Fanghua Ye

School of Data and Computer Science

Sun Yat-sen University

Introduction

- **Education**

- **Research Interests**
  - Social network analysis, data mining, machine learning

- **Publications**
  - SIGMOD, ICDM, CIKM, TNNLS, Information Sciences

Research Focus

Graph Data Mining

Community Detection

Influential Nodes

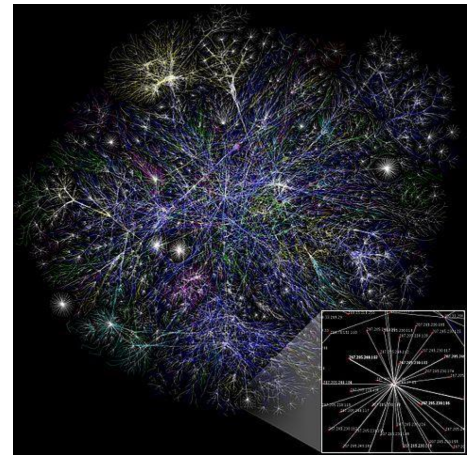
Network Embedding

Complex Network

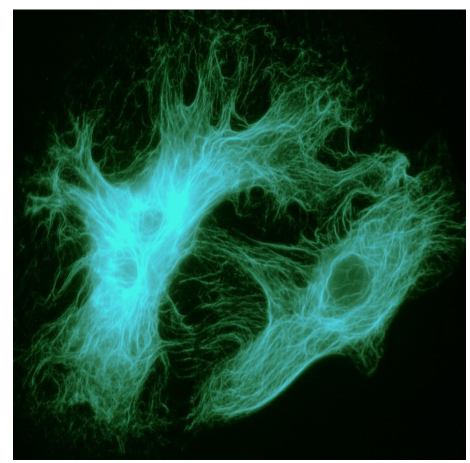
- **Complex network is everywhere**

Social

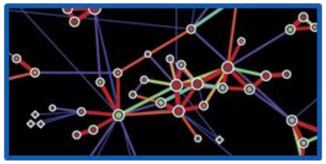
Internet



Protein



History



Computational Social Science [Giles] — 2012

Computational Social Science [Lazer et al.] — 2009

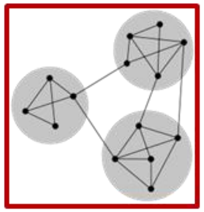
2009 — Social Influence Analysis [Tang, Sun]

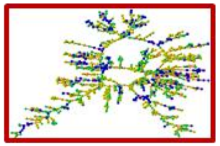
2007 — Spread of Obesity, Happiness [Christakis, Fowler]

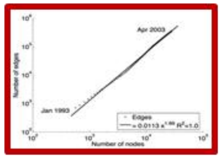
2005 — Densification [Leskovec, Kleinberg, Faloutsos]

2003 — Link Prediction [Liben-Nowell, Kleinberg]
Influence Maximization [Kempe, Kleinberg, Tardos]

2002 — Community Detection [Girvan, Newman]



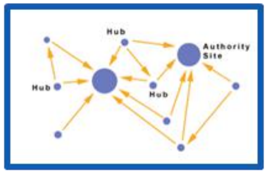


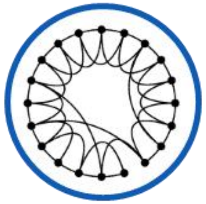


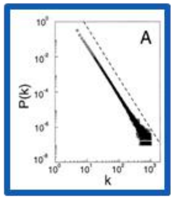
Scale Free [Barabási, Albert & Faloutsos et al.] — 1999

Small World [Watts, Strogatz] — 1998

HITS [Kleinberg]&PageRank [Brin&Page] — 1997





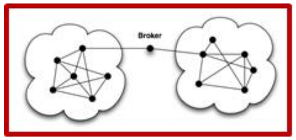


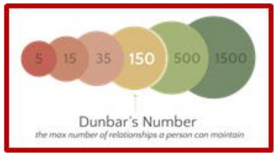
1995 — Structural Hole [Burt]

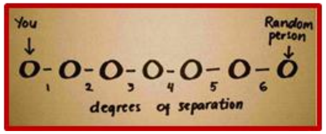
1992 — Dunbar's Number [Dunbar]

1973 — Weak Tie [Granovetter]

1967 — Six Degrees of Separation [Milgram]







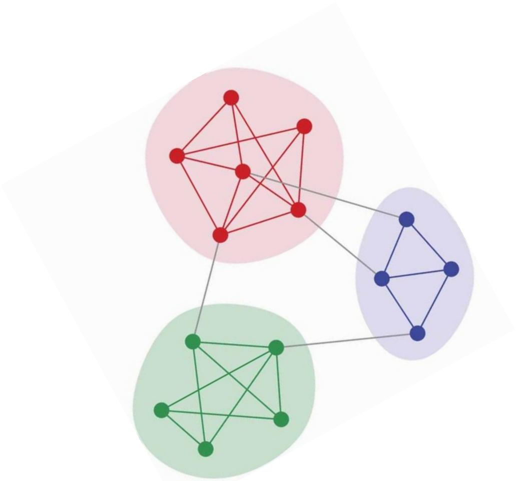
Community Detection

Community Structure

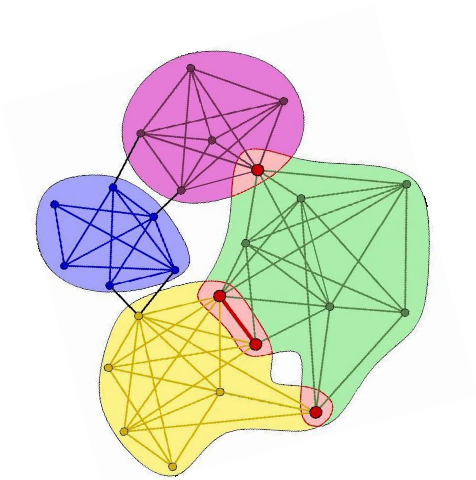
- Groups of nodes with dense internal connections and sparse external connections

Community Detection

- The procedure of finding the community structure



Disjoint



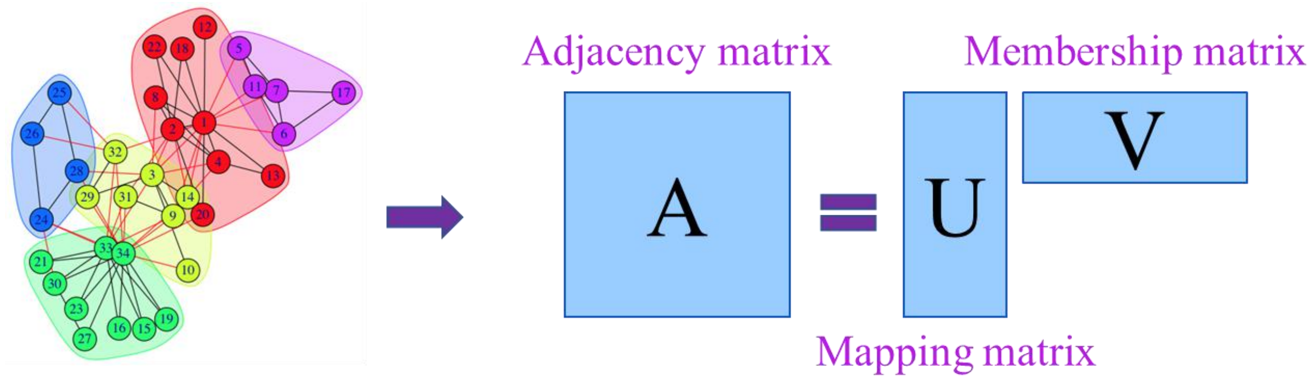
Overlapping

Paper

- **ICDM 2018**
  - Adaptive Affinity Learning for Accurate Community Detection
- **CIKM 2018**
  - Deep Autoencoder-like Nonnegative Matrix Factorization for Community Detection
- **ICDM 2019**
  - Discrete Overlapping Community Detection with Pseudo Supervision

NMF

◆ NMF for community detection



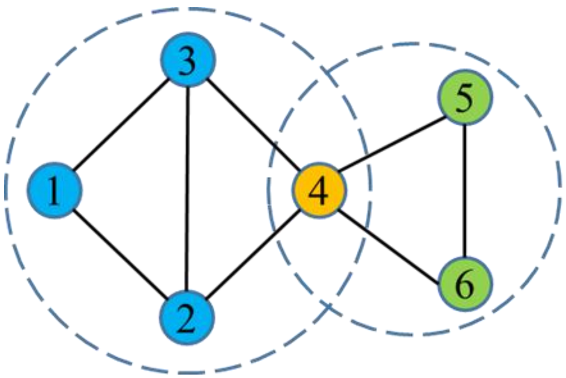
- Determine the community membership according to matrix V
- For undirected networks, set $V = U^T$

$$A \approx UV$$

$$A \approx UU^T$$

NMF

◆ An example



$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix} \qquad \mathbf{V}^T = \begin{bmatrix} 0.77 & 0.00 \\ 0.88 & 0.19 \\ 0.88 & 0.19 \\ 0.51 & 0.83 \\ 0.00 & 0.83 \\ 0.00 & 0.83 \end{bmatrix}$$

Disjoint communities

Community 1	Community 2
1, 2, 3	4, 5, 6

Overlapping communities

Community 1	Community 2
1, 2, 3, 4	4, 5, 6

* Threshold: 0.5

Sketch

- **ICDM 2018**
  - Learn similarity between nodes to improve community detection via graph regularization
- **CIKM 2018**
  - Deep AE-like architecture to capture hierarchical information
- **ICDM 2019**
  - Learn discrete overlapping community memberships to avoid post-processing

Adaptive Affinity Learning for Accurate Community Detection

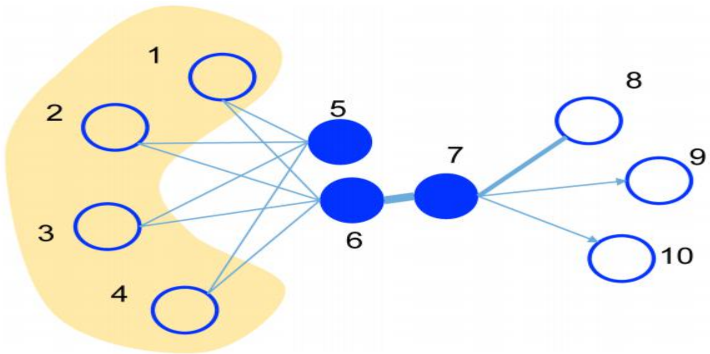
Fanghua Ye, Shenghui Li, Zhiwei Lin, Chuan Chen, Zibin Zheng

[ICDM 2018]

ICDM 2018

◆ Motivation

- Learn similarity between nodes (denoted by matrix **S**) adaptively



- Improve community detection via graph regularization

$$\|\mathbf{A} - \mathbf{U}\mathbf{U}^T\|_F^2 + \gamma \mathrm{Tr}(\mathbf{U}^T \mathbf{L_S} \mathbf{U}) \qquad \mathbf{U} \in \mathbb{R}_+^{n \times k}$$

Adaptive Affinity Learning

- Embed nodes into a low-dimensional space $\mathbb{R}^{k'}$
- Learn similarity between nodes in this low-dimensional space

Transformation matrix

$$\min_{\mathbf{S},\mathbf{Q}} \sum_{i,j=1}^{n} \|\mathbf{Q}^T\mathbf{a}_i - \boxed{\mathbf{Q}^T}\mathbf{a}_j\|_2^2 s_{ij} + \alpha\|\mathbf{S}\|_F^2,$$

$$\text{s.t.} \quad \mathbf{Q}^T\mathbf{A}\mathbf{A}^T\mathbf{Q} = \mathbf{I}_{k'}, \forall i, \sum_{j=1}^{n} s_{ij} = 1, s_{ij} \geq 0, s_{ii} = 0$$

ICDM 2018

Theorem in Graph Theory

- The multiplicity k of the eigenvalue 0 of the Laplacian matrix L_S equals the number of connected components

$$\delta_i \geq 0 \quad \Rightarrow \quad \sum_{i=1}^{k} \delta_i = 0$$

Ky Fan's Theorem

$$\sum_{i=1}^{k} \delta_i = \min_{\mathbf{Q}^T \mathbf{A} \mathbf{A}^T \mathbf{Q} = \mathbf{I}_k} \mathrm{Tr}(\mathbf{Q}^T \mathbf{A} \mathbf{L}_{\mathbf{S}} \mathbf{A}^T \mathbf{Q})$$

$$\Downarrow$$

$$k' = k$$

Deep Autoencoder-like Nonnegative Matrix Factorization for Community Detection

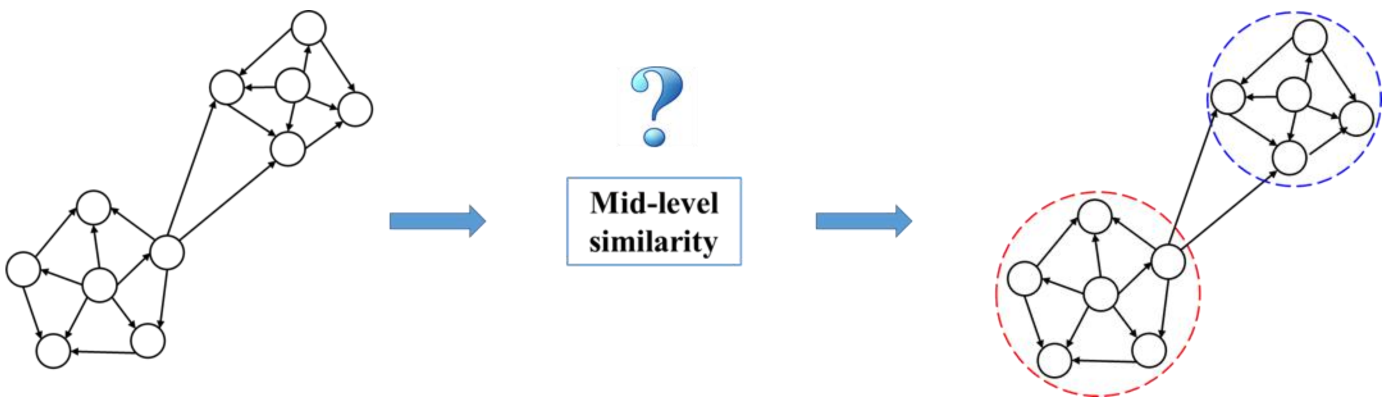
Fanghua Ye, Chuan Chen, Zibin Zheng

[CIKM 2018]

Motivation One

$$\mathbf{A} \approx \mathbf{UV}$$

- Traditional NMF is shallow
- **V** captures the community-level similarity between nodes
- **A** captures the link-level similarity between nodes
- Real-life networks consist of complicated patterns, the mapping between **A** and **V** should be hierarchic



Motivation Two

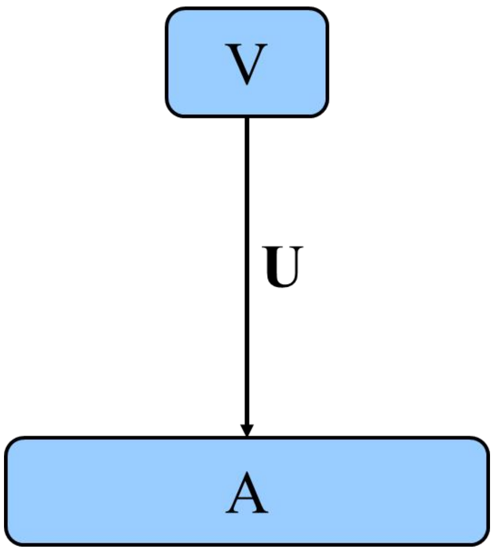
$$\mathbf{A} \approx \mathbf{UV}$$

- NMF reconstructs **A** from **V** with the aid of **U**
- An ideal **V** should be predictable from **A** with the aid of **U** as well, which is ignored by traditional NMF-based approaches

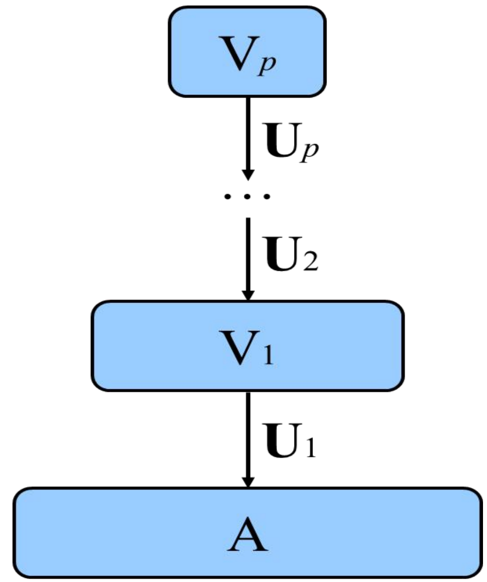
$$\mathrm{V} \stackrel{?}{=} \mathrm{U}^{T}\,\mathrm{A}$$

Deep NMF

- Further factorize the mapping **U** to learn hierarchical similarity between nodes from low level to high level



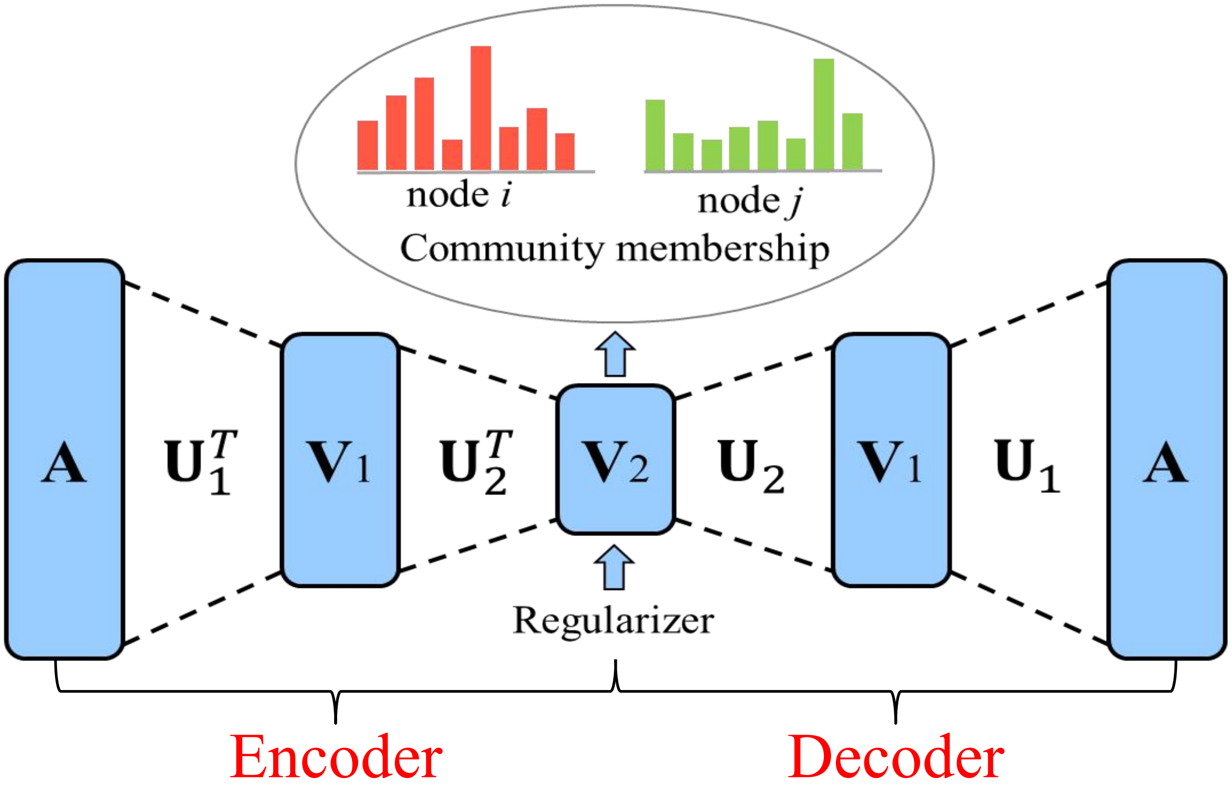
$$\mathbf{A} \approx \mathbf{U}\mathbf{V}$$



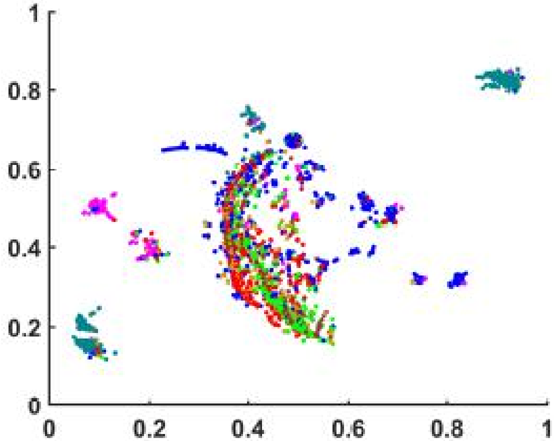
$$\mathbf{A} \approx \mathbf{U}_1\mathbf{U}_2 \cdots \mathbf{U}_p\mathbf{V}_p$$

Deep AE-like NMF

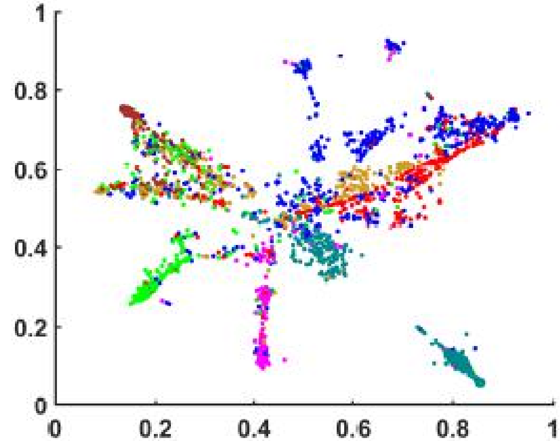
- Incorporate a symmetric encoder component into deep NMF to strengthen its representation learning ability



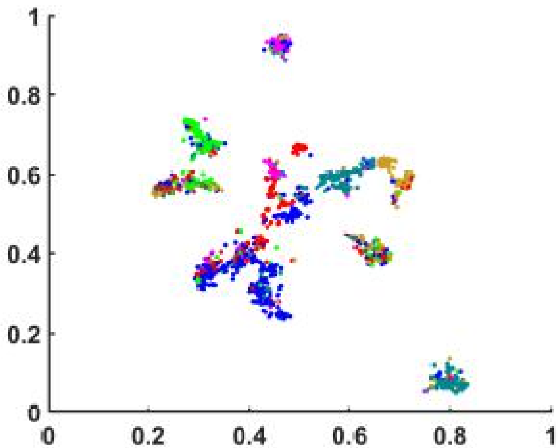
Visualization



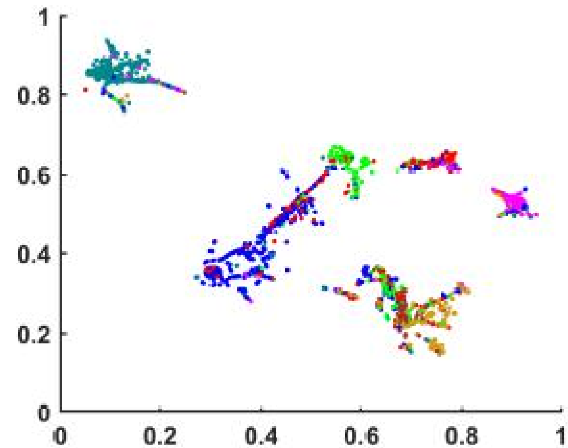
(a) Cora - before training



(b) Cora - layer 1



(c) Cora - layer 2



(d) Cora - layer 3

Discrete Overlapping Community Detection with Pseudo Supervision

Fanghua Ye, Chuan Chen, Zibin Zheng, Rong-Hua Li, Jeffrey Xu Yu

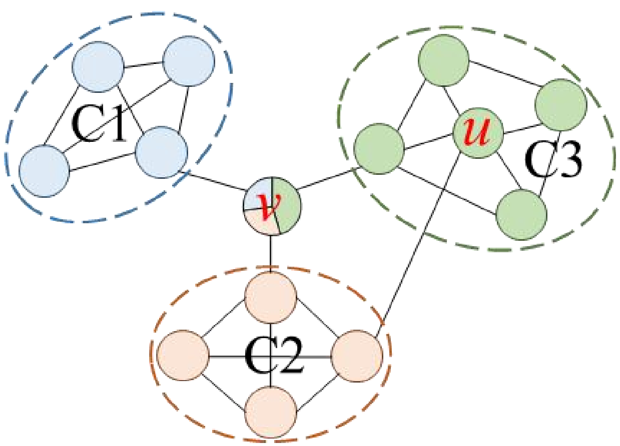
[ICDM 2019]

ICDM 2019

◆ Motivation

$$\mathbf{A} \approx \mathbf{U}\mathbf{U}^T$$

- To bypass the cumbersome post-processing step when using NMF for overlapping community detection



(a) Network structure

	C1	C2	C3
u	0	0.4	0.9
v	0.3	0.3	0.5

U

(b) Community memberships

Discrete Overlapping Community Detection

- Introduce a rotation matrix Q with orthogonal constraint

$$\min_{\mathbf{U},\mathbf{F},\mathbf{Q}} \|\mathbf{A} - \mathbf{U}\mathbf{U}^T\|_F^2 + \alpha\|\mathbf{U} - \mathbf{F}\mathbf{Q}\|_F^2,$$

$$\text{s.t. } \mathbf{U} \geq \mathbf{0} \wedge \boxed{\mathbf{Q}^T\mathbf{Q} = \mathbf{I}_k} \wedge \mathbf{F} \in \{0,1\}^{n\times k}$$

$$\Downarrow$$

$$\mathbf{F}\mathbf{F}^T = \mathbf{F}\mathbf{Q}\mathbf{Q}^T\mathbf{F}^T = (\mathbf{F}\mathbf{Q})(\mathbf{F}\mathbf{Q})^T \approx \mathbf{U}\mathbf{U}^T$$

$$\Downarrow$$

$$\mathbf{F}\mathbf{F}^T \approx \mathbf{U}\mathbf{U}^T \approx \mathbf{A}$$

- Why not $||A - FF^T||_F^2$?

Discriminative Pseudo Supervision

- Treat **F** as the pseudo ground-truth to learn a discriminative prediction function
- To exploit discriminative information in an unsupervised manner

$$\mathcal{L}(\mathbf{W}, \mathrm{b}; \mathbf{F}, \mathbf{A}) = \|\mathbf{F} - \overbrace{\phi^T(\mathbf{A})\mathbf{W} - \mathbf{1}_n \mathbf{b}^T\|_F^2 + \gamma}^{\boxed{\text{Kernel regression}}}\|\mathbf{W}\|_F^2$$

$$\mathcal{L}(\mathbf{B}, \mathrm{b}; \mathbf{F}, \hat{\mathbf{K}}) = tr(\mathbf{F}^T \mathbf{S} \mathbf{F})$$

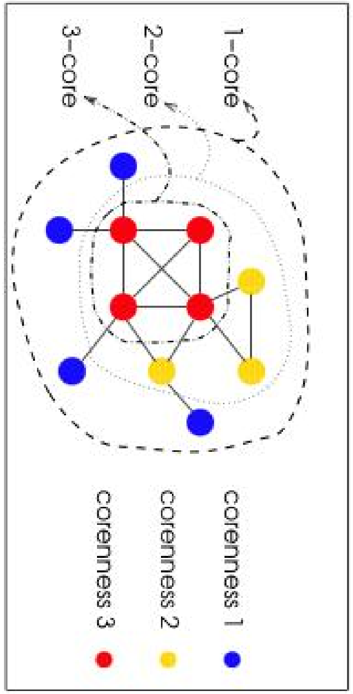
SIGMOD 2018

Database Flavor

Definition 1. *Let* $H = (V_H, E_H)$ *and* $H' = (V_{H'}, E_{H'})$ *be two communities. If* $f_i(H) \leq f_i(H')$ *for all* $i = 1, \cdots, d$, *and there exists* $f_i(H) < f_i(H')$ *for a certain* i, *we call* H' *dominates* H, *denoted by* $H \prec H'$.

Definition 2. *Given a multi-valued graph* $G = (V, E, X)$ *and an integer* k. *A skyline community with a parameter* k *is an induced subgraph* $H = (V_H, E_H)$ *of* G *such that it satisfies the following properties.*

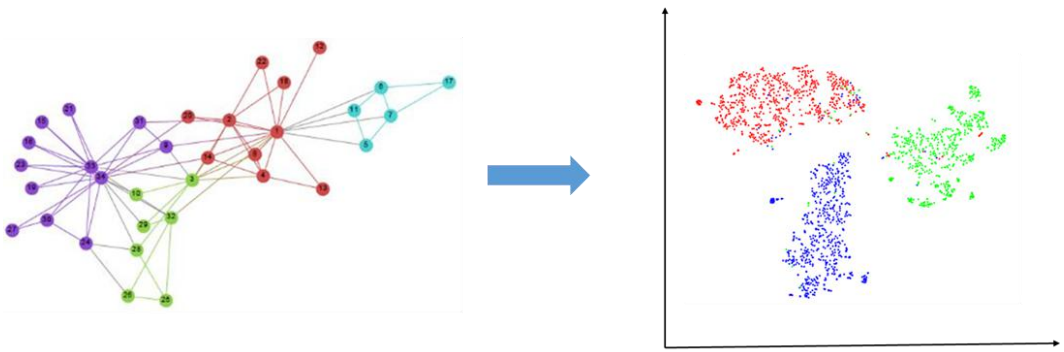
- *Cohesive property:* H *is a connected* k*-core;*
- *Skyline property: there does not exist an induced subgraph* H' *such that* $H \prec H'$;
- *Maximal property: there does not exist an induced supergraph* H' *such that (1)* H' *is a connected* k*-core, (2)* H' *contains* H, *and (3)* $f_i(H') = f_i(H)$ *for all* $i = 1, \cdots, d$.

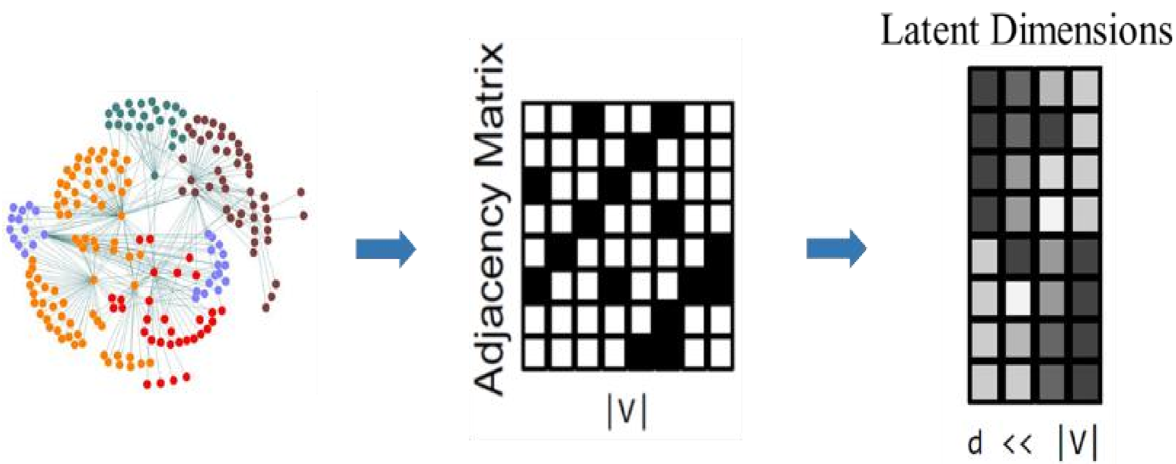


On-going: Network Embedding

Network Embedding

- **Dimension Reduction or Representation Learning**





Network Embedding

Existing Methods

Isomap	2000	Science	A Global Geometric Framework for Nonlinear Dimensionality Reduction
LLE	2000	Science	Nonlinear dimensionality reduction by locally linear embedding
LE	2001	NIPS	Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering
EdgeCluster	2009	CIKM	Scalable Learning of Collective Behavior Based on Sparse Social Dimensions
DeepWalk	2014	KDD	DeepWalk: Online Learning of Social Representations
GraRep	2015	CIKM	GraRep: Learning Graph Representations with Global Structural Information
HNE	2015	KDD	Heterogeneous Network Embedding via Deep Architectures
LINE	2015	WWW	LINE: Large-scale Information Network Embedding
DDRW	2016	ACL	Discriminative Deep RandomWalk for Network Classification
MMDW	2016	IJCAI	Max-Margin DeepWalk: Discriminative Learning of Network Representation
DNGR	2016	AAAI	Deep Neural Networks for Learning Graph Representations
HOPE	2016	KDD	Asymmetric Transitivity Preserving Graph Embedding
SDNE	2016	KDD	Structural Deep Network Embedding
node2vec	2016	KDD	node2vec: Scalable Feature Learning for Networks

Network Embedding

- **Existing Methods--Category**

Random Walk	Efficiency
Matrix Factorization	Interpretability
Deep Learning	Nonlinearity

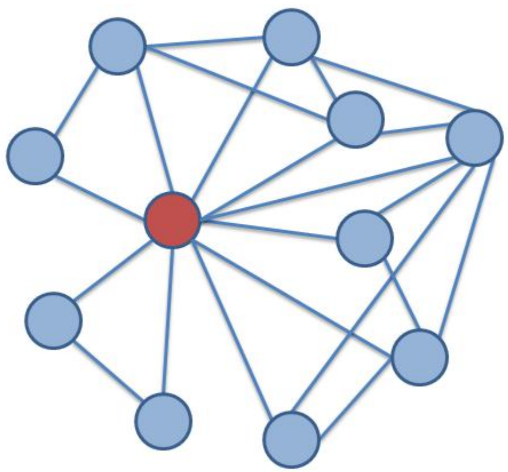
Degree Preserving Network Embedding

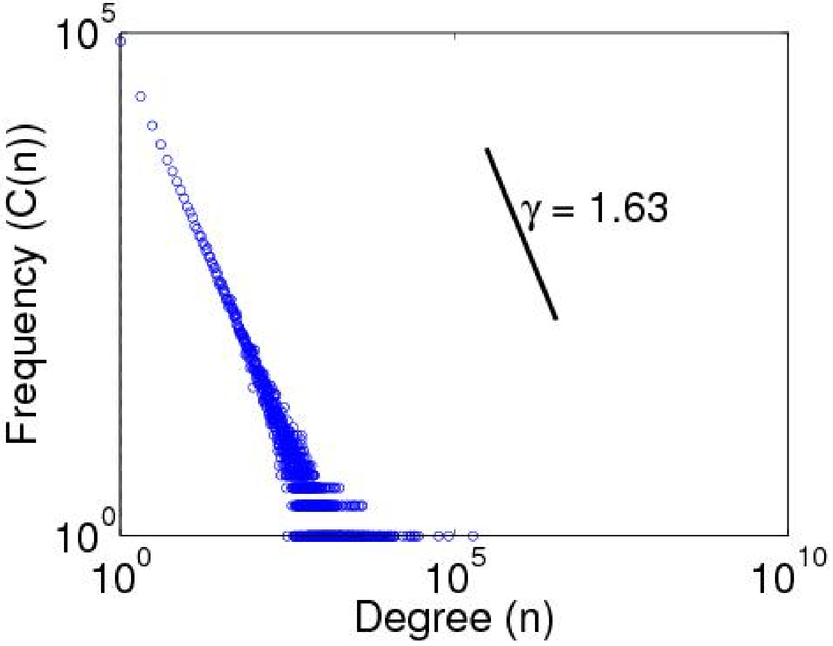
[On-going]

Motivation

◆ Local and Global

- ➢ Degree is a local information
- ➢ Degree (distribution) is a global information





Method

- **Degree Preserving**

$$A1_n = d.$$

$$\min_{U} \|A - UU^T\|_F^2, s.t.\ U \geq 0, UU^T 1_n = d$$

$$d = (d_1, d_2, \cdots, d_n)^T$$

High-order similarity matrix

- **Network Embedding**

$$U = SP$$

Inductive

$$\min_{P,U} \|U - SP\|_F^2 + \alpha\|A - UU^T\|_F^2$$

$$s.t.\ P \geq 0, U \geq 0, UU^T 1_n = d.$$

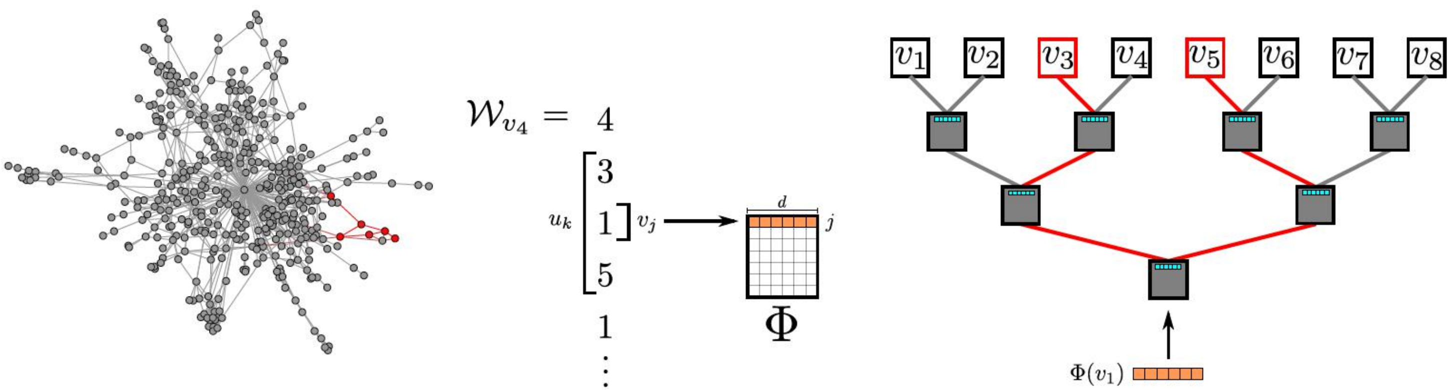
Multiwalk: Multiple Random Walkers induced Network Embedding

[On-going]

Skip-Gram

Generate node sequences

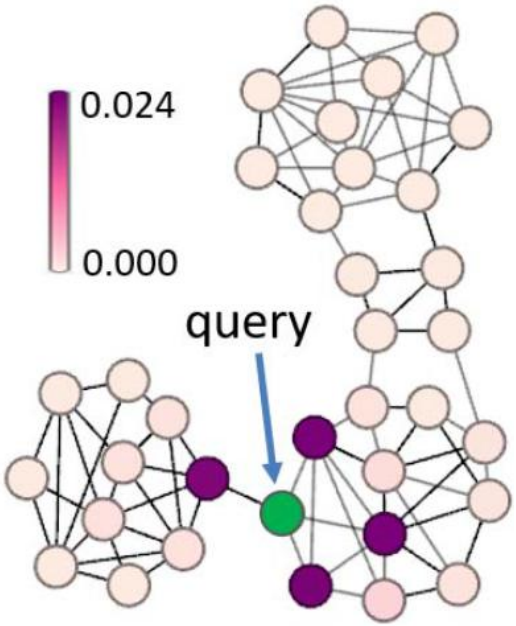
- Each node corresponds to a word
- Each node sequence corresponds to a sentence



(a) Random walk generation. (b) Representation mapping. (c) Hierarchical Softmax.

Motivation

- **Community structure is the most important**
  - Single walker is sensitive to boundary nodes
  - To be efficient to deal with large-scale networks



Method

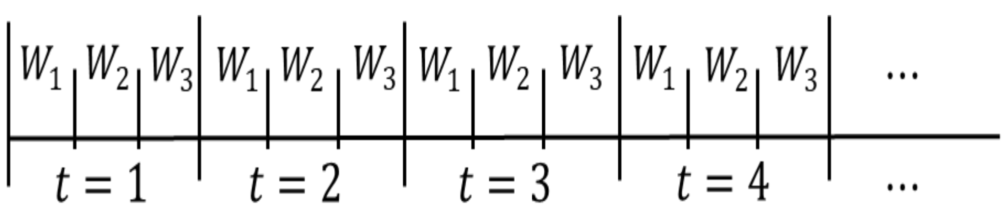
◆ Single Walker to Multiple Walkers

- Random walk with restart

$$x^{(t+1)} = cP^T x^{(t)} + (1-c)v$$

$$\Downarrow$$

$$x_i^{(t+1)} = cP^T x_i^{(t)} + (1-c)v_i^{(t)} \qquad v_i^{(t)} = \frac{1}{k}\sum_{j=1}^{k} x_j^{(t)}$$

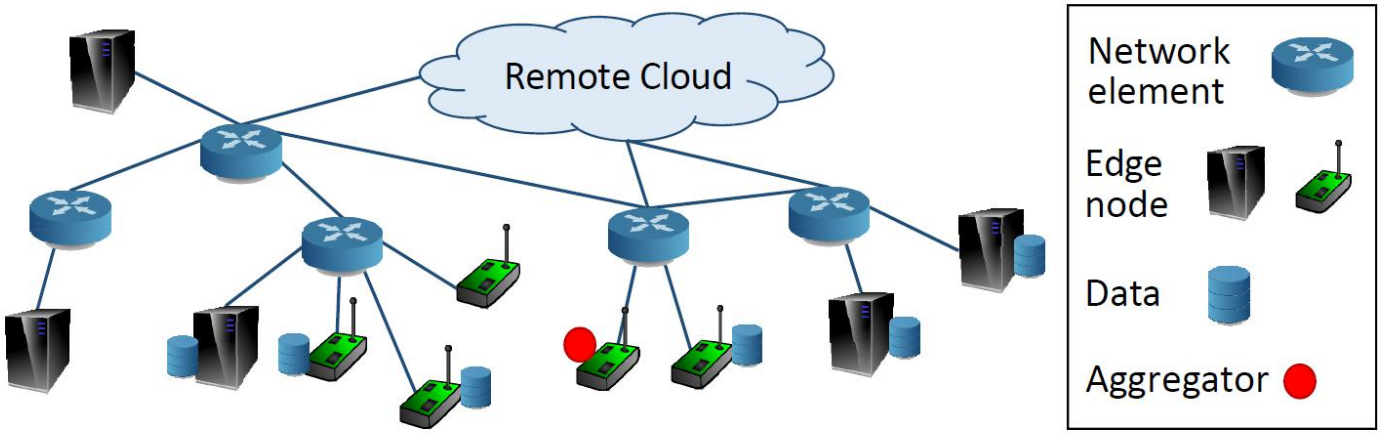


Future: Federated Learning

Federated Learning

◆ Concept

➢ The main idea is to build machine learning models based on data sets that are distributed across multiple devices while preventing data leakage



THANKS