

---

## **An Overview of Heterogeneous Information Networks based on Recommendation System**

**Tincy Thomas<sup>1</sup>, Bincy Mathew<sup>2</sup>, Aswathy Manoharan<sup>3</sup>, Nijo Joseph<sup>4</sup>, Binny S<sup>5</sup>**

<sup>1,2,3,4</sup> PG- Department of computer Application, Kristu Jyoti college of management and technology,

<sup>5</sup>Assistant Professor - Department of computer Application, Kristu Jyoti college of management and technology, Kottayam, Kerala

### **ABSTRACT**

Heterogeneous Information Networks (HINs) provide a natural way to represent different relationships among entities of different kinds, as a consequence they are valuable in many domains. Analysing and extracting understanding from HINs normally is based at the idea of meta paths, which are paths in the community schema denoting family members of different semantics among entities. Moreover, real-global HINs are frequently extremely big, containing thousands and thousands of nodes and edges. Therefore, exploring and analysing HINs not only requires interdisciplinary know-how, having the ability both to interpret and pick out suitable meta paths inside the network, but additionally to run the analysis in an efficient and scalable manner. However, there may be a loss of tools to facilitate this assignment. Most real system include a huge wide variety of interacting, multi-typed components, while most current researches version them as homogeneous networks, without distinguishing different forms of objects and links in the networks. In recent years, an increasing amount of works have been proposed to present helper data in recommender system to alleviate information sparsity and cold-start issues. Among them, heterogeneous information networks (HIN)-based recommender systems give a brought together way to deal with meld different helper data, which can be joined with standard suggestion calculations to successfully upgrade the presentation and interpretability of models, and accordingly have been applied in many kinds of recommendation tasks. This paper gives a far reaching and efficient overview of HIN-based recommender systems, including four perspectives: ideas, strategies, applications, and resources.

**Keywords—Heterogeneous Information Network, Data Mining, Meta Path, Recommendation system, Data Science.**

### **1. Introduction**

Many advanced applications depend on investigating a lot of information that include multiple kinds of entities and relationships between them. For example, data-driven science, which has become an exceptionally famous and effective paradigm for logical examination, is in light of computationally investigating huge and heterogeneous datasets. Additionally, the groundworks of the Fourth Industrial Revolution intensely depend on information science procedures for data-driven choice making in light of enormous datasets coming from multiple sources what's more, in different formats. With the fast improvement of the Internet time, people are over-whelmed by a large amount of insignificant data, which is moreover called Information overload. Information overload incredibly influences the efficiency of people in getting helpful information, and recommender systems mean to take care of this issue by providing users with a channel of things that they may be interested in. With many years of improvement, recommender systems have been effectively utilized in many fields like e-commerce and multimedia. Recommender systems frequently depend on historical user behaviour information, in any case, in many situations where users or items are refreshed often, authentic conduct information is very sparse. A more extreme case is cold-start, which intends to give proposals to users, items, or the entire system that have no interaction data. A typical way to deal with lighten the information sparsity and cold-start issue is to present helper data. For instance, in light of online person to person communication administration, we can coordinate social interests or social trust between users as helper information, which is likewise called social recommendation. Also, location-based social networking (LBSN) further adds area data to the social structure, which can likewise be

utilized to improve recommender systems. As a matter of fact, it has turned into a significant test to coordinate different helper data with recommender systems in a brought together way. Heterogeneous information networks (HINs) are complicated networks comprising of multiple kinds of nodes or edges. Since the recommender system itself can be viewed as a bipartite graph comprising of users and items, and parcel of helper data likewise has an intricate organization structure, a recommender system with assistant data can frequently be viewed as a complicated communication system. In this way, displaying such an interaction system with heterogeneous data networks not just normally preserves the entities and relationships in the recommender system, but also effectively incorporates different helper information, consequently successfully mitigating the information sparsity and cold start issues, and somewhat, working on the interpretability of recommender systems.

Lately, a developing number of works have attempted to utilize heterogeneous information networks to model recommender systems consistently, and apply them to increasingly complex recommendation tasks, like ternary recommendation, cross-domain recommendation and sequential recommendation. In spite of the fact that there have been many research works on HIN-based recommender systems, not a single one of them has deliberately figured out and characterized the current functions as well as explained the benefits, impediments and application situations of different methodologies. To connect this hole, we start from the idea, efficiently figure out all the delegate works in the field and characterize them, and make sense of the key works utilizing recipes and representations. Then, we present the application situations of various works, and put together the normal datasets and open-source tools. At last, we examine the future exploration headings. Note that there are a few overviews presenting recommender systems that combine heterogeneous information, but few of them focus on HIN-based recommender systems. Therefore, the main contributions of this study are as per the following:

- We give a complete and orderly investigation for HIN-based recommender systems.
- We classify more than 70 delegate HIN-based recommendation algorithms based on models or application situations, examine the shortcomings of existing works and present expected research directions.
- We provide usually utilized benchmark datasets with their factual information, open-source codes and tool compartments connected with HIN-based recommender systems.

## **2. Methodology**

### **2.1 Recommendation Systems**

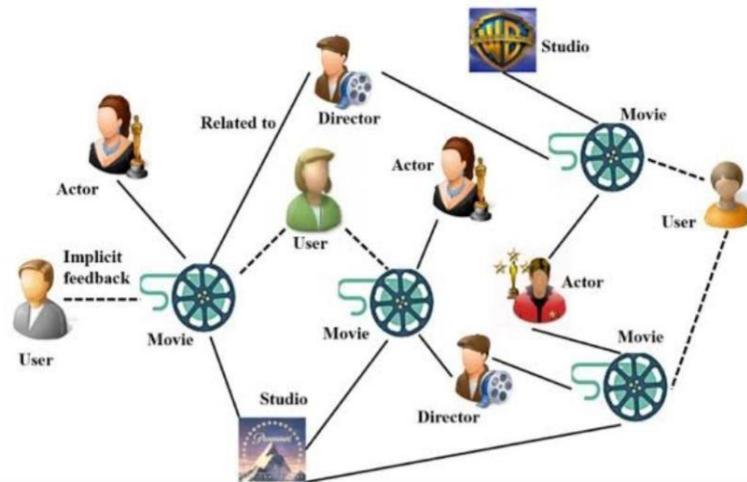
The recommender system is a compelling data sifting strategy that provides users with content that they may be interested in a customized way recommender systems divided into cooperative separating based proposal, content-based, information-based, and hybrid recommendation. Recommender frameworks helps purchasers with making item recommendation that are likely to be interest of consumer such as books, films, and restaurants. It utilizes an expansive scope of strategies from data recovery, measurements, and machine learning to look for similarities among things and client preferences. Collaborative filtering is one of the most famous procedures, which incorporates two kinds of approaches: memory-based techniques and model-based techniques. As of late, matrix factorization has shown its adequacy and effectiveness in recommender frameworks, which factorizes the client thing rating matrix into two low position users specific and item specific.

Recently, a few scientists aware about the significance of heterogeneous data for suggestions. The extensive data and rich semantics of HIN make it promising to create better suggestions. Fig. 1 shows an example of movie recommendation. The HIN not just holds back various kinds of items. Meta-path is all around used to investigate the semantics and concentrate relations among objects.

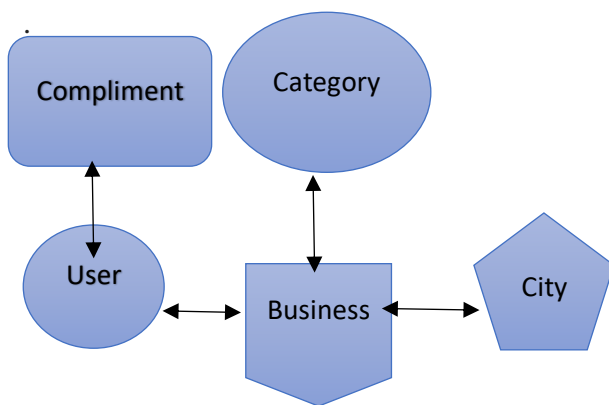
### **2.2 Heterogeneous Information Network**

Many genuine networks can be viewed as intricate organizations created of various sorts of elements and relations. Conventional exploration strategies model them as homogeneous organizations, disregarding the heterogeneity of items and links. To thoroughly display the rich underlying and

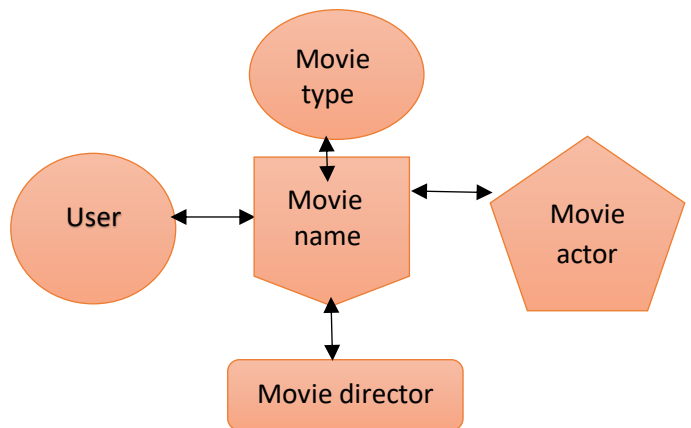
semantic data in complex networks, specialists have proposed the idea of heterogeneous information networks what's more, effectively applied it to the demonstrating of different information mining assignments.



**Fig 1: An example of HIN for movie Recommendation**

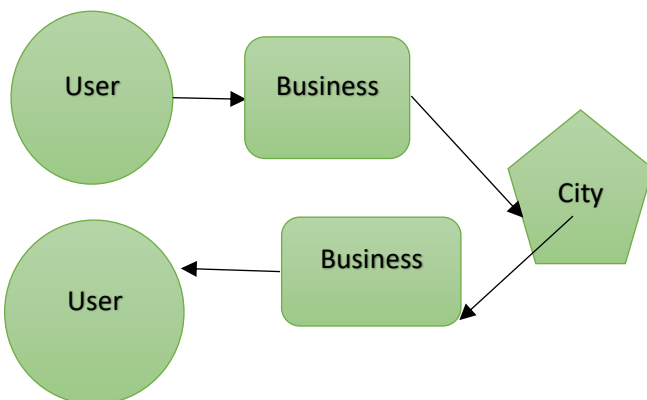


(a)The network schema extracted from Yelp Website

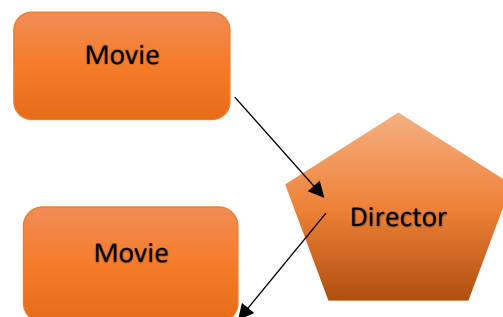


(b)The network schema extracted from Movie systems

**Fig 2: Network schemas extracted from the heterogeneous information networks built based on Yelp website and movie systems.**



(a)Meta-paths extracted from Yelp website



(b) Meta-paths extracted from Movie systems

**Fig 3: Meta-paths extracted from the network schemas of Yelp website and movie systems**

The core terms of heterogeneous data networks are characterized as follows:

### 2.2.1 Information Network

A directed graph  $G = (V, E)$  is utilized to address the information network  $V$  is collection of object node,  $E$  is the set of edges between the organization, and node  $V$  needs to meet the object mapping relationship.  $V \rightarrow A$ , and each node  $v \in V$  belongs to object type set  $A$ . Edge type  $E$  meet link mapping relationship.  $E \rightarrow R$ , which demonstrates that the kind of each edge  $e \in E$  belongs to link type set  $R$ . If  $|A| > 1$  and  $|R| > 1$ , demonstrating information network is an HIN. As is displayed in Figure 2, (a), (b) shows the HIN built based on Yelp site and movie system. We take the HIN based on a movie system for instance, the HIN comprises of three nodes belong to different type includes director, actor and movie. The edge connecting the two nodes addresses a certain connection between the two nodes. For instance, the edge associating the director and the film demonstrates that the director made the film, and the edge associating the film and actor shows that the actor showed up in this movie

### 2.2.2 Network schema

It very well may be characterized as  $TG = (A, R)$ , which is a directed graph characterized on  $A$  and  $R$ . In Figure 3, the network schema is more conceptual than HINs displayed in Figure 2, and depict different type object contained in Figure 2, and the connections between different object types.

### 2.2.3 Meta Path

It is characterized on  $TG (A, R)$  and can be expressed as  $A_1 \xrightarrow{R_1} A_2 \xrightarrow{R_2} \dots \xrightarrow{R_s} A_{s+1}$ . The blend relationship  $R$  somewhere in the range of  $A_1$  and  $A_{s+1}$  can be communicated as  $R = R_1 \circ R_2 \dots \circ R_s$  from the entity  $A_1$  to the entity  $A_{s+1}$ , where  $\circ$  is combination operator of the relationship.  $A_i$  represents the object type and  $R_i$  addresses the relationship type. As is displayed in Fig. 3, two instances of meta-ways separated from network schema of Yelp website and movie system. The meta-path “User-Business-City- Business-User (UBCiBU)” in Fig. 3(a) is the multiple users has been to the business in the same city. And the expression of the meta path “Movie-Director-Movie (MDM)” in the Fig. 3(b) shows different movies were made by the same director.

### 2.2.4 Meta Structure

It refers to a coordinated non-cyclic chart in the organization outline  $TG = (A, R)$ , which determines the beginning Entity  $A_1$  and the objective entity  $A_{s+1}$ , and is made out of numerous meta ways, which is likewise called a meta chart. Accordingly, contrasted and the meta-way, the meta-structure depicts the more significant level semantic in- development between two elements. A meta-design of the movie heterogeneous information networks, which is made out of the two meta-paths, addressing the relationship between two users' appraisals of a similar movie, or the scoring connection between two motion pictures coordinated by a similar director. Because of the conspicuous benefits in mining the rich design and semantic information behind genuine complex systems, heterogeneous information networks have been generally utilized in likeness estimation, Node clustering, grouping, interface forecast, classification, data combination, and different information mining undertakings, and accomplished great outcomes in the above undertakings.

## 2.3 HIN Based recommender system

The recommender system can be viewed as a connection forecast task on the data information networks. Some conventional recommender systems are in light of bipartite chart displaying, hard to utilize different assistant data. A few scientists have proposed strategies for combining specific kinds of assistant data, however they are in many cases not widespread. In ongoing years, HIN-based recommender system has effectively tackled the issue of displaying heterogeneous auxiliary information and user communication ways of behaving in a brought together way, which cannot just really reduce the information scantiness and cold beginning issues in the recommender systems, yet in addition essentially work on the interpretability of the recommender systems, so stand out enough to be noticed and applications. HIN-based proposal comprises of two primary advances: Firstly, we develop a heterogeneous information network in light of the user thing communication information and all assistant data. Then, we plan a proposal model appropriate for the particular heterogeneous information networks. Fig. 1 shows an illustration of HIN-based movie recommender system. It very



well may be seen from the above models that a recommender system in light of a heterogeneous information organization can completely coordinate different organized helper data like interactive information (U-M), social information (U), and attribute information (M-T). It can additionally actually hold rich significant level semantic information in auxiliary information and give interpretability by planning meta-paths and meta-structures. For instance, U-M-U compares to "users who have a similar survey history as the objective user", and can produce recommendation clarifications, for example, "users with a similar review history as you additionally like this movie". The development of heterogeneous, as a matter-of-fact Information organization can be utilized for organized information, yet additionally for unstructured information handled by element extraction and connection extraction, so it has areas of strength for a. In the wake of displaying the different kinds of entities and complex collaboration ways of behaving in the recommender system as a heterogeneous information network, the following thing to do is to plan a suggestion calculation reasonable for heterogeneous data organizations.

### **3. Benchmarks and Open-source tools**

The benchmark datasets generally utilized in heterogeneous information network-based recommender system. In addition, we summarize a few open-source assets in the field of heterogeneous information network-based recommender systems.

#### **3.1 Datasets**

Currently, there are many datasets accessible. We group the datasets into several categories as indicated by their application situations: movie networks, product networks, business organizations, music networks, and book networks. Moreover, we summarize the detailed information of each dataset. For example, node types, link types, normally utilized meta-paths, the existence of timestamps, the existence of properties, and related papers.

- MovieLens: There are various sizes of information like MovieLens 100K, MovieLens 1M and MovieLens 25M. Five kinds of nodes are available: client, movie, genre, age and occupation.
- Amazon: This dataset contains co-viewing and co-buying information between users, including nodes like user, business, class, brand and angle.
- Epinions: This dataset contains trust connections between user also, survey evaluations, including nodes like user, item and class.
- DBLP: This is a academic network that describes the relationship among authors and papers, including nodes like creator, paper, term and venue.

#### **3.2 Open-source code and Tools**

Open-source codes are of incredible worth to academic research, and we have gathered a few delegate executions of heterogeneous information network-based recommender systems. We gather open-source code for the previously mentioned HIN-based recommendation algorithms, with their URLs and programming stages. Early model code was for the most part carried out utilizing Python, MATLAB and C++, while neural network-based models utilized common deep learning structures.

### **4. Future aspects**

There may be found that the heterogeneous information network, has been used in a variety of recommender system models and task with its advantages in improving model performance. In this area, we will introduce a few potential future research aspects.

#### **4.1 Advanced models of HIN-based recommendation**

Now, the most progressive heterogeneous graph recommendation models addressed by the heterogeneous graph neural networks still face many difficulties. These difficulties come not just from the constraints of the graph neural network model itself yet additionally from inflexibility of modelling heterogeneous data. But the ongoing graph neural networks typically utilize a message passing mechanism, which has unfortunate problem to prepare for certain wrong links. For another, the current meta-way put together techniques actually depend with respect to rich master information

to physically plan meta-ways. Step by step instructions to naturally produce significant and meaningful meta-ways or meta-structures is the center test of such techniques. As of late, some work utilizes artificial intelligence and machine learning to create meta-paths, however the improvement of strategy and age effectiveness and their blend with recommendation undertakings are as yet the exploration focal points of future work.

#### **4.2 Developed HIN-based recommendation with external knowledge**

The heterogeneous neural networks introduced in our study just displaying the data of the suggestion situation itself and doesn't present outer information, for example, information base and enormous scope pre-preparing model. Integrating outside information can assist with working on the exhibition of recommendation calculations, however it likewise brings extra difficulties. The past pre-training models have shown their effectiveness in moving external knowledge to specific scenarios, and have been used in recommender systems. Contrasted with homogeneous charts, heterogeneous graph has rich semantic and primary properties, which represent extra challenges for the plan of pre-preparing models and for adjusting them on downstream tasks.

#### **4.3 Enterprise HIN-based recommendation**

Enterprise recommender systems frequently need to deal with super huge scale information and have high necessities for the continuous exhibition of suggestions, and subsequently numerous perplexing proposal models can't be utilized straightforwardly. There are still moves in how to plan a lightweight model that is quicker, more flexible and is appropriate for heterogeneous data organizations.

For another, genuine recommendation scenarios will proceed to create new intelligent information, and the quantity of clients and items will increment step by step. To gain information from new information, numerous proposal models should be retrained when the diagram structure changes, which is clearly unsatisfactory for enormous scope recommendation situations. There are challenges in how to apply it to heterogeneous data organizations, on the grounds that the last option needs to consider the effect of dynamic reports on the primary semantics of heterogeneous charts and requires selective forgetting.

### **CONCLUSION**

The demonstrating technique for heterogeneous information networks can successfully utilize the rich primary and semantic data in helper information and apply it to the displaying of recommender systems. It cannot just actually mitigate the virus start issue brought about by information inadequacy, yet in addition work on the precision and interpretability of proposal models. This paper efficiently researches the related works and assets of HIN-based recommender systems and calls attention to some future research directions. We trust that this review would be able be advantageous for researchers of heterogeneous information networks and recommender systems, and advance the improvement of this field.

### **Reference**

1. Shaohua Fan, Junxiong Zhu, Xiaotian Han, Chuan Shi, Linmei Hu, Biyu Ma, Yong liang Li Meta path-guided heterogeneous graph neural network for intent recommendation Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (2019), pp. 2478-2486
2. Zeakis, T. Vergoulis, and D. Skoutas. 2020. SPHINX: A System for Meta path-based Entity Exploration in Heterogeneous Information Networks. In Proceedings of the 46th International Conference on Very Large Data Bases (VLDB 2020).
3. C. Shi, Y. Li, P. S. Yu, and B. Wu. 2016. Constrained-meta-path-based ranking in heterogeneous information network. Knowl. Inf. Syst. 49, 2 (2016), 719–747.
4. Y. Li, J. Zhang, Y. Sun, and P. S. Yu. 2017. A Survey of Heterogeneous Information Network Analysis. IEEE TKDE 29, 1 (2017), 17–37.

5. L. Yao, J. Tang J. Li, and Z. Su. 2008. ArnetMiner: Extraction and Mining of Academic Social Networks. In Proceedings of the 14th ACM SIGKDD. ACM, 990–998.
6. The Gene Ontology Consortium. 2018. The Gene Ontology Resource: 20 years and still Going strong. *Nucleic Acids Research* 47, D1 (11 2018), D330–D338.
7. Y. Zhu, and P. S. Yu. 2015. Top-k Similarity Join in Heterogeneous Information Networks. *IEEE Trans. Knowl. Data Eng.* 27, 6 (2015), 1710–1723.
8. L. Hou U., K. Yao, and H. Mak, “Pathsimext: Revisiting pathsim in heterogeneous information networks,” in *WAIM*, 2014, pp. 38–42.
9. C. Aggarwal, M. H. Tsai and T. Huang, “Ranking in heterogeneous social media,” in *WSDM*, 2014, pp. 613–622.
10. J. Han, “Mining heterogeneous information networks by exploring the power of links,” in *Discovery Science*, 2009, pp. 13–30
11. Yao, Zhao, Yao, Quanming, Jianda, Yangqiu, Dik Lun Lee, 2017. Meta-graph-based recommendation fusion over heterogeneous information networks. In: *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 635–644.
12. Ye, Liqiang, Yao, Mengqiu, Zhenyu, Wang, Jianming, Xiao, Jing, 2020. A heterogeneous information network based cross domain insurance recommendation system for cold start users. In: *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 2211–2220.
13. Yin, Yu, Junliang, Gao, Jundong, Hongzhi, Liu, Huan, 2018. Adaptive implicit friends identification over heterogeneous network for social recommendation. In: *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pp. 357–366.
14. Andreas, Pfadler, Wang, Jizhe, Wang, Lifeng, Huang, Dik Lun Lee, 2020. Billion-scale recommendation with heterogeneous side information at taobao. In: *2020 IEEE 36th International Conference on Data Engineering (ICDE)*. IEEE, pp. 1667–1676.
15. Tuan-Anh Nguyen, Pham, Li, Xutao, Zhang, Zhenjie, 2016. A general recommendation model for heterogeneous networks. *IEEE Trans. Knowl. Data Eng.* 28 (12), 3140–3153.