Module III: Graph Representation Learning

11:15 am - 12:00 pm

Module 3 Overview
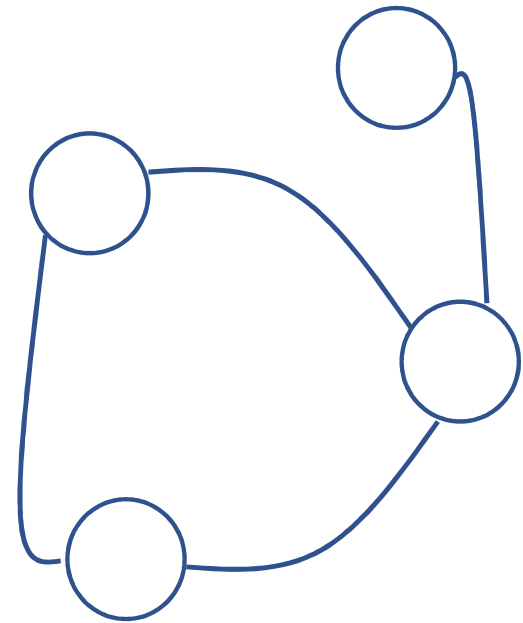
Graph representation learning

What and why

A brief history of Graph Embedding

- Skip-gram based
- Graph Neural Networks (GNN)
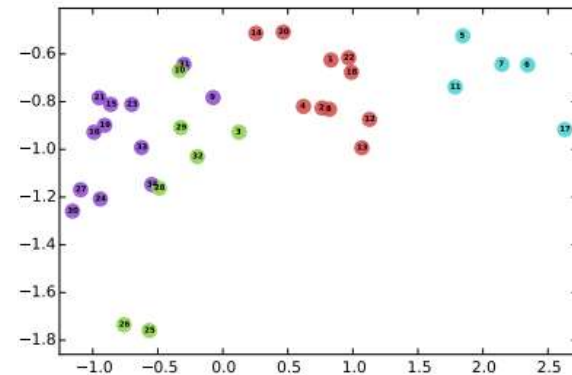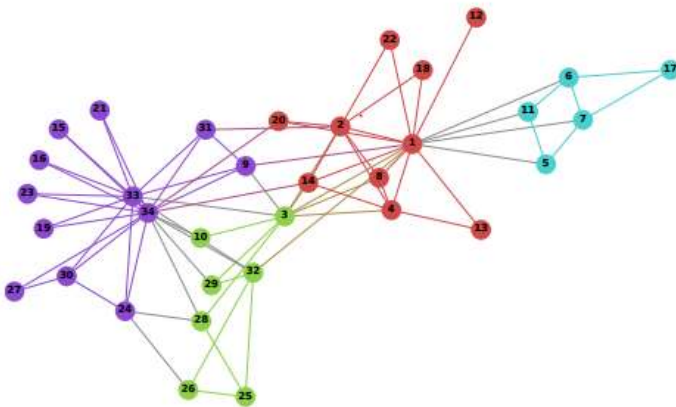
Lab 3 – Multi-sense similarity

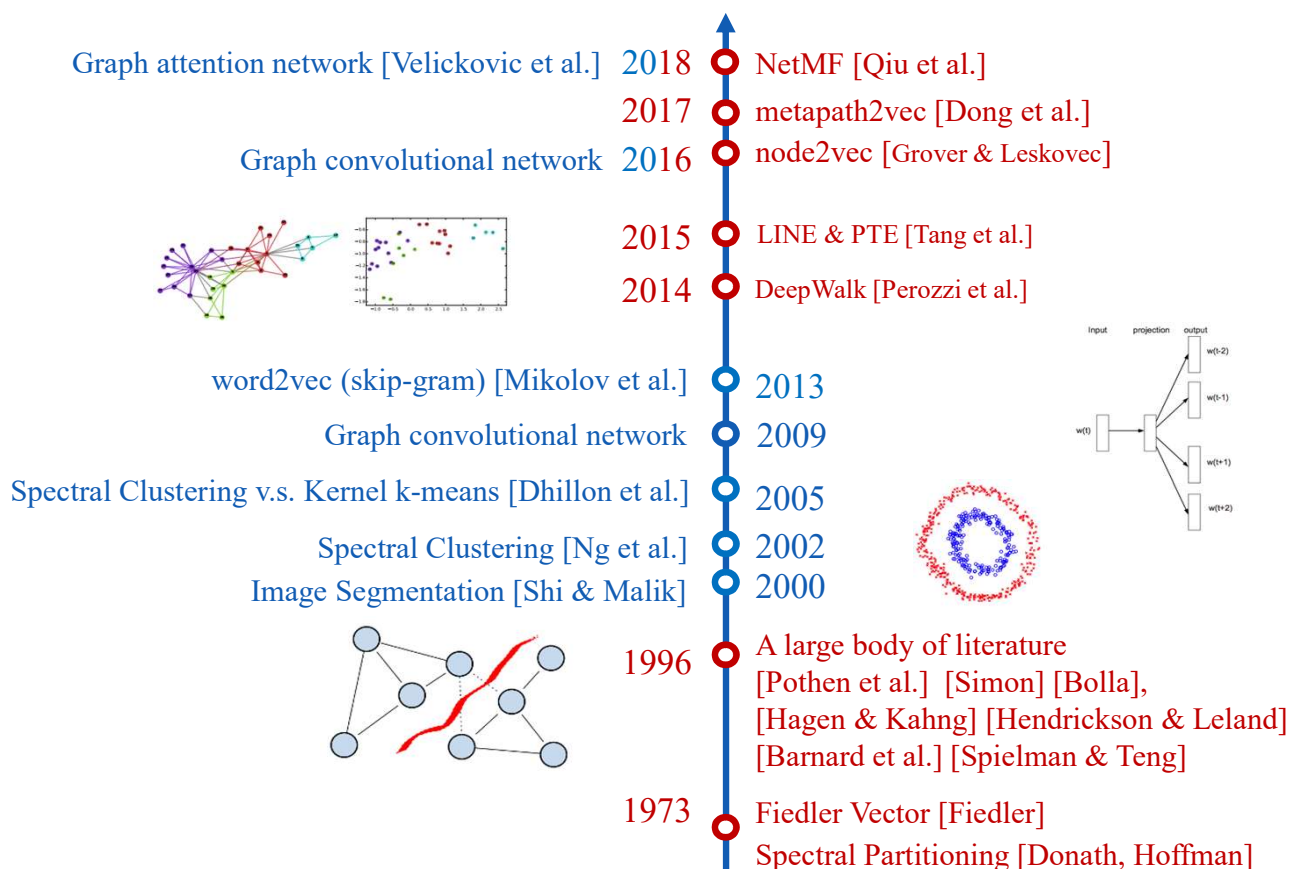# What and Why - Graph Representation learning

Problem (Graph representation learning, network embedding, graph embedding)

- Input: a network $G = (V, E)$

- Output: $X \in R^{|V| \times k}$, $k \ll |V|$, k-dim vector X_v for each node v.
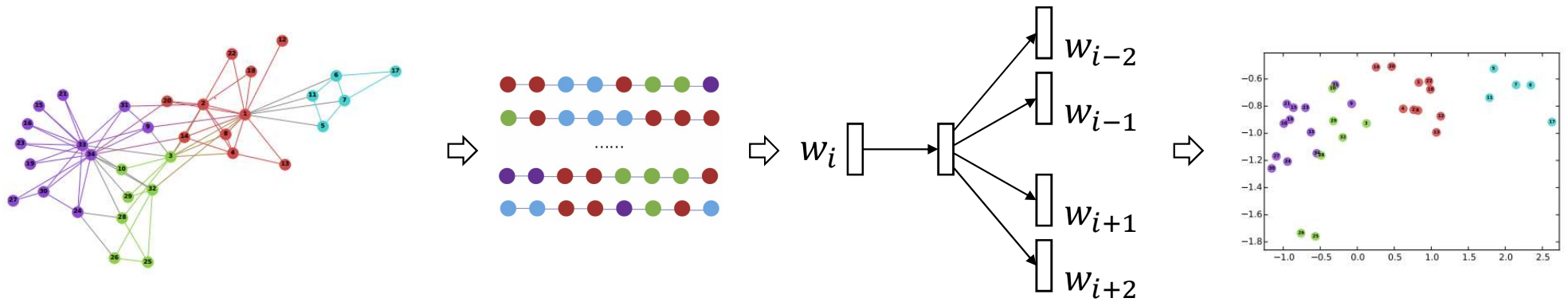
Each node ➔ a latent low-dimension vector,
network structure information (encoded into) ➔ distributed node representations
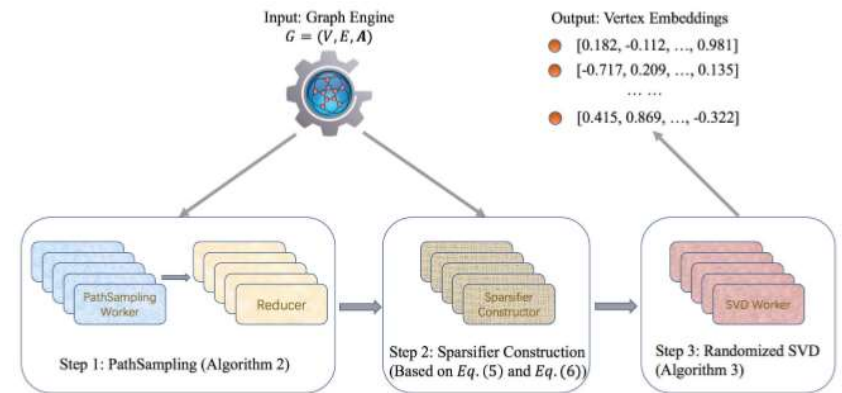
A brief history of graph embedding

Graph attention network [Velickovic et al.] 2018 ○ NetMF [Qiu et al.]

2017 ○ metapath2vec [Dong et al.]

Graph convolutional network 2016 ○ node2vec [Grover & Leskovec]

2015 ○ LINE & PTE [Tang et al.]

2014 ○ DeepWalk [Perozzi et al.]

word2vec (skip-gram) [Mikolov et al.] ○ 2013

Graph convolutional network ○ 2009

Spectral Clustering v.s. Kernel k-means [Dhillon et al.] ○ 2005

Spectral Clustering [Ng et al.] ○ 2002

Image Segmentation [Shi & Malik] ○ 2000

1996 ○ A large body of literature
[Pothen et al.] [Simon] [Bolla],
[Hagen & Kahng] [Hendrickson & Leland]
[Barnard et al.] [Spielman & Teng]

1973 ○ Fiedler Vector [Fiedler]
Spectral Partitioning [Donath, Hoffman]

Skip-gram based graph embedding

Skip-gram based graph embedding

Matrix Factorization

NetMF

NetSMF

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

- LINE $\quad\quad \log\left(\dfrac{\mathrm{vol}(G)}{b}\boldsymbol{D}^{-1}\boldsymbol{A}\boldsymbol{D}^{-1}\right)$

- PTE $\quad\quad\;\; \log\left(\begin{bmatrix}\alpha\,\mathrm{vol}(G_{\mathrm{ww}})(\boldsymbol{D}_{\mathrm{row}}^{\mathrm{ww}})^{-1}\boldsymbol{A}_{\mathrm{ww}}(\boldsymbol{D}_{\mathrm{col}}^{\mathrm{ww}})^{-1}\\\beta\,\mathrm{vol}(G_{\mathrm{dw}})(\boldsymbol{D}_{\mathrm{row}}^{\mathrm{dw}})^{-1}\boldsymbol{A}_{\mathrm{dw}}(\boldsymbol{D}_{\mathrm{col}}^{\mathrm{dw}})^{-1}\\\gamma\,\mathrm{vol}(G_{\mathrm{lw}})(\boldsymbol{D}_{\mathrm{row}}^{\mathrm{lw}})^{-1}\boldsymbol{A}_{\mathrm{lw}}(\boldsymbol{D}_{\mathrm{col}}^{\mathrm{lw}})^{-1}\end{bmatrix}\right)-\log b$

- node2vec $\quad \log\left(\dfrac{\frac{1}{2T}\sum_{r=1}^{T}\left(\sum_{u}\boldsymbol{X}_{w,u}\underline{\boldsymbol{P}}_{c,w,u}^{r}+\sum_{u}\boldsymbol{X}_{c,u}\underline{\boldsymbol{P}}_{w,c,u}^{r}\right)}{b\left(\sum_{u}\boldsymbol{X}_{w,u}\right)\left(\sum_{u}\boldsymbol{X}_{c,u}\right)}\right)$

Heterogeneous Skip-gram graph embedding
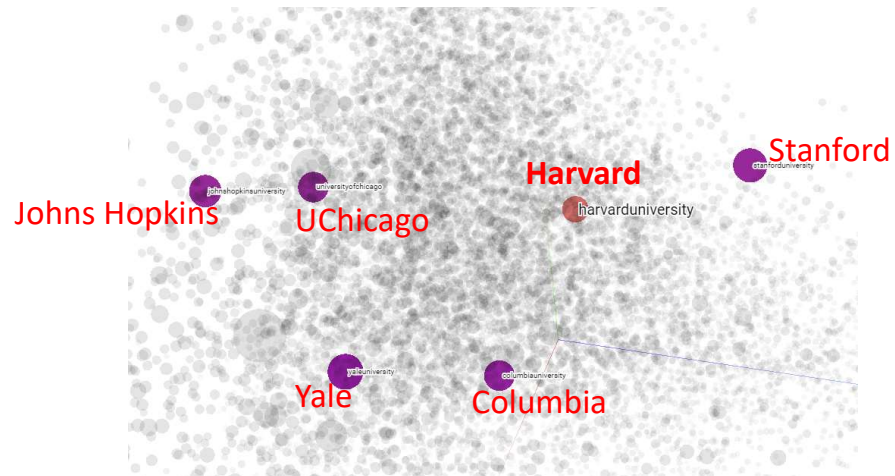
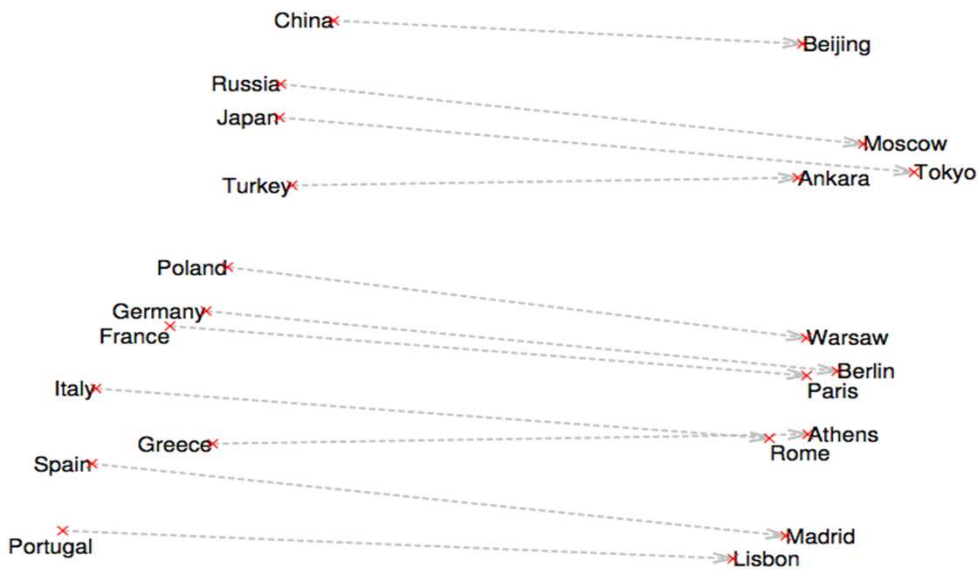

meta paths

APA

APVPA

OAPVPAO

meta-path-based random walks

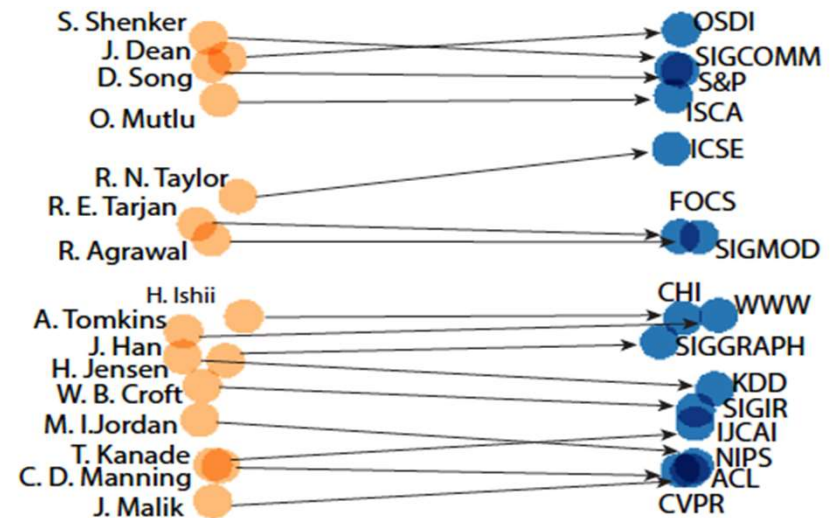heterogeneous skip-gram
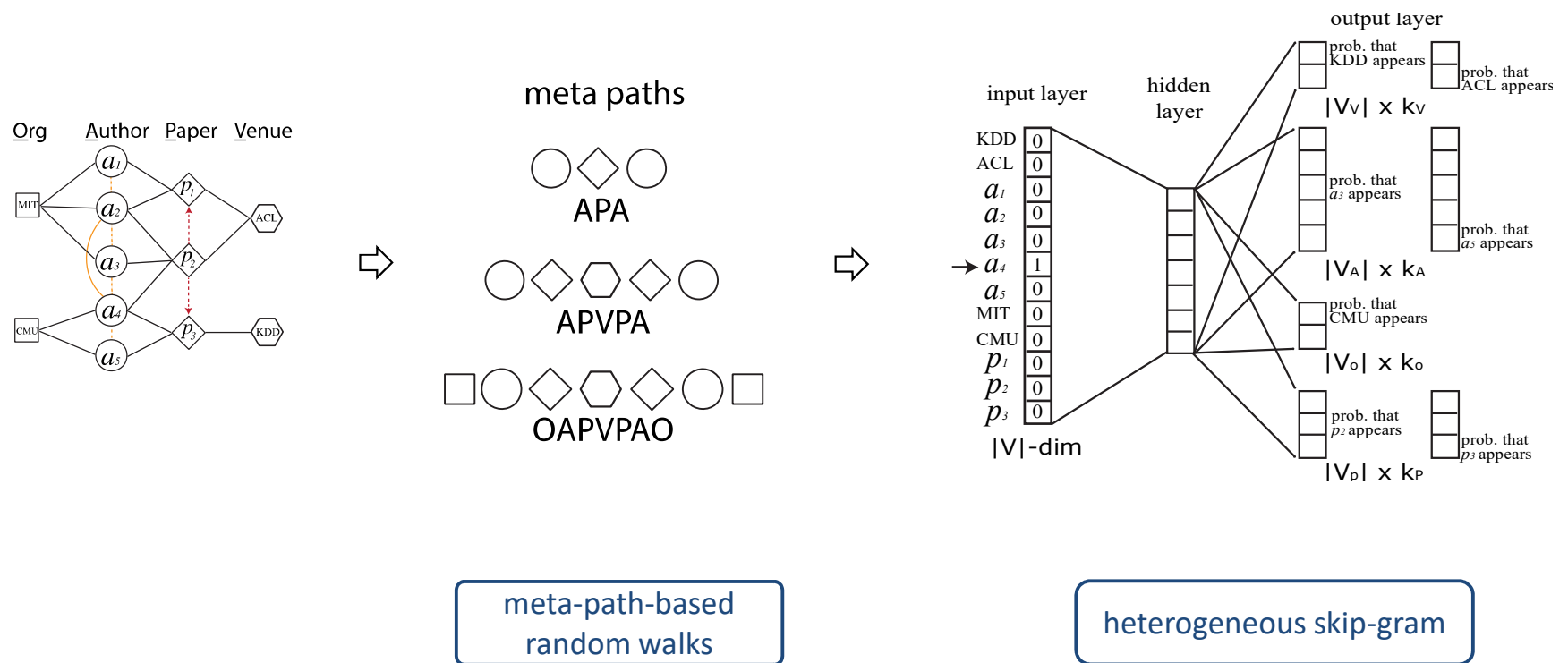
Embedding Heterogeneous Academic Graph
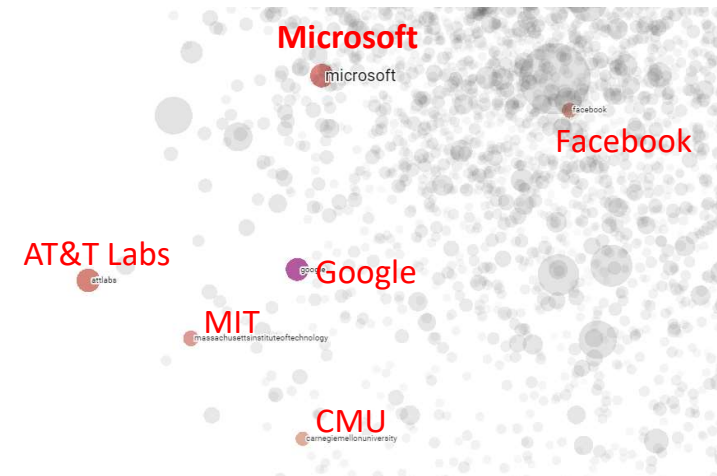
Embedding Heterogeneous Academic Graph

word2vec [Mikolov, 2013]

Metapath2vec++ [Dong et al., 2017]

Lab 3: NetworkSimilarity in MAG

- Task : who are the most similar ones to me
- Homogeneous Network Embedding using NetSMF

  - Affiliation – affiliation
  - Venue – venue
  - Field of study - field of study

Lab 3: NetworkSimilarity in MAG

- GitHub Repository
  https://github.com/graph-knowledgegraph/KDD2019-HandsOn-Tutorial

- Import Databricks notebook
  - 3.NetworkSimilarityDemo.py (https://github.com/graph-knowledgegraph/KDD2019-HandsOn-Tutorial/blob/master/Module_III/3.NetworkSimilarityDemo.py)

- Run 3.NetworkSimilarityDemo