

# Future Research on Graphs

---

Jeffrey Xu Yu

*The Chinese University of Hong Kong*

---

# Two Issues

- The integration of SQL and graph processing for graph analytics
  - The network embedding
-

# Graph Processing by SQL

## Graph Analytics

PageRank, Shortest Distance,  
Weakly Connected Component,  
Keyword Search,  
Label Propagation,  
Topological Sort, etc.

## Our Enhanced Recursive Query

### SQL'99 recursion

#### Monotonic RA

$\sigma, \pi, \times$   
 $\cup, \bowtie$

Stratified Program  
Least Fixed Point

#### Non-monotonic RA

- MV-Join
- MM-Join
- Anti-Join
- Union-by-Update

XY-Stratified Program  
Iterative Fixed Point

Generated  
SQL/PSM

# We Have Implemented on Spark

## Graph Analytics Tasks

PageRank, Shortest Distance,  
Weakly Connected Component,  
Hyperlink-Induced Topic Search,  
Label Propagation,  
Topological Sort, etc.

### Enhanced Recursive SQL Queries

#### SQL, SQL/PSM

RDBMS  
e.g. Oracle,  
PostgreSQL

SQL-on-Hadoop  
e.g. Hive, Impala

#### Datalog

Datalog Sys  
e.g. Socialite,  
DeALS

#### Java, C/C++

Graph Systems  
e.g. GraphX,  
GraphLab

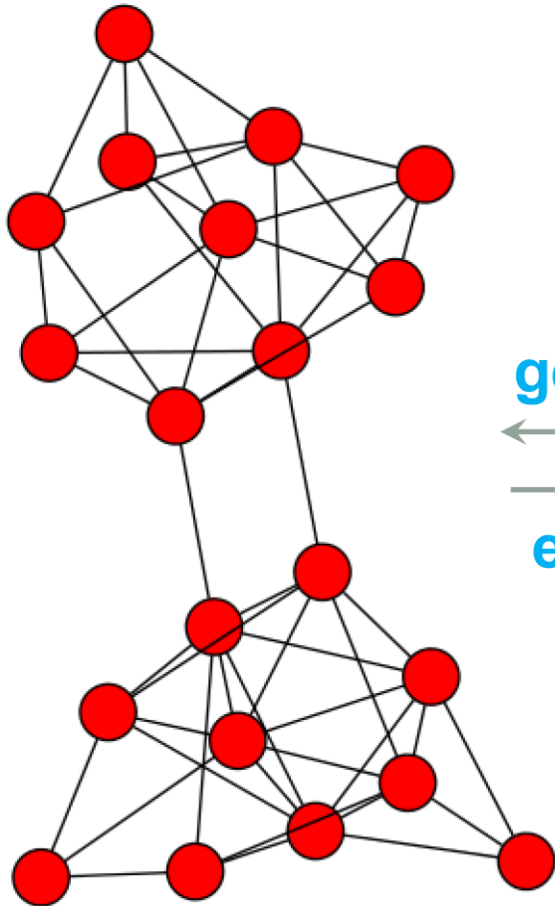
### Infrastructure Platform

Standalone

Distributed

# Network Representations

$$G = (V, E)$$



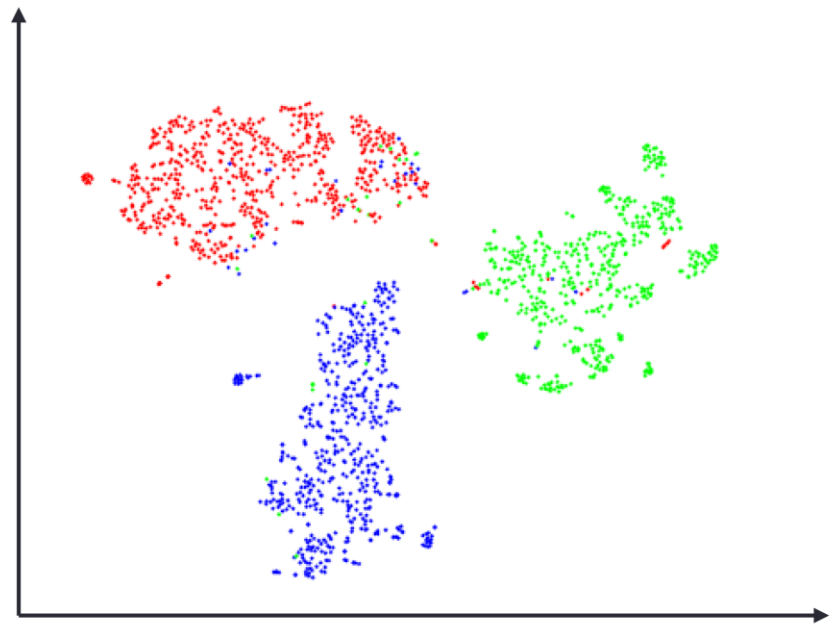
generate



embed

$$G = (V)$$

Low-dimensional Vector Space



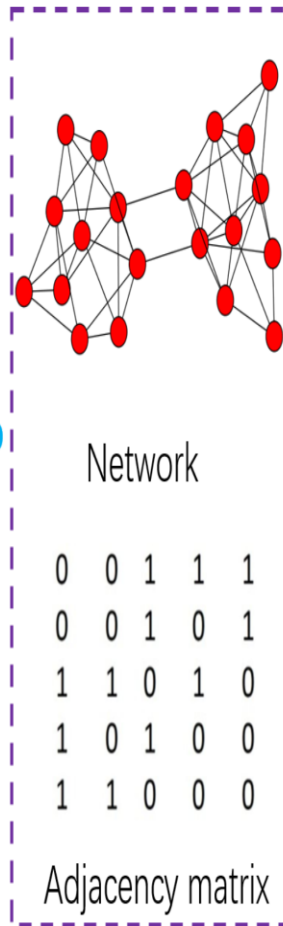
- Easy to parallel
- Can apply classical ML methods

## Traditional topology based network analysis

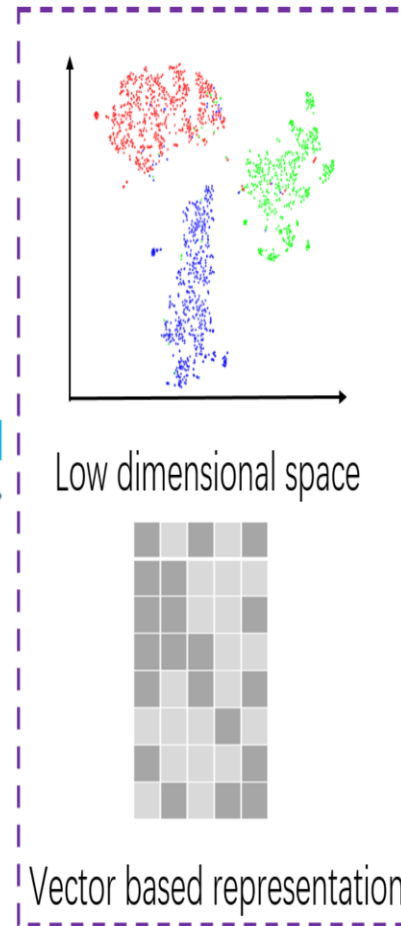
## Network embedding based network analysis

- ❑ Node importance
- ❑ Community detection
- ❑ Network distance
- ❑ Link prediction
- ❑ Node classification
- ❑ Network evolution
- ❑ ...

applied to



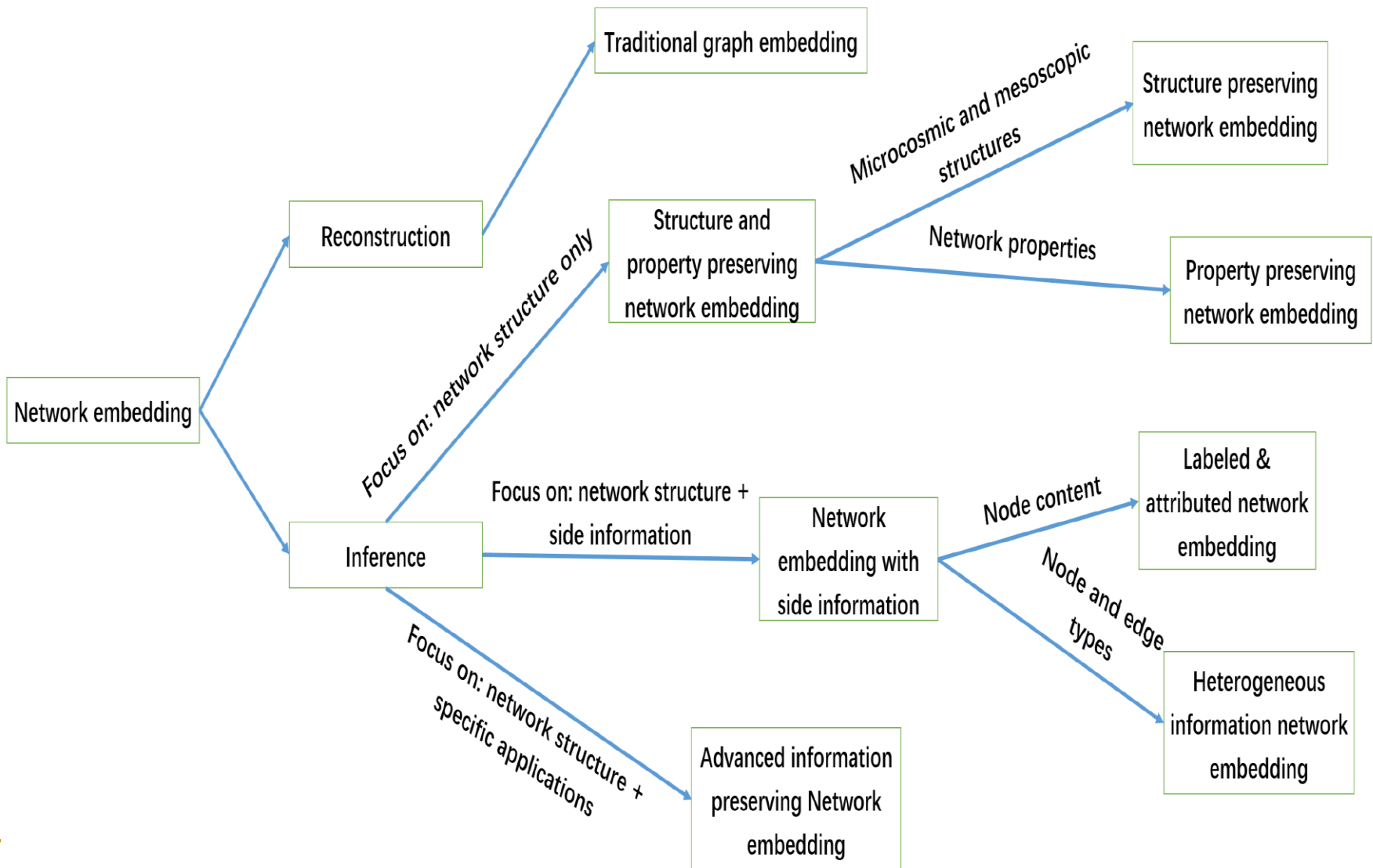
embed



applied to

- ❑ Node importance
- ❑ Community detection
- ❑ Network distance
- ❑ Link prediction
- ❑ Node classification
- ❑ Network evolution
- ❑ ...

# Network Embedding



Taken from the survey by Cui et al.

---

# Main Techniques

- Matrix Factorization
- Random Walk
- Deep Neural Networks



---

# The KDD'17 Tutorials

- Learning Representations of Large-Scale Networks,  
Jian Tang, Cheng Li, and Qiaozhu Mei
- Network Embedding: Enabling Network Analytics and  
Inference in Vector Space, Peng Cui, Jian Pei, and  
Wenwu Zhu

---

# The Surveys on Graph Embedding

- [Graph Embedding Techniques, Applications, and Performance: A Survey](#), Palash Goyal and Emilio Ferrara, CoRR, May, 2017
  - [Representation Learning on Graphs: Methods and Applications](#), William L. Hamilton, Rex Ying, and Jure Leskovec, CoRR, Sep., 2017
  - [A Survey on Network Embedding](#), Peng Cui, Xiao Wang, Jian Pei, and Wenwu Zhu, CoRR Nov., 2017
  - [A Comprehensive Survey of Graph Embedding: Problems, Techniques and Applications](#), Hongyun Cai, Vincent W. Zheng, and Kevin Chen-Chuan Chang, CoRR, Feb., 2018
  - [Knowledge Graph Embedding: A Survey of Approaches and Applications](#), TKDE, Vol. 29, No. 12, 2017
-