# A Survey of Heterogeneous Information Network Analysis

15352204 lind8@mail2.sysu.edu.cn

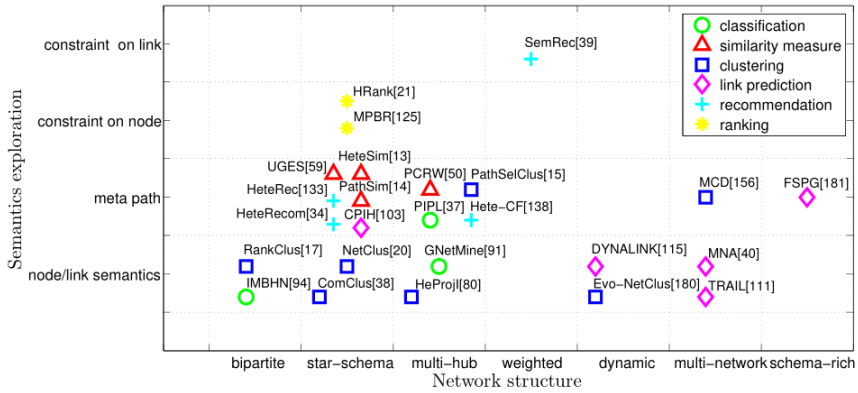
March 18, 2018

## 1 Motivation

Most of the network analysis focus on homogeneous information networks, however the real-world networks are build by multi-typed objects and links, which is called heterogeneous information network (HIN). It contains rich semantics and is an effective tool to fuse more information. Usually, the techniques using in homogeneous networks could not be used directly in HIN, due to the multi-typed objects and links and rich semantics. That is also one of the difficulties in HIN analysis.

## 2 Contributions

- Introduces in-depth understanding of HIN;
- Summarizes contemporary developments of HIN in most data mining tasks;
- Presents the advanced topics and future works of HIN.



# 3 RESEARCH DEVELOPMENTS in HIN

## 3.1 Similarity Measure

- Categorized into two types: feature based approaches and link based approaches.
- Take the meta path connecting into account

**PathSim** Similarity of same-typed objects (based on symmetric paths)

**RelSim** meta-path-based relation similarity

**PCRW** entity proximity in a labeled directed graph

**HeteSim** the relevance of any object pair under arbitrary meta path

**AvgSim** two random walk processes along the given meta path and the reverse meta path

## 3.2 Clustering

- Objects in a cluster share the same topic
- A cluster may include different types of objects
- Utilize Multi-type link, objects' attribute

**Sun et al.'s paper** Considering the incompleteness of objects attributes, different types of links

**Qi et al.'s paper** with outlier links

**TCSC** density-based, vertex attributes

**LSA-PTM**

**CHINC** uses general-purpose knowledge as indirect supervision

**Sun et al.'s paper** semi-supervised clustering algorithm

**Luo et al.'s paper** introduce the concept of relation-path to measure the similarity between same-typed objects; SemiRPClus: semi-supervised learning in HIN

**HeProjI** ranking based, arbitrary schema

## 3.3 Classification

Predict class labels of multiple types of objects simultaneously

**Technique**

- Transductive regression model
- Latent-representation-based classification
- Random-walk-based classification
- Meta-path-based classification
- Ranking-based classification

## 3.4 Link Prediction

- View as simple binary problem: where link exists or not.
- Based on structural properties, attribute information, meta path
- Difference between single HIN and multiple aligned HIN
- Difference between static HIN vs dynamic HIN

**Problems** Dependencies among multiple types of links

**Common tools**

**Probabilistic models**

- MRIP: multi-relational HIN;
- TFGM: semi-supervised learning, a latent topic layer;
- Dong et al.'s paper: transfer-based ranking factor graph model.

**Matrix factorization** JMF joint manifold factorization

## 3.5 Ranking

evaluates object importance or popularity

**Problems**

- In homogeneous networks, only consider the same type of objects, e.g. PageRank, HITS;
- Different meta paths carry different semantic meanings, may have different results;

**Common tools**

- Co-ranking problem on bipartite graphs, co-Rank, co-HITS
- Multi-relation network, multiRank, HAR
- Path based ranking methods, HRank
- HIN in social media network, SocialRank

## 3.6 Recommendation

Search for similarities among items and customer preference.

**Technique**

**Collaborative filtering** Memory-based methods and model-based methods

**Matrix factorization** Factorizes the user-item rating matrix into two low rank user-specific and item-specific matrices

**Meta-path based**

- HeteRecom: semantic-based recommendation system
- SemRec: with attribute values
- context-dependent matrix factorization model: utilizing different contexts information
- cluster-based citation recommendation framework

## 3.7 Information Fusion

Merging information from multiple HIN

## 3.8 Prerequisite

HIN alignment Align the HINs via the shared common information entities, but perfect HIN alignment is a challenging problem

**Common Techniqes**

- Approximated HIN based on various node attribute
- View alignment problem as an anchor link prediction problem
- Partial network alignment methods based on supervised learning setting and PU learning setting
- Transfer-based factor graph model

**Application**

- For link prediction, recommendation and community detection, HIN alignment helps overcome the cold start problem
- Refine the clustering results of the shared entities with information in other aligned networks

# 4 My thinking

We can see that the traditional network analysis is not suit for HIN, and this paper introduce network analysis and application in HIN. My idea is, is there a way to avoid using HIN to represent network with most of the information well-keeping, instead, using homogeneous network with vertex/link attribute. From the very beginning, the HIN is a man-made network from unstructured text data. Select a certain type object which we care most from the multi-type objects, and other type of objects are considered as the attributes information of the main-type object. This will benefit HIN alignment for only predicting the anchor link prediction between same type objects. However, there is still some information loss, e.g. the rich semantic in meta-path. Another useful HIN analysis is network embedding. Recently, BL-MNE: Emerging Heterogeneous Social Network Embedding through Broad Learning with Aligned Autoencoder using DIME (Deep aligned autoencoder based embedding) is one of the research work in aligned heterogeneous network embedding.