北京郵電大學
Beijing University of Posts and Telecommunications

UIC
UNIVERSITY OF ILLINOIS AT CHICAGO

Recent Developments of Deep Heterogeneous Information Network Analysis --Part I Introduction

Chuan Shi

shichuan@bupt.edu.cn

Beijing University of Posts and Telecommunications

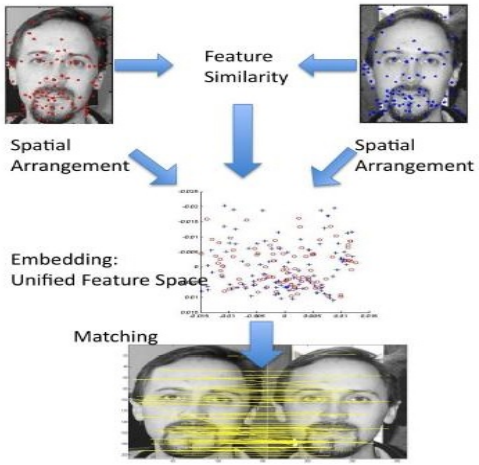
Philip S. Yu

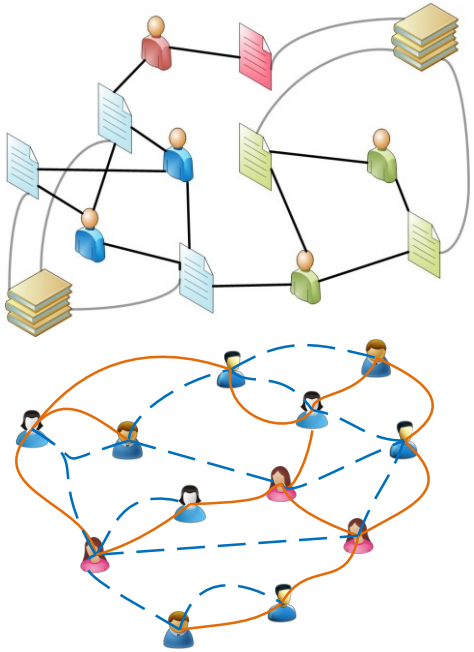
psyu@uic.edu

University of Illinois at Chicago

Motivation

- Roadmap of Data Mining Research





Feature based mining

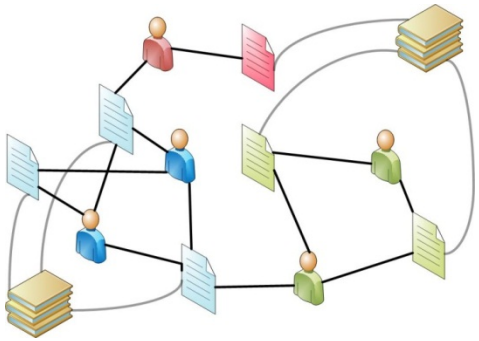
Link based mining

Homogeneous Networks

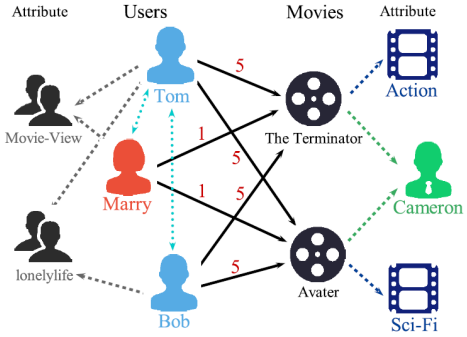
Heterogeneous Networks

Heterogeneous Information Networks(HIN)

- Contain multiple object types and/or multiple link types.

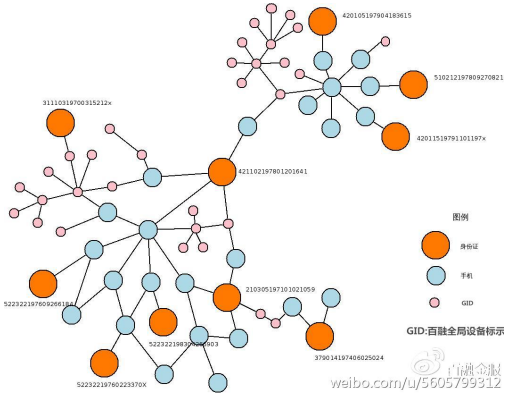


Bibliographic data



Movie data

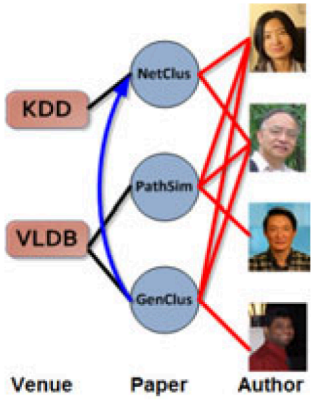
Social network data



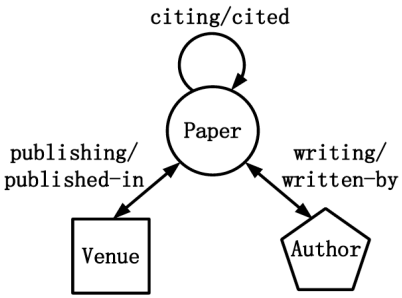
Knowledge graph

Basic Concepts

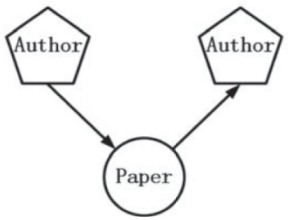
- Network schema
  - Meta-level description of a network
- Meta path (Sun VLDB2011)
  - A relation sequences connecting object pairs
  - Contain rich semantics



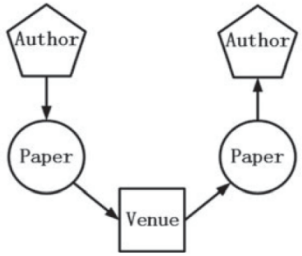
(a) Network instance



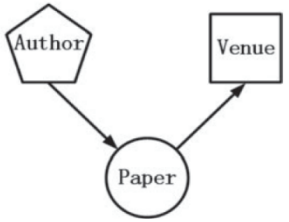
(b) Network schema



(a) APA



(b) APVPA

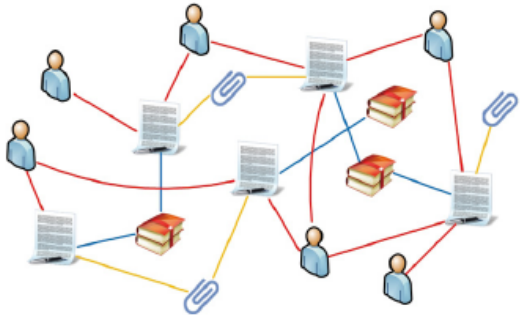


(c) APV

Yizhou Sun, Jiawei Han, Xifeng Yan, Philip S. Yu, Tianyi Wu. PathSim: Meta Path-Based Top-k Similarity Search in Heterogeneous Information Networks. VLDB pp. 992-1003, 2011.

Basic Concepts

- Constrained Meta path
  - Node constrained Meta Path (Li, KAIS 2016)



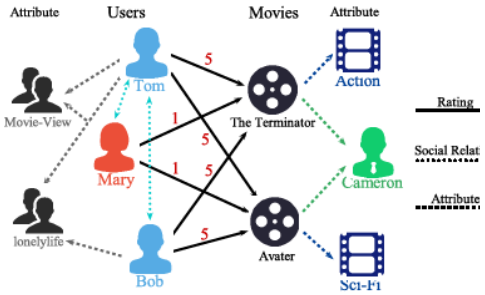
(a) Heterogeneous network

(b) Network schema

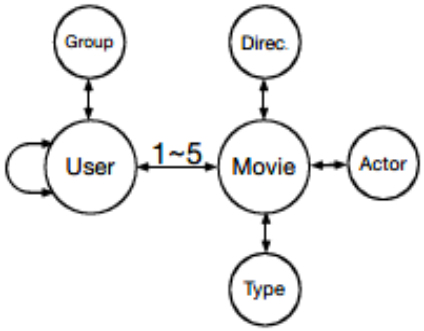
constraint on objects

$$APA|P.L = \text{“}DM\text{”}$$

  - Link constrained Meta Path (Shi, CIKM 2015)



(a) Heterogeneous network



(b) Network schema

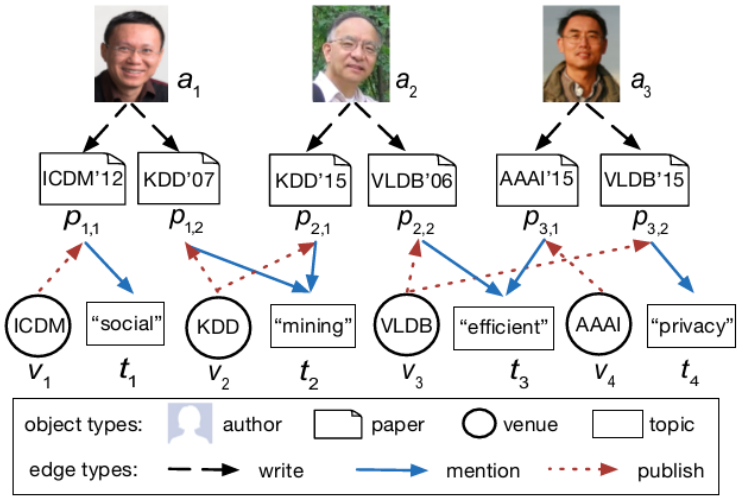
constraint on relations

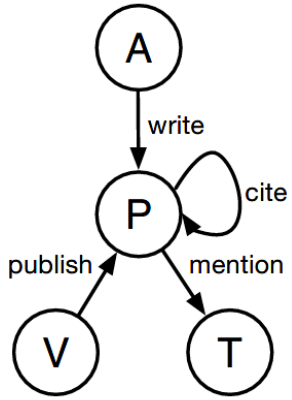
$$\mathrm{U(i)M(j)U}|i = j$$

Chuan Shi, Yitong Li, Philip S. Yu, Bin Wu. Constrained-Meta-Path based Ranking in Heterogeneous Information Network KAIS2016.
Chuan Shi, Zhiqiang Zhang, Ping Luo, Philip S. Yu, Yading Yue, Bin Wu. Semantic Path based Personalized Recommendation on Weighted Heterogeneous Information Networks. CIKM 2015.

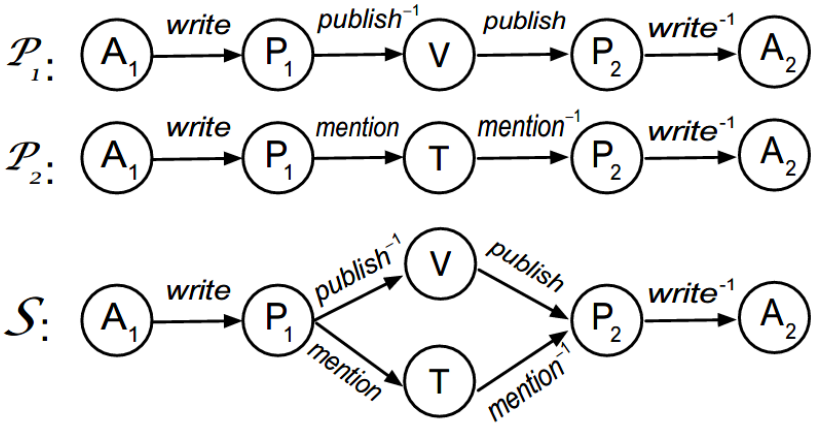
Basic Concepts

- Meta structure/graph (Huang, KDD 2016; Zhao, KDD 2017)





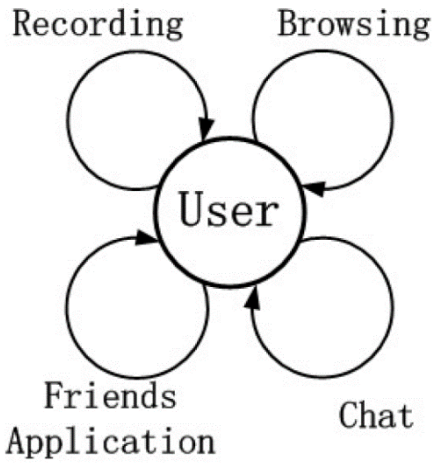
(a) Schema



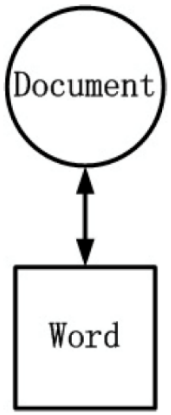
(b) Meta Path, Meta Structure

Zhipeng Huang, Yudian Zheng, Reynold Cheng, Yizhou Sun, Nikos Mamoulis, Xiang Li. Meta structure: Computing relevance in large heterogeneous information networks. KDD 2016.
Huan Zhao, Quanming Yao, Jianda Li, Yangqiu Song, Dik Lun Lee. Meta-graph based recommendation fusion over heterogeneous information networks. KDD 2017.

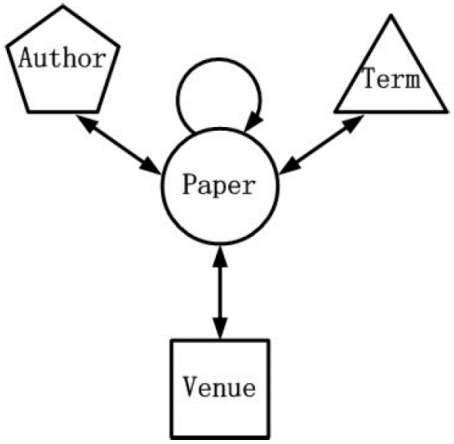
More Examples in Literatures



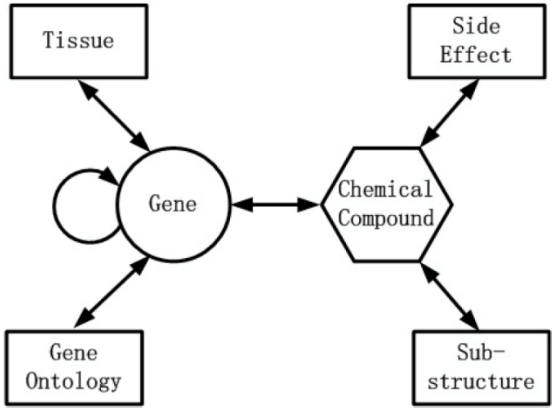
Multi-relational network



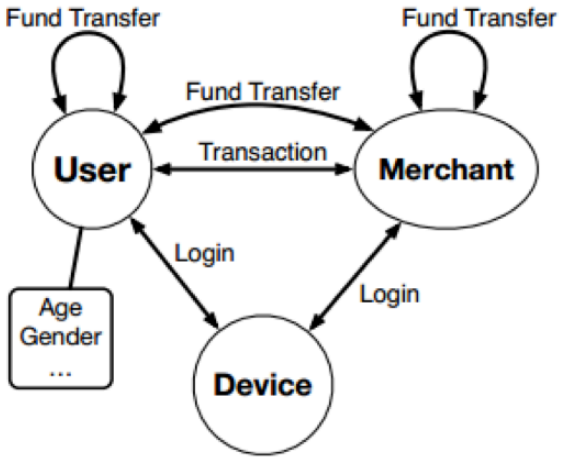
Bipartite network



Star-schema network



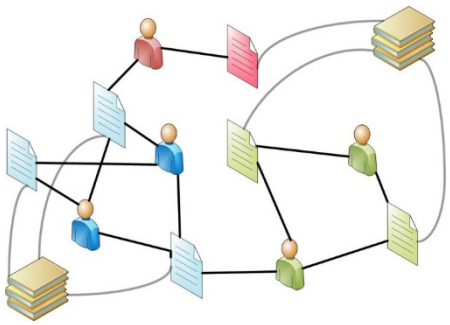
Multiple-hub network



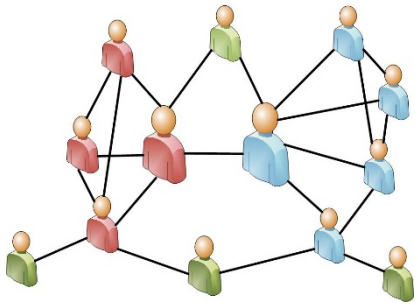
Attributed network

Comparisons with Related Concepts

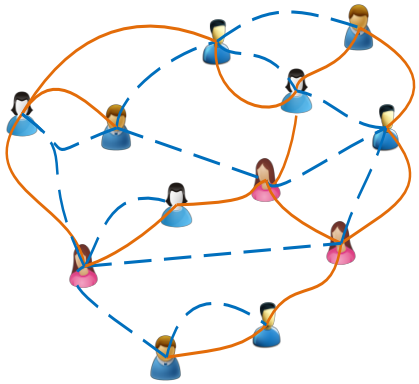
- Heterogeneous network
  - vs homogeneous network
  - vs multi-relational network, multi-dimensional/mode network, composite network
  - vs complex network



Heterogeneous network



Homogeneous network

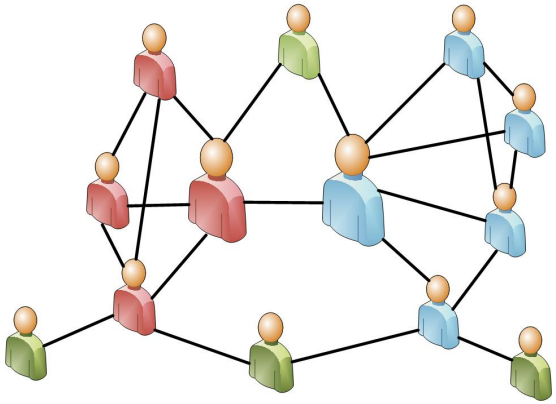


Multi-relational network

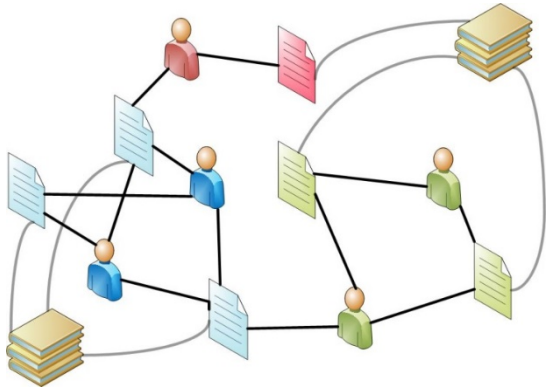
Complex network

Why mine HIN

- Advantages
  - Ubiquitous
  - Comprehensive information
  - Rich semantics
- Challenges
  - Complex structure
  - Mine semantics



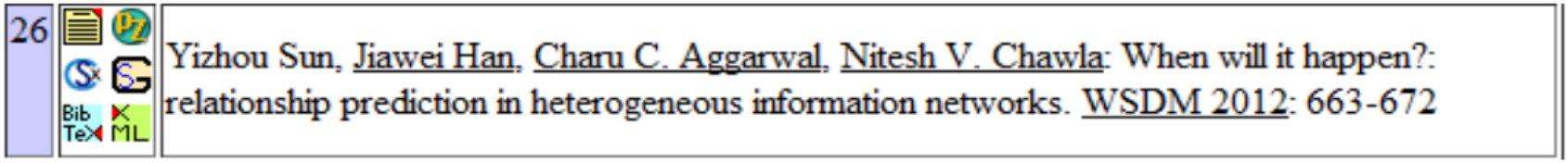
Homogeneous network



Heterogeneous network

What can be mined from HIN

- DBLP: A Computer Science bibliographic database



A sample publication record in DBLP (>1.8 M papers, >0.7 M authors, >10 K venues), …

Knowledge hidden in DBLP Network	Mining Functions
How are CS research areas **structured**?	Clustering
Who are the **leading** researchers on Web search?	Ranking
What are the most essential **terms, venues, authors** in **AI**?	Classification + Ranking
Who are the **peer** researchers of Jure Leskovec?	Similarity Search
Whom **will** Christos Faloutsos **collaborate with**?	Relationship Prediction
Which types of **relationships** are most **influential** for an author to decide her topics?	Relation Strength Learning
How was the field of Data Mining **emerged** or **evolving**?	Network Evolution
Which authors are **rather different** from his/her peers in IR?	Outlier/anomaly detection

[Cited from KDD 2012 Keynote by Prof. Jiawei Han]

Outline

- Metapath based data mining
  - Metapath based similarity measure (VLDB2011, TKDE2014)
  - Metapath based recommendation (CIKM2015, WSDM2014, KAIS2016, KDD2017)
  - Automatic generation of metapaths (SDM2016, TBD2018)
- Heterogeneous information network embedding
  - Shallow models (KDD2017, CIKM2017, KDD2018, TKDE2018, AAAI2018, AAAI2018, KDD2019)
  - Deep models (IJCAI2018, KDD2019, WWW2019)
- Applications (KDD2017, AAAI2019, KDD2019)
- Conclusion and future work