac{1}{T}\sum_{r=1}^{T}\left(\boldsymbol{D}^{-1}\boldsymbol{A}\right)^{r}\right)\boldsymbol{D}^{-1}\right)$$

ProNE: More fast & scalable network embedding



1. Zhang et al. ProNE: Fast and Scalable Network Representation Learning. In **IJCAI 2019**

Embedding enhancement via spectral propagation

$$R_d \leftarrow D^{-1}A(I_n - \tilde{L})\, R_d$$

$\widetilde{L} = U g(\Lambda) U^T$ is the spectral filter of $L = I_n - D^{-1}A$

$D^{-1}A(I_n - \tilde{L})$ is $D^{-1}A$ modulated by the filter in the spectrum

The idea of **Graph Neural Networks**

1. Zhang et al. ProNE: Fast and Scalable Network Representation Learning. In **IJCAI 2019**

Performance

20 Threads (DeepWalk, LINE, node2vec) — 1 Thread (ProNE)

Dataset	*DeepWalk*	*LINE*	*node2vec*	*ProNE*
PPI	272	70	828	**3**
Wiki	494	87	939	**6**
BlogCatalog	1,231	185	3,533	**21**
DBLP	3,825	1,204	4,749	**24**
Youtube	68,272	5,890	>5days	**627**
	19hours	98mins		10mins

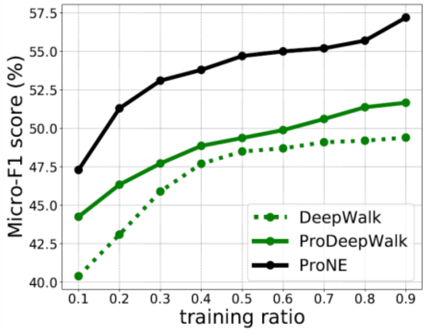
Dataset	training ratio	0.01	0.03	0.05	0.07	0.09
DBLP	DeepWalk	49.3	55.0	57.1	57.9	58.4
	LINE	48.7	52.6	53.5	54.1	54.5
	node2vec	48.9	55.1	57.0	58.0	58.4
	GraRep	50.5	52.6	53.2	53.5	53.8
	HOPE	**52.2**	55.0	55.9	56.3	56.6
	ProNE (SMF)	50.8	54.9	56.1	56.7	57.0
	ProNE	48.8	**56.2**	**58.0**	**58.8**	**59.2**
	($\pm\sigma$)	($\pm$1.0)	($\pm$0.5)	($\pm$0.2)	($\pm$0.2)	($\pm$0.1)
Youtube	DeepWalk	38.0	40.1	41.3	42.1	42.8
	LINE	33.2	35.5	37.0	38.2	39.3
	ProNE (SMF)	36.5	40.2	41.2	41.7	42.1
	ProNE	**38.2**	**41.4**	**42.3**	**42.9**	**43.3**
	($\pm\sigma$)	($\pm$0.8)	($\pm$0.3)	($\pm$0.2)	($\pm$0.2)	($\pm$0.2)

ProNE offers **10-400X** speedups
(**1** thread vs **20** threads)

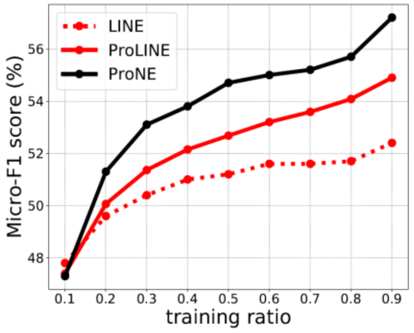
ProNE embeds **100,000,000** nodes by **1** thread:
29 hours with performance superiority

1. Zhang et al. ProNE: Fast and Scalable Network Representation Learning. In **IJCAI 2019**

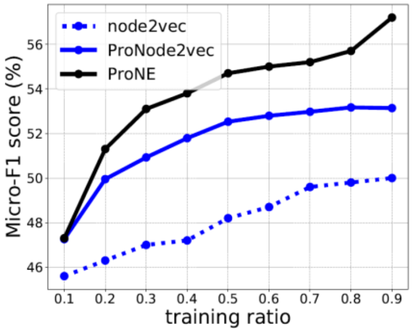
A general embedding enhancement framework



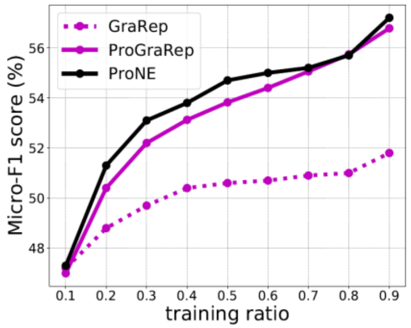
(a) ProDeepWalk



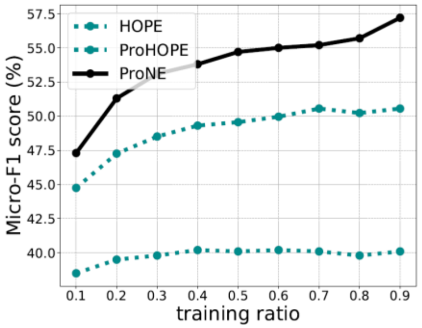
(b) ProLINE



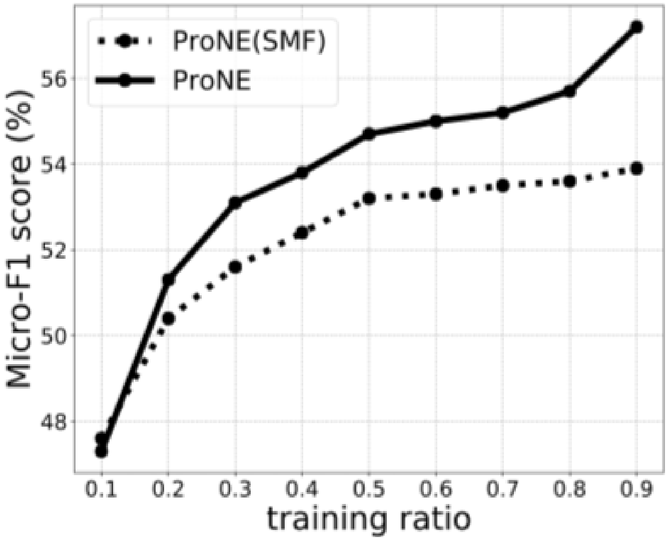
(c) ProNode2vec



(d) ProGraRep

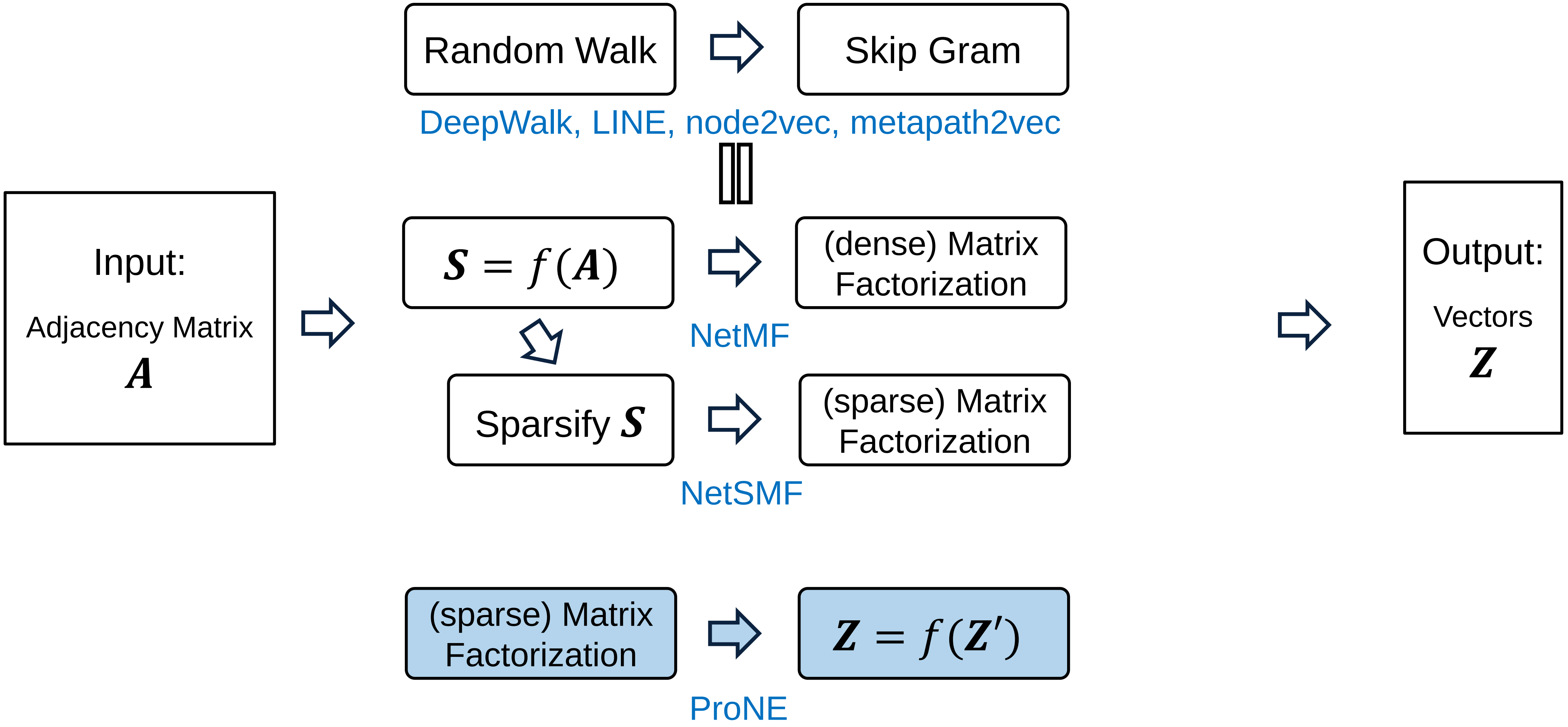


(e) ProHOPE



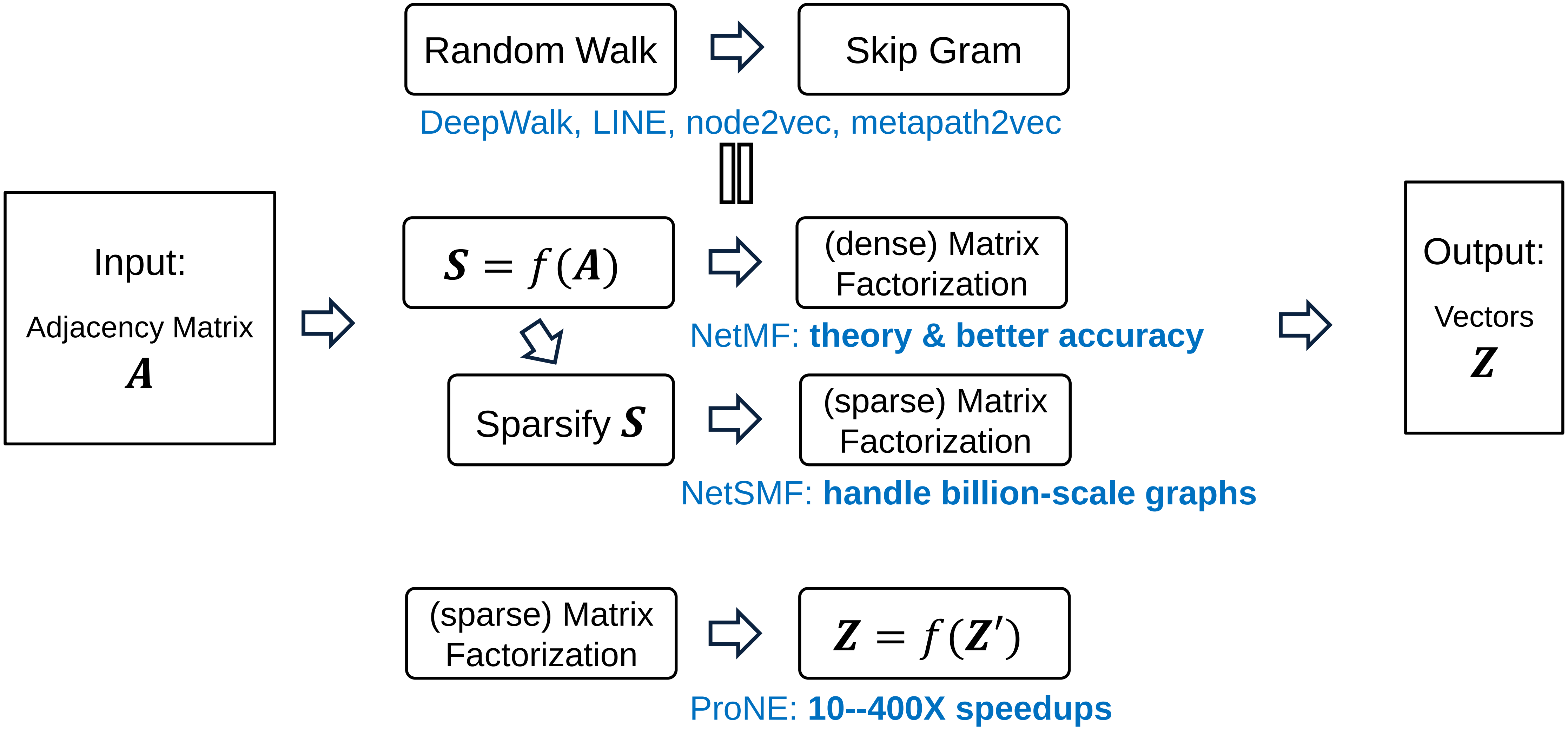
1. Zhang et al. ProNE: Fast and Scalable Network Representation Learning. In **IJCAI 2019**

Network Embedding



Factorize A, and then incorporate network structures via spectral propagation

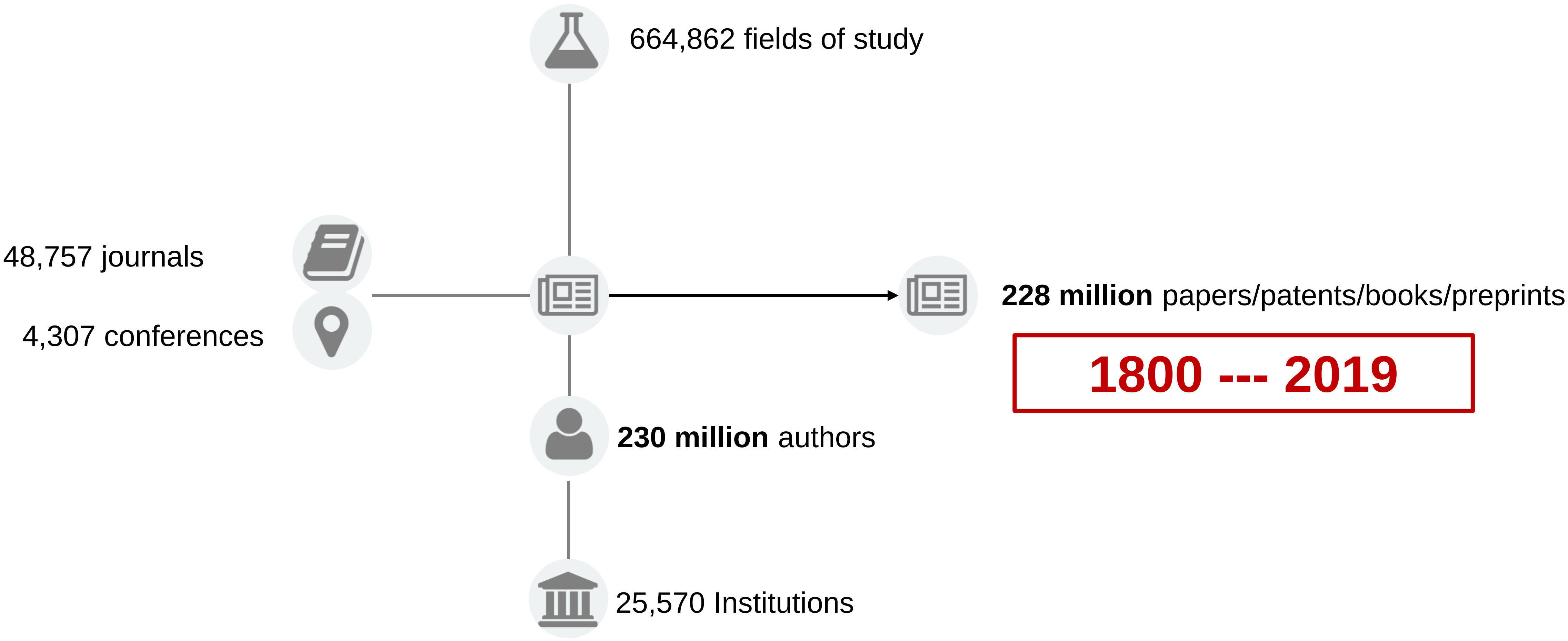
Network Embedding



References

1. Jiezhong Qiu, Yuxiao Dong, Hao Ma, Jian Li, Chi Wang, Kuansan Wang, and Jie Tang. *NetSMF: Large-Scale Network Embedding as Sparse Matrix Factorization*. **WWW** 2019.
2. Jie Zhang, Yuxiao Dong, Yan Wang, Jie Tang, and Ming Ding. *ProNE: Fast and Scalable Network Representation Learning*. **IJCAI** 2019.
3. Jiezhong Qiu, Yuxiao Dong, Hao Ma, Jian Li, Kuansan Wang, and Jie Tang. *Network Embedding as Matrix Factorization: Unifying DeepWalk, LINE, PTE, and node2vec.* **WSDM** 2018.
4. Yuxiao Dong, Nitesh V. Chawla, Ananthram Swami. *metapath2vec: Scalable Representation Learning for Heterogeneous Networks*. **KDD** 2017.
5. Fanjing Zhang, et al. OAG: Toward Linking Large-scale Heterogeneous Entity Graphs. ACM ***KDD*** 2019.
6. Wu, Shi, Dong, Huang, Chawla. Neural Tensor Decomposition. **WSDM** 2019.

Microsoft Academic Graph



https://academic.microsoft.com as of Sep. 2019
The graph data is open!

Thank you!

Papers & data & code available at https://ericdongyx.github.io/

ericdongyx@gmail.com

Joint work with Jiezhong Qiu, Jie Zhang, Jie Tang (Tsinghua University)
Hao Ma (MSR & Facebook AI) and Kuansan Wang (MSR)