Data Mining: Advanced Techniques

Introduction

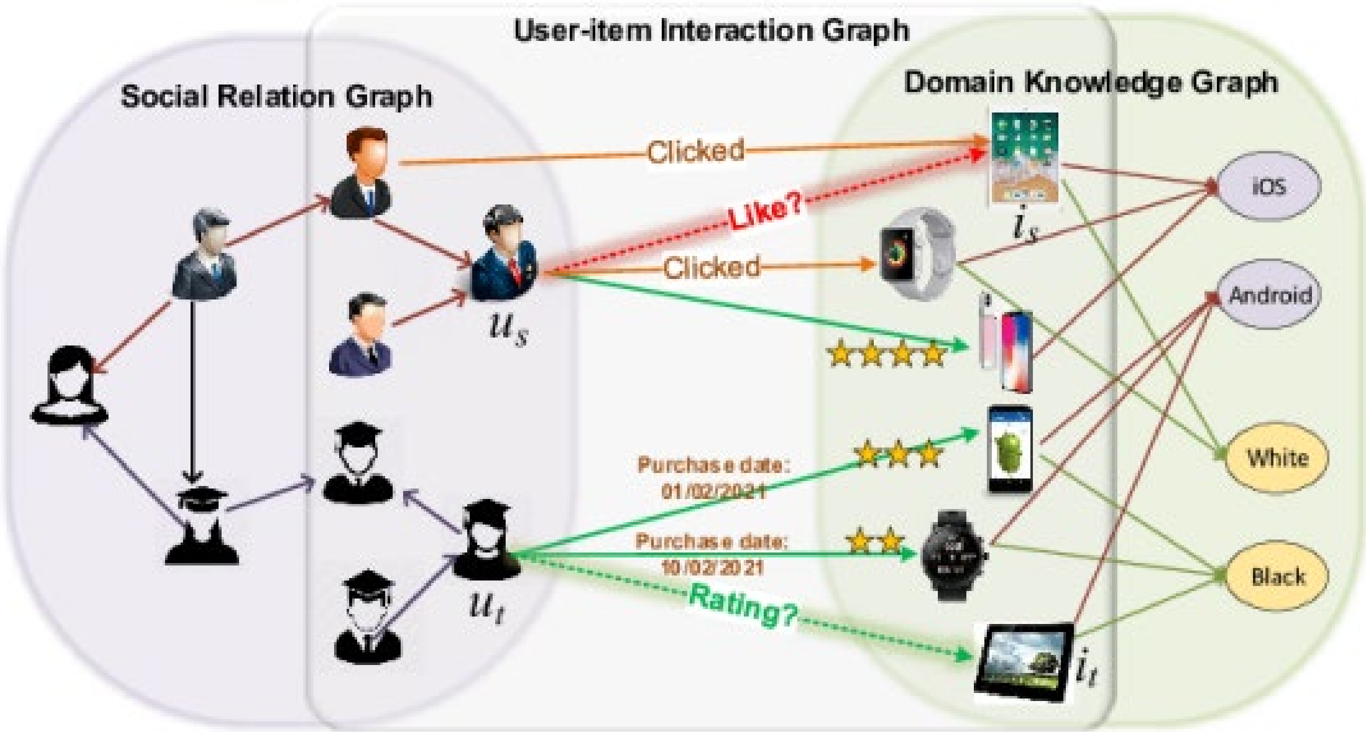
- 主讲教师: 陈佳伟, sleepyhunt@zju.edu.cn
  - https://jiawei-chen.github.io/
- TA: 陈思睿, chenthree@zju.edu.cn

Why graph mining?

- A graph is a mathematical structure used to model pairwise relations between objects.
- Graph is ubiquitous in our world, e.g., social networks, molecular graph, transactions, Internet of Things (IoT), etc.
- Mining on graph can utilize both features and relations of individuals
- Graph+ML has been widely applied in many fields

Why graph mining?

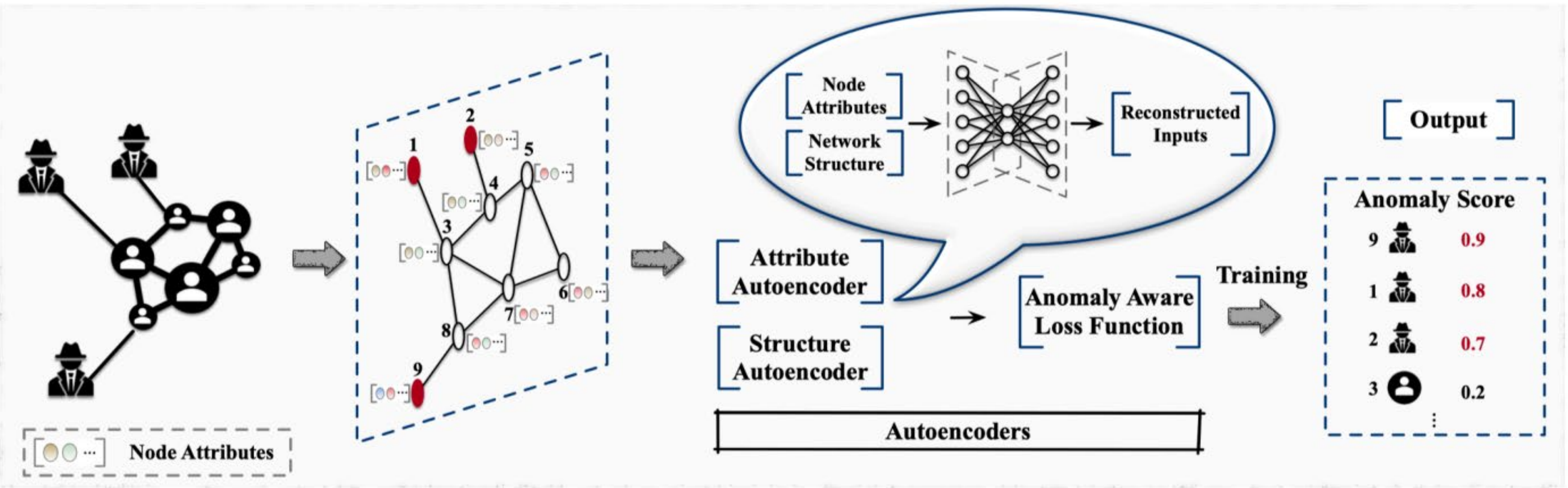
- Recommender system



Slide from Sheng Zhou

Why graph mining?

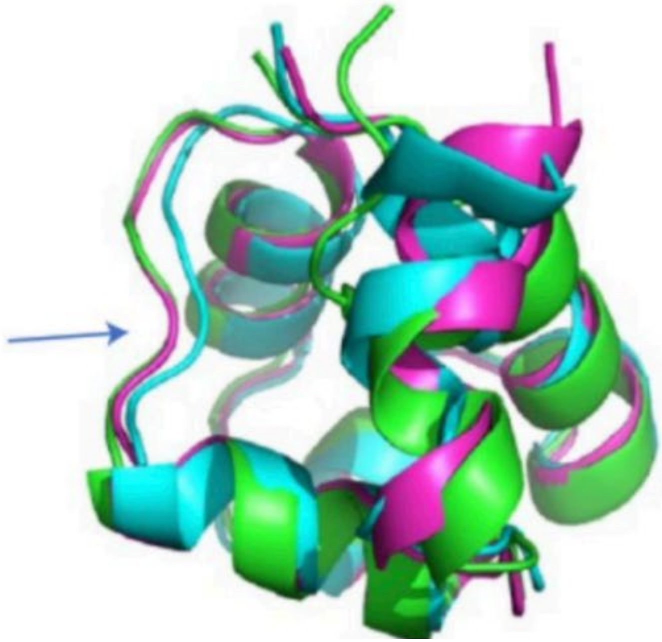
- Anomaly detection

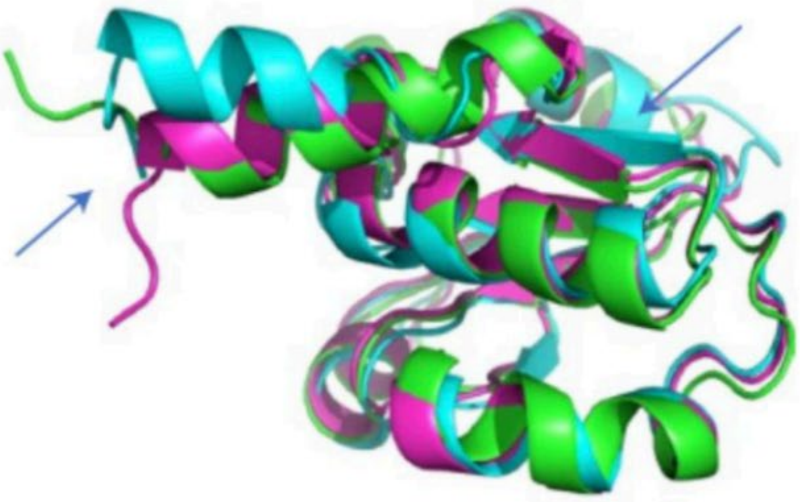


Slide from Sheng Zhou

Why graph mining?

- Protein structure prediction

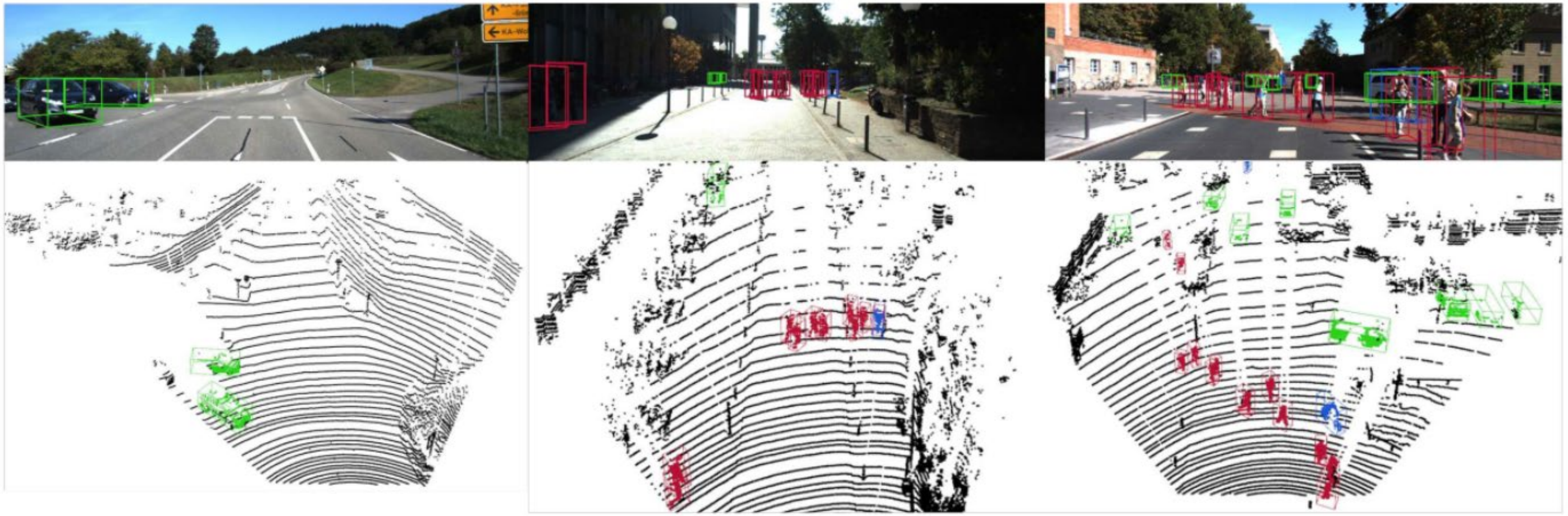




Slide from Sheng Zhou

Why graph mining?

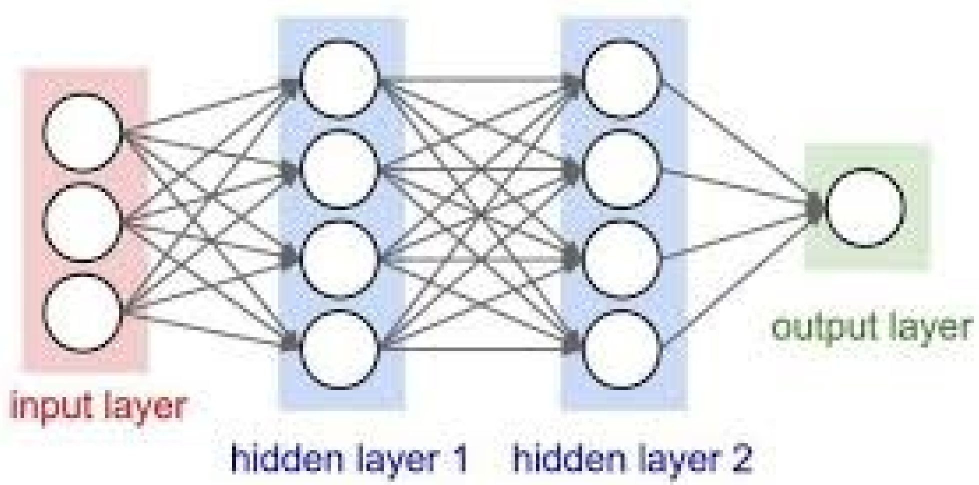
- Computer vision



Slide from Sheng Zhou

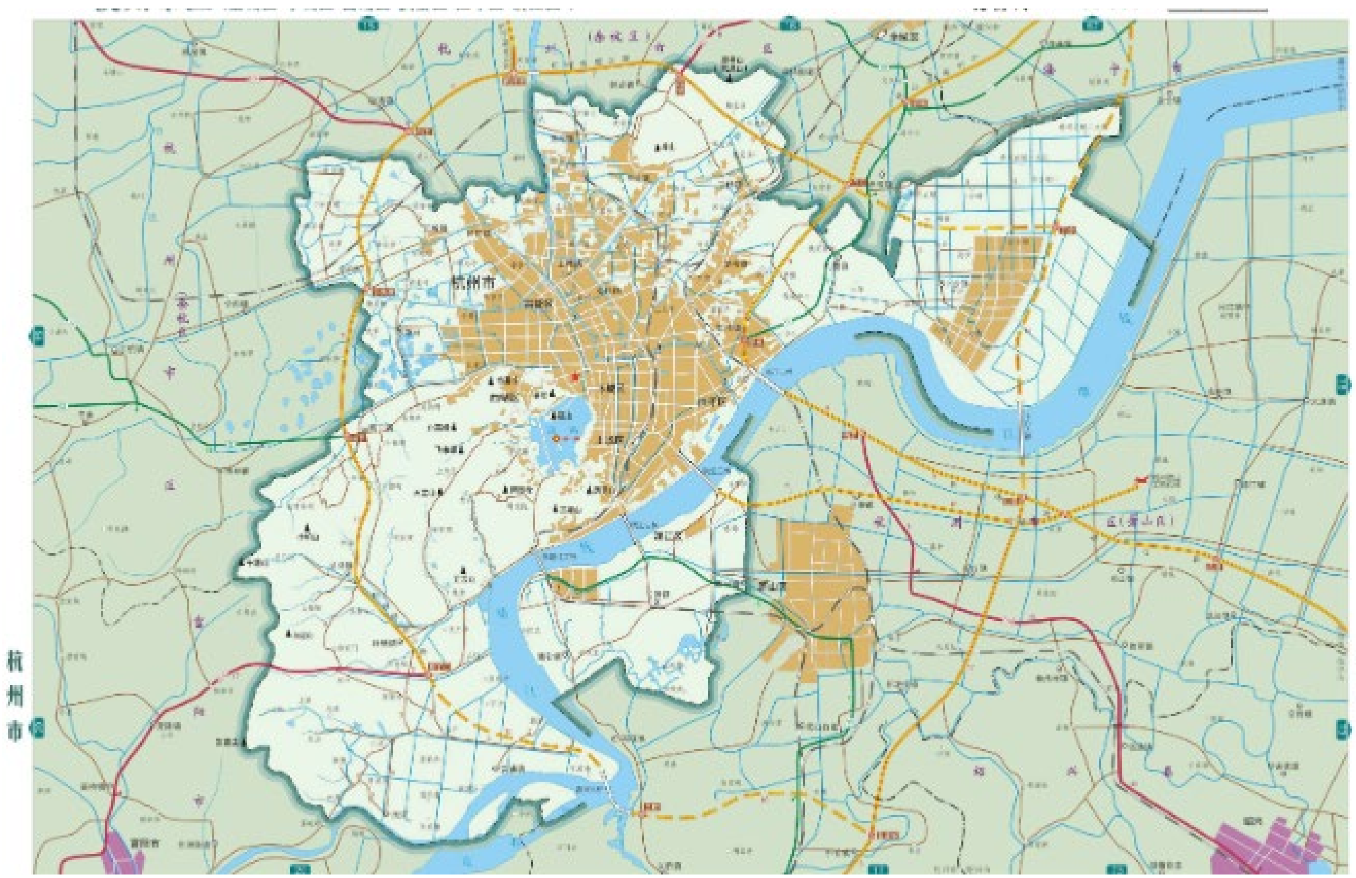
Why graph mining?

- Neural network



Why graph mining?

- Traffic



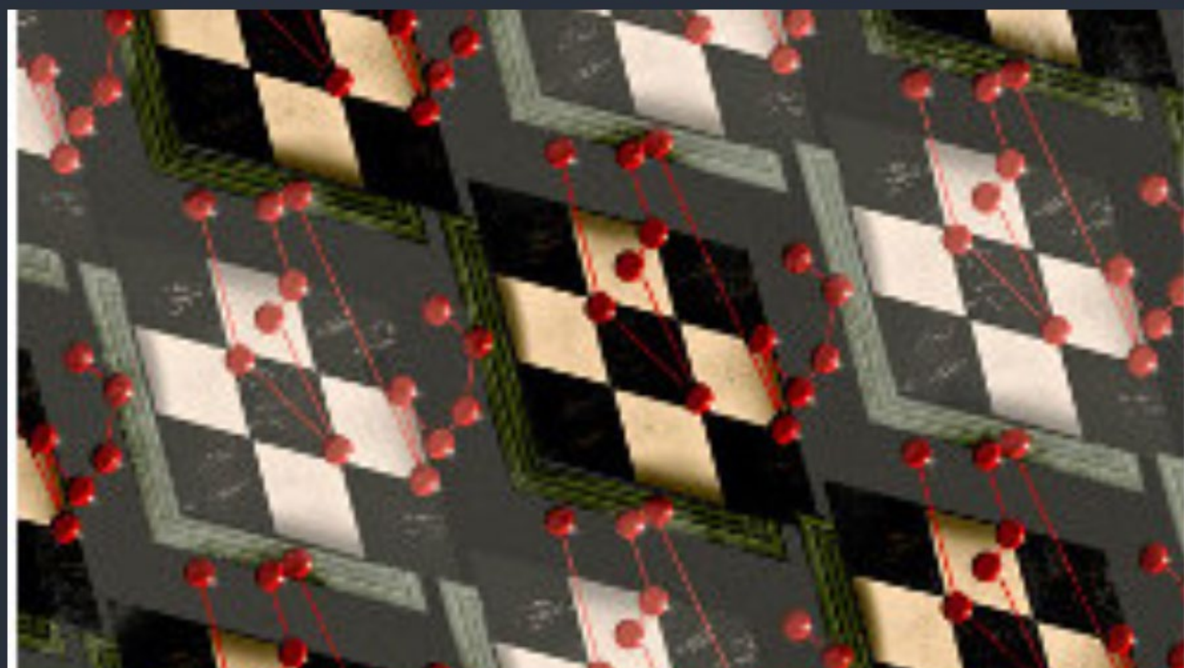
Why graph mining?

The Future Is Big Graphs: A Community View on Graph Processing Systems

Ensuring the success of big graph processing for the next decade and beyond.

By Sherif Sakr, Angela Bonifati, Hannes Voigt, Alexandru Iosup, Khaled Ammar, Renzo Angles, Walid Aref, Marcelo Arenas, Maciej Besta, Peter A. Boncz, Khuzaima Daudjee, Emanuele Della Valle, Stefania Dumbrava, Olaf Hartig, Bernhard Haslhofer, Tim Hegeman, Jan Hidders, Katja Hose, Adriana Iamnitchi, Vasiliki Kalavri, Hugo Kapp, Wim Martens, M. Tamer Özsu, Eric Peukert, Stefan Plantikow, Mohamed Ragab, Matei R. Ripeanu, Semih Salihoglu, Christian Schulz, Petra Selmer, Juan F. Sequeda, and Joshua Shinavier

Posted Sep 1 2021



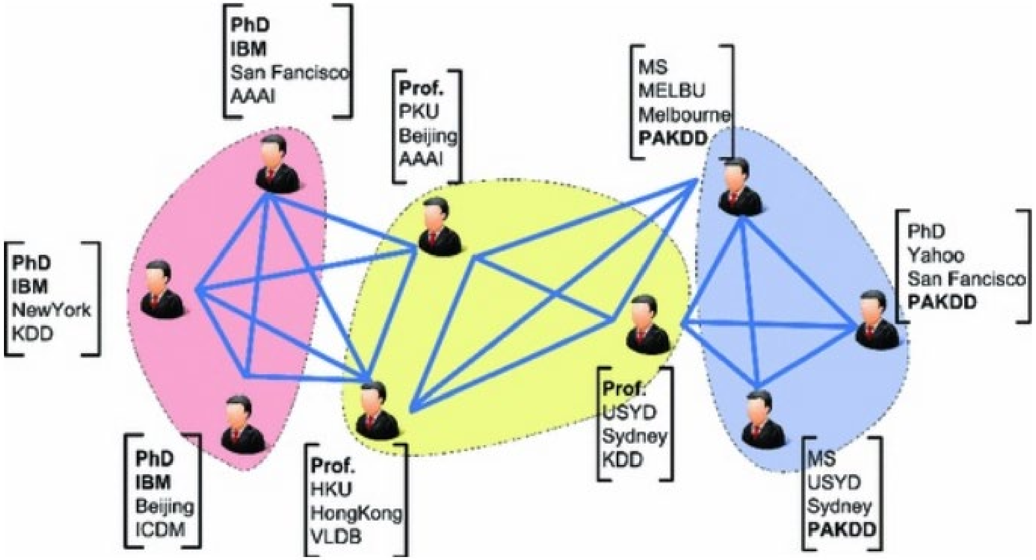
Cover Article of 《Communications of the ACM》

Graph Mining

- Graph data
- Classic graph representation learning
- Graph neural network
- Promising directions

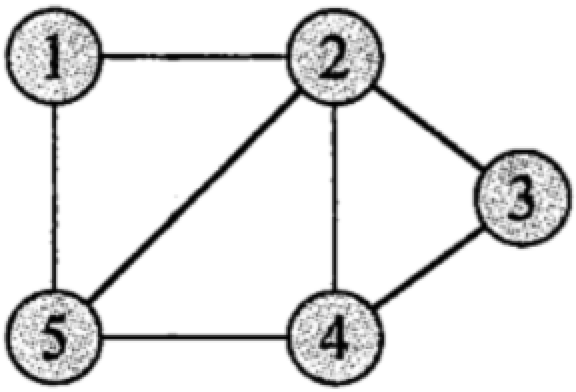
What is graph?

- A graph is a mathematical structure used to model pairwise relations between objects.
- Graph can be formally defined as: $G = \{V, E, X\}$
  - V denotes the set of nodes
  - E denotes the set of edges between nodes
  - X denotes the set of features of nodes



What is graph?

- Edges can be also be represented as a adjacency matrix A



$$\begin{array}{c|ccccc|}
 & 1 & 2 & 3 & 4 & 5 \\
\hline
1 & 0 & 1 & 0 & 0 & 1 \\
2 & 1 & 0 & 1 & 1 & 1 \\
3 & 0 & 1 & 0 & 1 & 0 \\
4 & 0 & 1 & 1 & 0 & 1 \\
5 & 1 & 1 & 0 & 1 & 0 \\
\hline
\end{array}$$

- The degree of a node is represented as:
  - $d_i = \sum_{j=1}^{n} A_{ij}$
  - the number of edges connected to the node

What is graph?

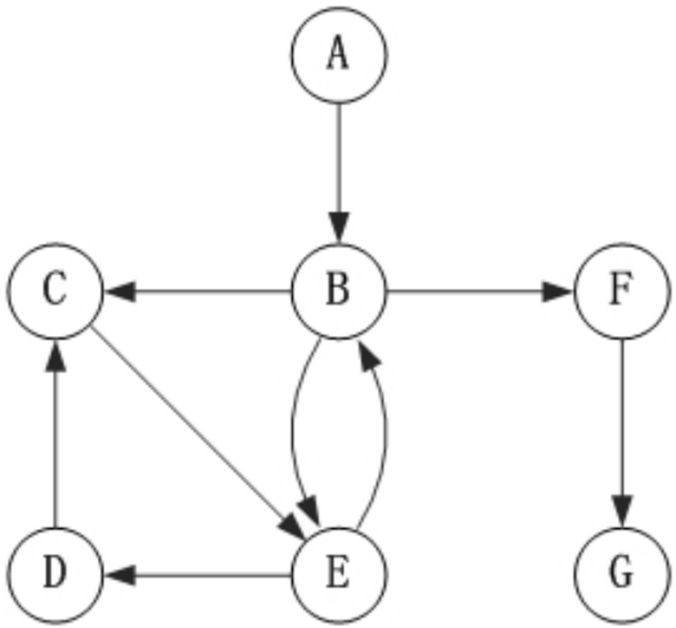
- Complex graph
  - Directed Graph
  - Weighted Graph
  - Heterogeneous Graph
  - Multiplex Graph
  - Dynamic Graph
  - …….

Slide from Sheng Zhou

What is graph?

- Directed Graph
  - Asymmetric relations
  - A-> B does not suggest B-> A

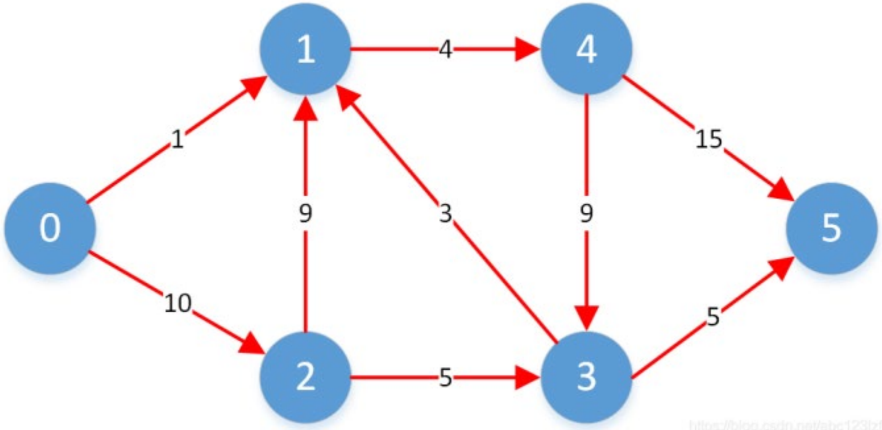
- Applications
  - Transaction
  - Follow in social network
  - Citation



Slide from Sheng Zhou

What is graph?

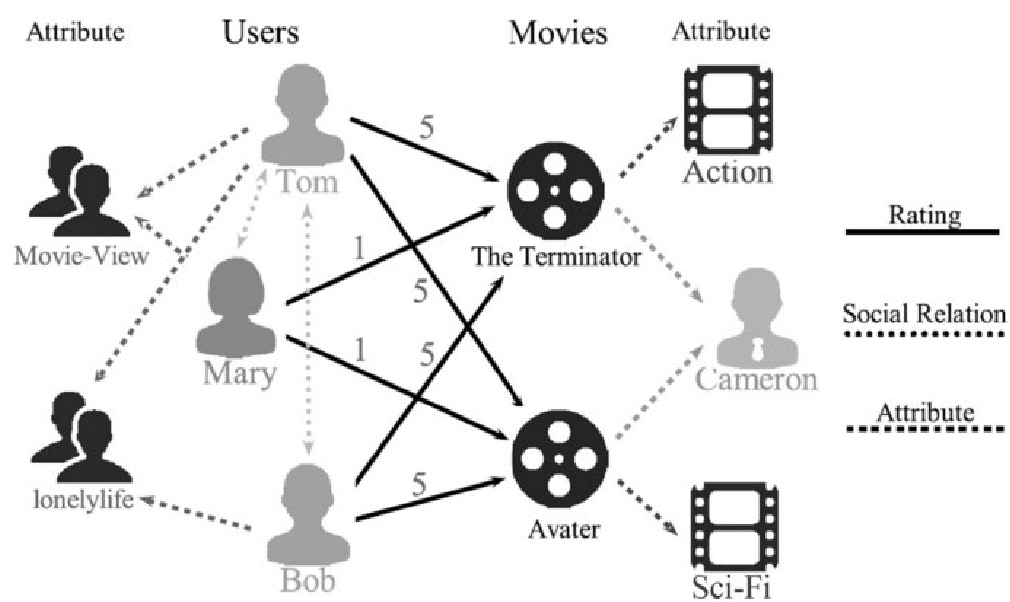
- Weighted Graph
  - Each edge is companied with a weight
  - The relationships are not equivalent
- Applications
  - Friendship
  - Transaction
  - Traffic flow



Slide from Sheng Zhou

What is graph?

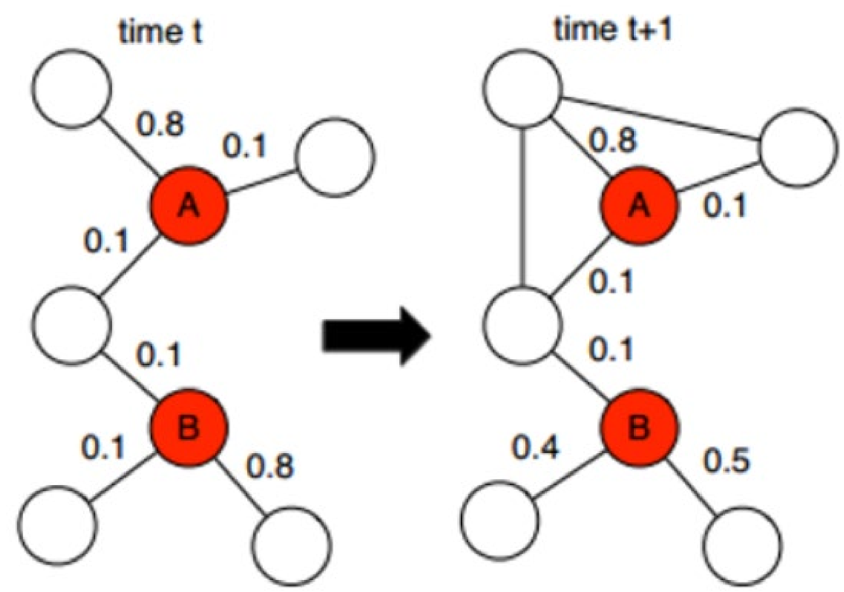
- Heterogeneous Graph
  - Multiply types of nodes/edges
  - Diverse features of nodes/edges
- Applications
  - Knowledge graph
  - Social network
  - Citation network



Slide from Sheng Zhou

What is graph?

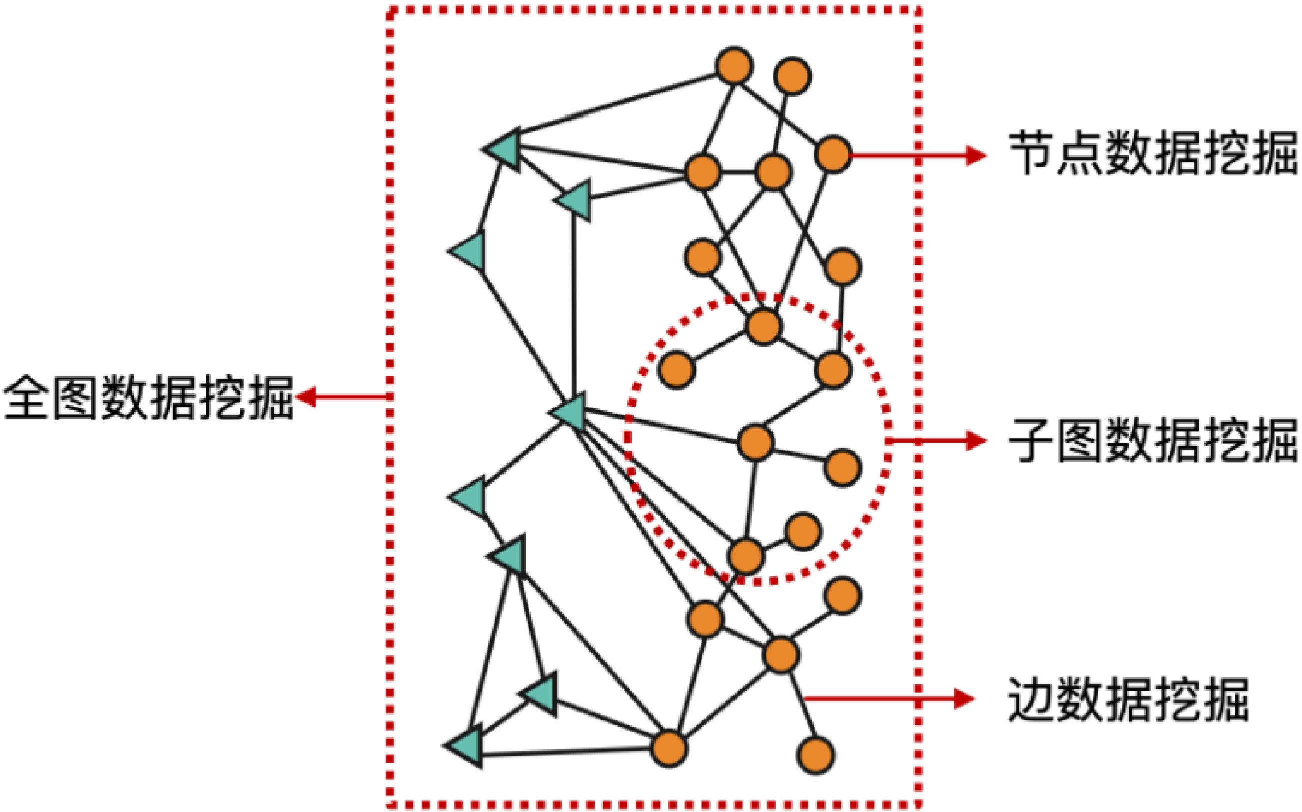
- Dynamic (Temporal) Graph
  - Graph structure or features evolves with time
- Applications
  - Social network
  - Citation network
  - Transportation network



Slide from Sheng Zhou

Graph mining

- Task of graph mining



Slide from Sheng Zhou

Graph mining

- Node classification
  - Given a graph $G = \{V, E, X\}$ and labels of partial nodes, predict the labels of the rest.
  - Different from traditional ML: i.i.d assumption does not hold
  - KEY: utilize the relations
- Applications
  - User profile
  - Products classification

Slide from Sheng Zhou

Graph mining

- Link prediction
  - Given a graph $G = \{V, E, X\}$, predict missing edges
- Applications
  - Item recommendation
  - Complete the knowledge graph
  - Friend recommendation

Slide from Sheng Zhou

Graph mining

- Important motif mining
  - a motif is a small, recurring, and statistically significant subgraph within a graph
  - the process of identifying important motif for understanding the properties and functions of graphs
- Applications
  - Graph interpretability
  - molecule

Graph mining

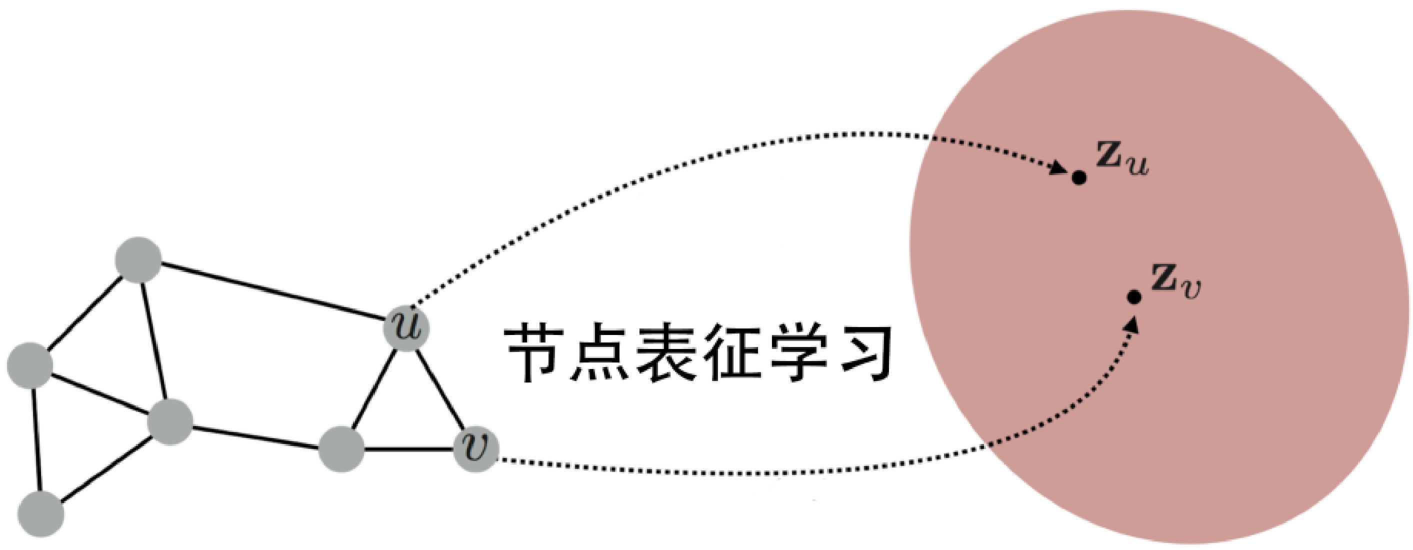
- Graph classification
  - assigning a label to an entire graph based on its structure, properties, or attributes.
- Applications
  - Drug discovery
  - Fraud detection
  - Protein function prediction

Graph Mining

- Graph data
- Classic graph representation learning
- Graph neural network
- Promising directions

Graph Representation Learning

- Representation learning
  - map each node/graph into continuous low-dimensional vectors space
  - preserve structural and relational information of the graph



Graph Representation Learning

- Motivation
  - Original graph is hard to be utilized
    - Sparsity of the adjacent matrix
    - Diverse information
    - Relations between nodes
  - Benefits of embedding
    - Uniform representation
    - Low-dimension benefit memorization
    - Compatible to ML

Graph Representation Learning

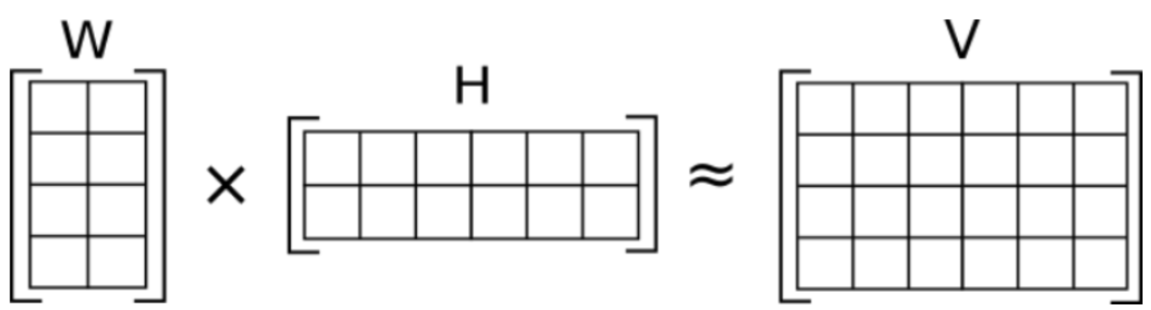
- Classic graph representation learning
  - Matrix factorization
  - Random walk

Graph Representation Learning

- Matrix factorization
  - factorizing a matrix into a product of two lower-dimensional matrices
  - The goal is to approximate the original matrix by capturing its underlying structure and patterns

W, H: embedding of nodes

V: graph features



Slide from Sheng Zhou

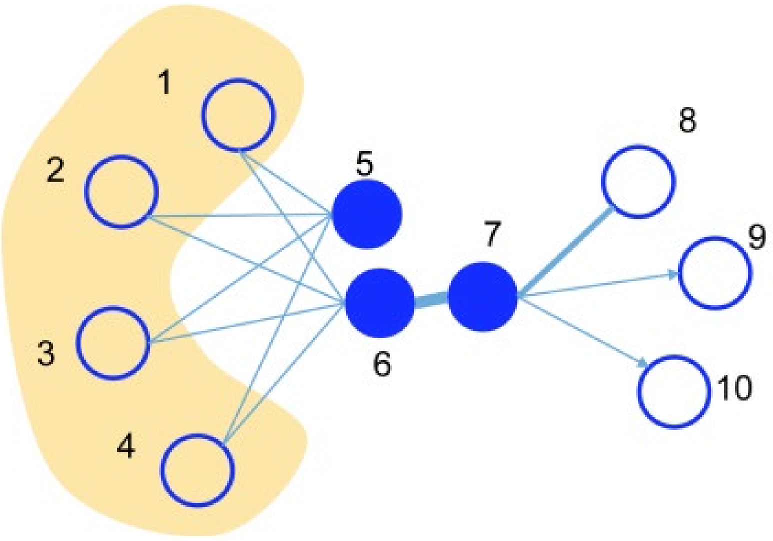
Matrix factorization

- Keys: how to construct the matrix V
  - Neighbor relations
  - High-order relations
  - Page-rank
  - ……
- The inner product of node embeddings capture this information

Slide from Sheng Zhou

Matrix factorization

- Neighbor relations
  - First-order Proximity: whether the two nodes are connected
  - Second-order Proximity: Proximity of the neighbors of two nodes

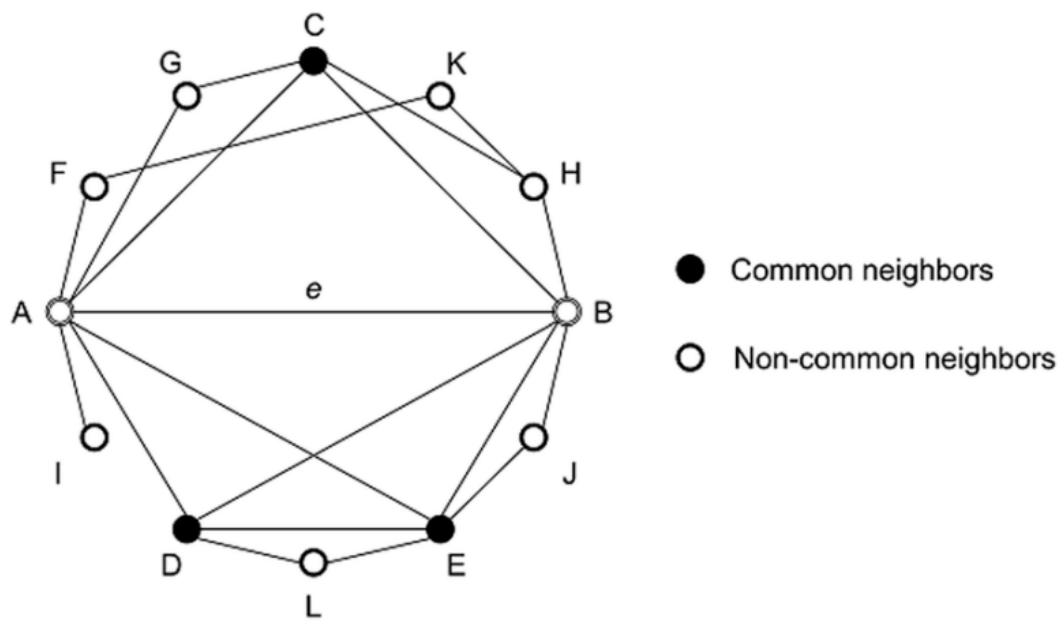


WWW'15: LINE: Large-scale Information Network Embedding.

Matrix factorization

- High-order neighbor relations
  - Common Neighbors: the number of common neighbors between two nodes

$$S^{CN} = A^2$$



Matrix factorization

- High-order neighbor relations
  - Adamic-Adar: integrate degrees into common neighbors

$$A(x,y) = \sum_{u \in N(x) \cap N(y)} \frac{1}{\log |N(u)|}$$

  - Matrix form:

$$\mathbf{S}^{AA} = \mathbf{A} \cdot \mathbf{D} \cdot \mathbf{A}$$
$$\mathbf{D}_{ii} = 1 / \sum_j (A_{ij} + A_{ji})$$

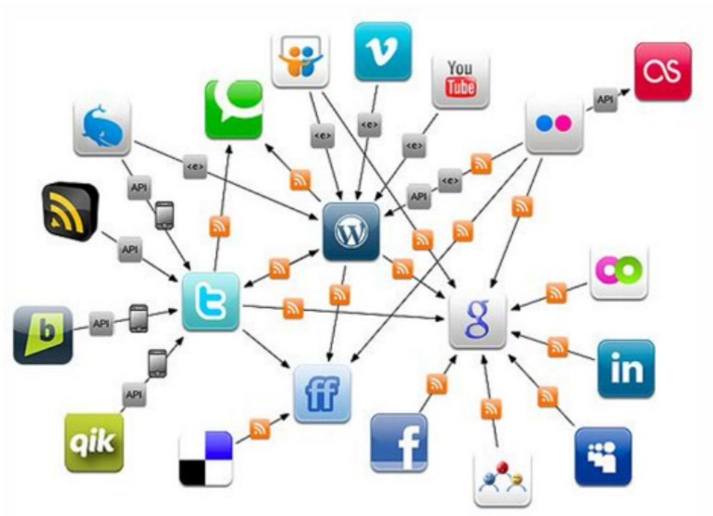
Matrix factorization

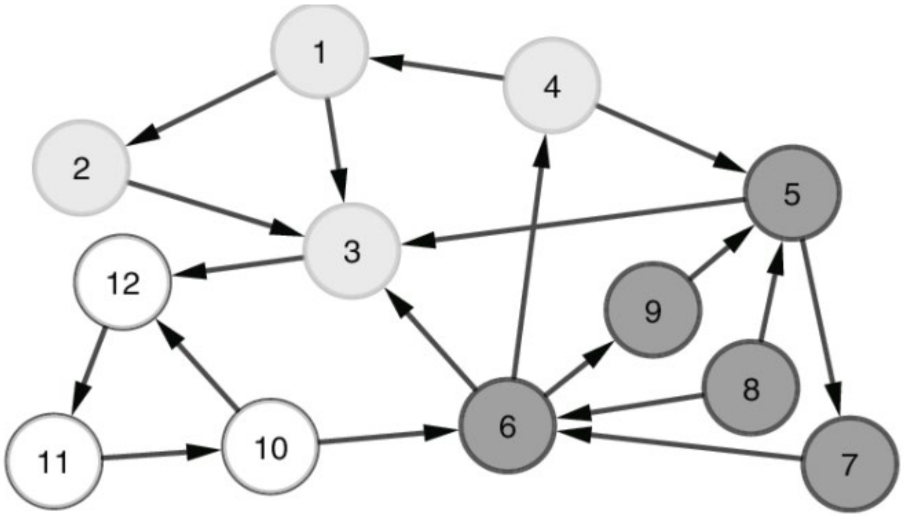
- High-order neighbor relations
  - Katz distance takes into account the total number of paths between two nodes in a network
  - The importance of each path is weighted according to the length of the path

$$S^{Katz} = \sum_{l=1}^{\infty} \beta^l A^l$$

Matrix factorization

- Page-rank
  - How to evaluate the importance and relevance of pages





Matrix factorization

- Basic assumptions of Page-rank:
  - In a webpage, each hyperlink represents a vote from the source page to the target page.
  - An important page often receives more votes.
  - An important page receives vote from other important pages.

Matrix factorization

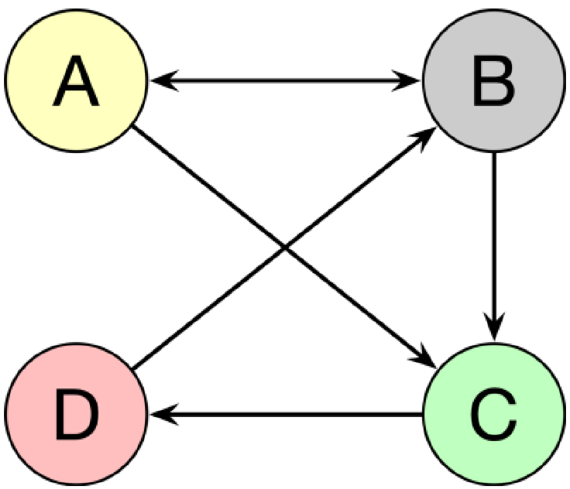
- Page-rank:
  - Each edge in the directed graph represents the flow of node importance.
  - Let the important of a node i be r_i, the out-degree of the node is d_i, the flow of each edge (i, j) is $M_{ij} = \frac{r_i}{d_i}$.
  - The final importance of the nodes is determined by the steady state

$$r = \mathrm{W} \cdot r \qquad W_{ij} = \frac{A_{ji}}{d_j}$$

  - Utilize r or W for constructing matrix V

Matrix factorization

- Page-rank --- an example:



$$r_A = r_B/2$$
$$r_B = r_A/2 + r_D$$
$$r_C = r_A/2 + r_B/2$$
$$r_D = r_C$$

	A	B	C	D
A	0	$r_A/2$	$r_A/2$	0
B	$r_B/2$	0	$r_B/2$	0
C	0	0	0	r_C
D	0	r_D	0	0

Slide from Sheng Zhou

Matrix factorization

- Page-rank --- another perspective:
  - Considering a “Web crawler” with random walker
    - Locate at i-th node
    - Randomly choose a hyperlink for walking
    - Walk from i-th node to another j-th node
    - Continuously repeat the above processes
  - The probability of the node that the crawler finally locates at:

$$p(t+1) = \mathrm{W} \cdot p(t)$$

Slide from Sheng Zhou

Matrix factorization

- Rooted Page-rank:
  - Weaknesses of page-rank:
    - Some nodes may have no out-edges
    - There exists non-connected components in a graph
  - Rooted Page-rank adds a small probability to randomly walk to an arbitrary node

$$r_j = \sum_{i \to j} \beta \frac{r_i}{d_i} + (1 - \beta) \frac{1}{N}$$

Slide from Sheng Zhou

Graph Representation Learning

- Classic graph representation learning
  - Matrix factorization
  - Random walk

Slide from Sheng Zhou

Random Walk

- Node representation via random walk
    - Performing random walk on graph to generate multiple sequences of nodes
    - utilizing node2vec technique to get the node representation

Slide from Sheng Zhou

Random Walk

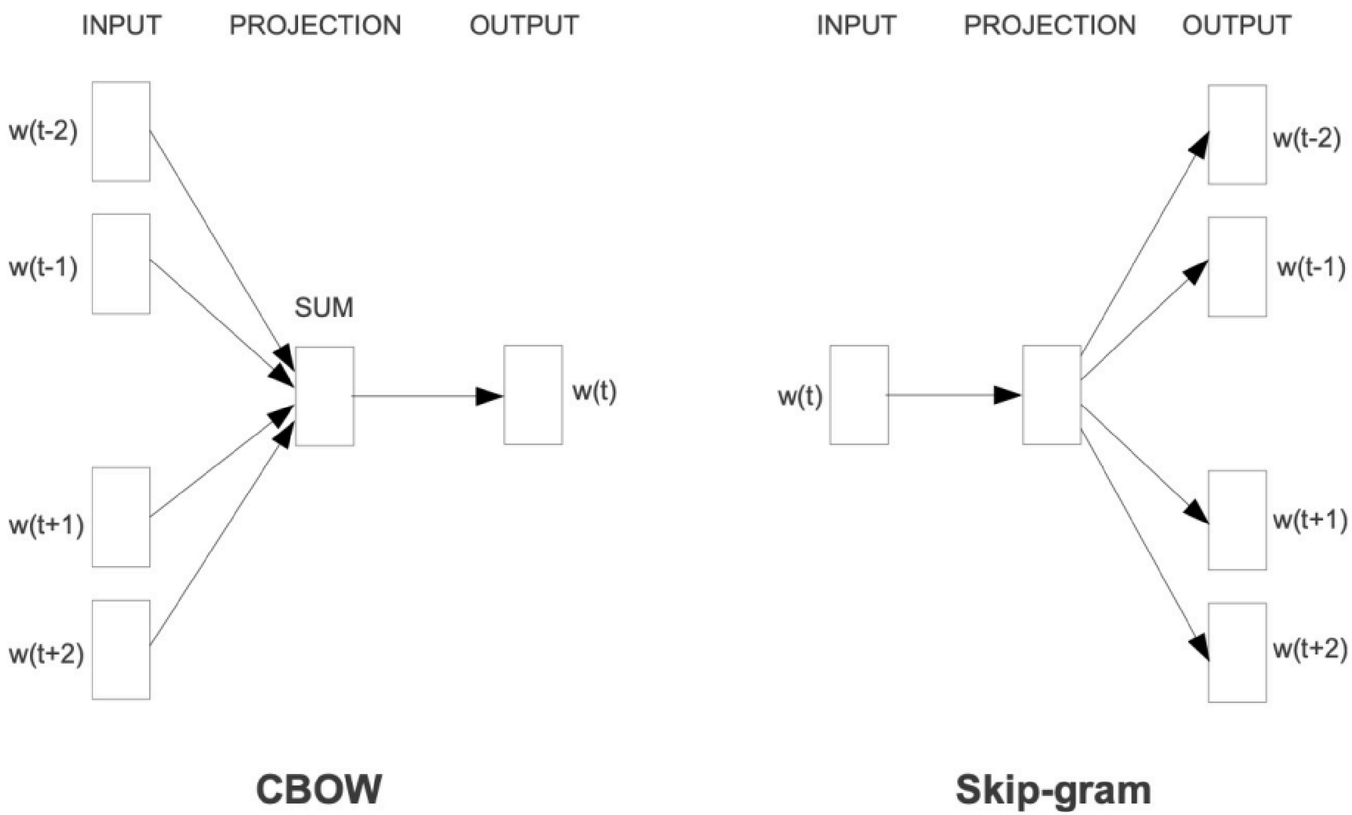
- Node2vec
  - words are represented as continual low-dimensional vectors capturing semantic information
  - Learning representation from the contexts (sentences)

  - Two types for learning word representation:
    - CBOW: Given the context of the word w_t, which includes w_{t-l}, …, w_{t-1}, w_{t+1}, …, w_{t+l}, predict the current word w_t.
    - Skip-gram: Given the word w_t, predict w_{t-l}, …, w_{t-1}, w_{t+1}, …, w_{t+l}.

Slide from Sheng Zhou

Random Walk

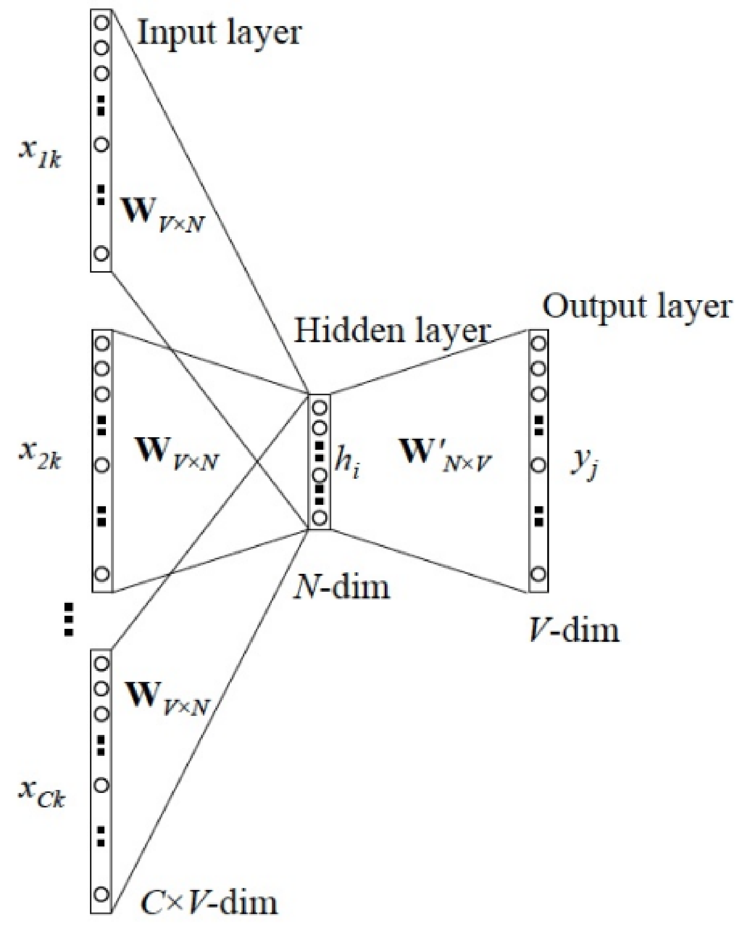
- CBOW and Skip-gram:



Slide from Sheng Zhou

Random Walk

- CBOW:
  - Input: one-hot vectors (ID) of the contexts (C is the number of words in Contexts)
  - One-hot vectors of contexts are multiplied by a matrix W
  - Average these vectors to get the latent vector h_i

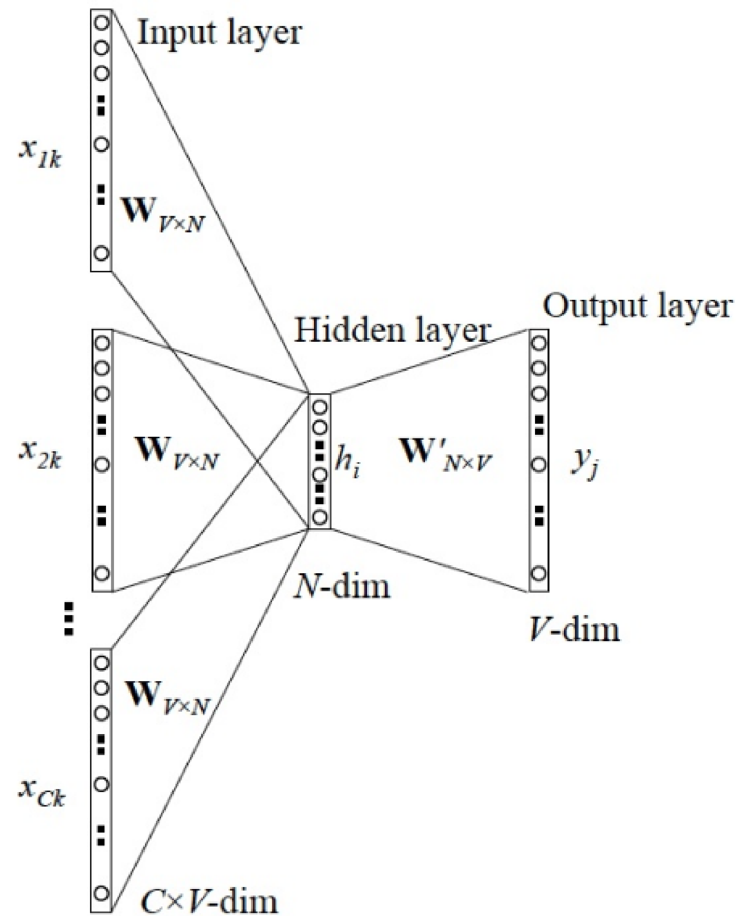


Slide from Sheng Zhou

Random Walk

- CBOW:
  - h_i is multiplied by a matrix W' and then transformed using the softmax function to get the prediction probability of words.
  - Optimizes the model with:

$$\min \quad \mathcal{L}_{CBOW} = -\log(y_j)_p$$

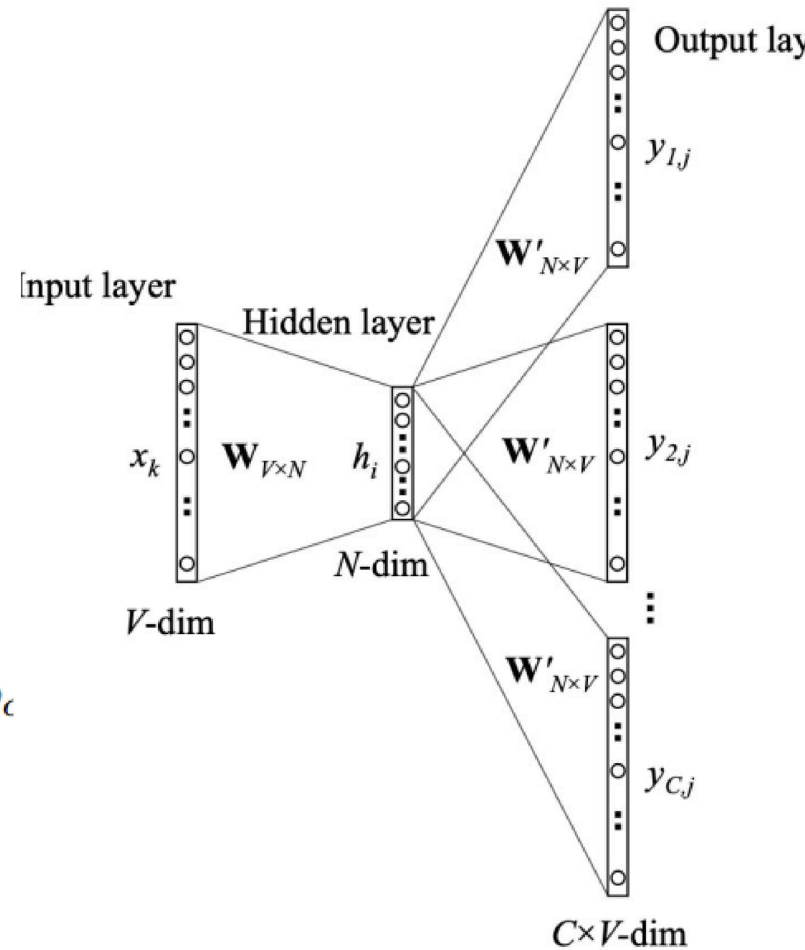


Slide from Sheng Zhou

Random Walk

- Skip-gram:
  - Utilize the current word to predict the contexts.
  - Optimizes the model with:

$$\min \quad \mathcal{L}_{Skip-gram} = -\log \prod_{c=1}^{C} (y_{c,j})_{p_c}$$



Slide from Sheng Zhou

Random Walk

- From the perspective of co-occurring words, Skip-gram optimizes:

$$P(D=1 \mid w,c) = \sigma(\vec{w}\cdot\vec{c}) = \frac{1}{1+e^{-\vec{w}\cdot\vec{c}}}$$

$$P(D=0 \mid w,c) = \sigma(-\vec{w}\cdot\vec{c}) = \frac{1}{1+e^{\vec{w}\cdot\vec{c}}}$$

- Due to the large-scale of the space of words, a strategy is proposed for acceleration:
  - Negative Sampling

Slide from Sheng Zhou

Random Walk

- word2vec utilizes skip-gram with negative-sampling:

$$\log \sigma(\vec{w} \cdot \vec{c}) + k \cdot \mathbb{E}_{c_N \sim P_D} \left[\log \sigma \left(-\vec{w} \cdot \vec{c}_N \right) \right]$$

$$P_D(c) = \frac{\#(c)}{|D|}$$

Slide from Sheng Zhou

Random Walk

- DeepWalk is on the basis of Word2vec with leveraging random walk to generate sequences of nodes
  - The first deep-learning model on the graph
  - Capture high-order neighbor relations:
    - Graph neighbors -> word contexts
  - Unsupervised learning
  - Easy parallelization
  - Fast adaptation

Slide from Sheng Zhou

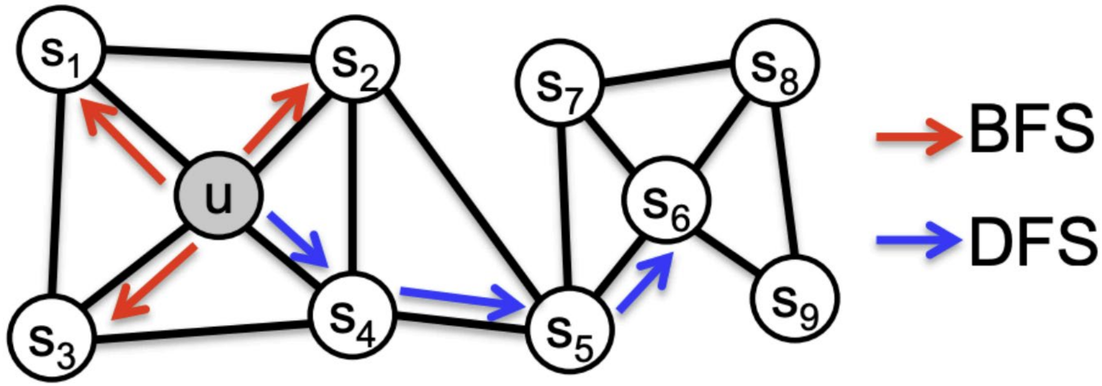
Random Walk

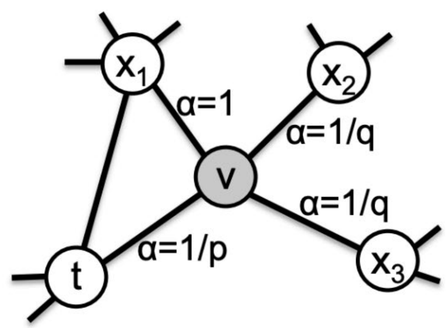
- Deepwalk:
  - Performing random walk from each node to generate multiple sequences of nodes
  - Utilizing word2vec (skip-gram) to learn node representations

Slide from Sheng Zhou

Random Walk

- node2vec:





node2vec: Scalable feature learning for networks

A Grover, J Leskovec - Proceedings of the 22nd ACM SIGKDD …, 2016 - dl.acm.org

… node2vec, an algorithmic framework for learning continuous feature representations for nodes in networks. In node2vec, … We demonstrate the efficacy of node2vec over existing state-of…

☆ 保存 99 引用 被引用次数：12659 相关文章 所有 24 个版本