Heterogeneous Information Networks: the Past, the Present, and the Future

Yizhou Sun
UCLA
Los Angeles, CA
yzsun@cs.ucla.edu

Jiawei Han
UIUC
Urbana-Champaign, IL
hanj@illinois.edu

Xifeng Yan
UCSB
Santa Barbara, CA
xyan@cs.ucsb.edu

Philip S. Yu
UIC
Chicago, IL
psyu@uic.edu

Tianyi Wu
Meta
Bellevue, WA
tianyiwu@fb.com

ABSTRACT

In 2011, we proposed PathSim to systematically define and compute similarity between nodes in a heterogeneous information network (HIN), where nodes and links are from different types. In the PathSim paper, we for the first time introduced HIN with general network schema and proposed the concept of meta-paths to systematically define new relation types between nodes. In this paper, we summarize the impact of PathSim paper in both academia and industry. We start from the algorithms that are based on meta-path-based feature engineering, then move on to the recent development in heterogeneous network representation learning, including both shallow network embedding and heterogeneous graph neural networks. In the end, we make the connection between knowledge graphs and HINs and discuss the implication of meta-paths in the symbolic reasoning scenario. Finally, we point out several future directions.

PVLDB Reference Format:
Yizhou Sun, Jiawei Han, Xifeng Yan, Philip S. Yu, and Tianyi Wu. Heterogeneous Information Networks: the Past, the Present, and the Future. PVLDB, 15(12): 3807 - 3811, 2022.
doi:10.14778/3554821.3554901

1 INTRODUCTION

With the advent of different types of networked data, such as the Web network, social networks, and bibliographic networks, there was a huge demand for analytical tools on such data, on which the traditional data mining and machine learning tools are no longer applicable due to the violation of i.i.d. assumptions. Seminal papers such as PageRank [34], HITS [26], and SimRank [23] set foot in networked data and laid foundations on research in this direction.

These studies, however, focus only on networks with one type of nodes and one type of links, which are true for Web and social networks. But there are many complicated networks where the nodes and links belong to multiple types. For example, in bibliographic network (illustrated in Fig. 1), we have different types of nodes, such as authors, papers, venues, and keywords, and different types of links, such as paper$\xrightarrow{\text{cites}}$ paper, author$\xrightarrow{\text{writes}}$ paper, and paper$\xrightarrow{\text{published in}}$ venue. It is critical to preserve these type information, as they are associated with different semantic meanings. The innovations of our PathSim paper [45] are three-folds: (1) we proposed a general definition of *heterogeneous information networks* (HINs) and introduced the concept of *network schema* to describe the meta structure of an HIN; (2) we proposed *meta-path* to systematically capture the high-order relationship between any two nodes in the networks, which is defined as a sequence of relations between objects following the schema; and (3) we proposed a specific meta-path-based similarity measure called *PathSim* to quantify the similarity between two nodes for a specific meta-path or a set of meta-paths.

PathSim sets a solid foundation for HIN schema design and meta-path-based algorithmic framework. Its impact has been across a broad spectrum of research communities including database, data mining, machine learning, artificial intelligence, network science, and graph neural networks (deep learning), as well as diverse applications in e-commerce, biomedical domains, academic graphs, and cybersecurity. Moreover, industry has embraced the concept of HIN and/or meta-path in their products and systems including Microsoft, Amazon, Meta, Alibaba, and Twitter to name a few.

When PathSim was first proposed 11 years ago, it still falls into the feature engineering regime. Recently, representation learning, which aims at automatically extracting "good" features from data, receives wide attention in both academia and industry. Nevertheless, HIN and meta-path still play a critical role in this new era, due to two reasons. First, with the development of all kinds of new IT techniques, applications of HINs are expanding rapidly, covering social media, e-commerce, healthcare, and cybersecurity. Second, the concept of meta-path touches the fundamentals of how to understand the semantic meanings of relations between objects, and it is frequently used as the basic semantic unit for proximity definition in heterogeneous network embedding and message passing in heterogeneous graph neural networks.

Most of the HIN applications have a relative simple schema, meaning with a small number of node types and link types, until the era of knowledge graph (KG) comes (again). Interestingly, although we mentioned the potential application of "knowledge network" in our PathSim paper, the deeper discussions on the connections to

This work is licensed under the Creative Commons BY-NC-ND 4.0 International License. Visit https://creativecommons.org/licenses/by-nc-nd/4.0/ to view a copy of this license. For any use beyond those covered by this license, obtain permission by emailing info@vldb.org. Copyright is held by the owner/author(s). Publication rights licensed to the VLDB Endowment.
Proceedings of the VLDB Endowment, Vol. 15, No. 12 ISSN 2150-8097.
doi:10.14778/3554821.3554901

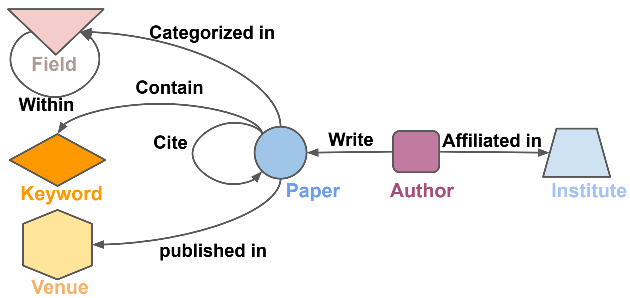

Figure 1: Network schema of a bibliographic network, where different shapes of nodes denote different node types and labels on the links denote link types.

KG came much later [51]. KGs have a much richer network schema, with hundreds or even thousands of relation types. In addition to the flat structure which we call instance view between entities, KGs could also have hierarchical structure which we call ontological view between concepts or between entities and concepts [16]. KGs enable the storage of knowledge in addition to data, and to some extent blur the boundary of data and knowledge. Meta-path in this setting naturally connects to the conjunction of predicates, and enables symbolic reasoning in the world of representation learning, pointing to a very exciting and promising new research frontier.

In this retrospective paper, we will review the development of HINs and meta-path in Section 2, summarize the recent advances on heterogeneous network embedding in Section 3 and heterogeneous graph neural networks in Section 4, discuss the connections between knowledge graphs, symbolic reasoning and heterogeneous information networks in Section 5, and finally conclude the paper in Section 6.

2 HETEROGENEOUS INFORMATION NETWORKS AND META-PATHS

As introduced in Section 1, earlier studies focus on homogeneous information networks (in contrast to HINs). The main observation by us was that many information networks do have multiple types of objects and relations, and we proposed solutions to several special types of HINs, such as RankClus for bi-typed networks [46] and NetClus for HINs with star schema [48]. In PathSim, for the first time, we formally gave the definition of a general HIN and introduced the concept of network schema, which was inspired from the entity-relationship (ER) graph from the database community. With the network schema, which can be considered as a meta-level graph with object types as nodes and relation types as edges, we can explicitly write out what the object types are and what types of (binary) relations are valid between object types. An example of network schema can be found in Figure 1.

How to define similarity between objects in a network is a fundamental problem to tasks such as classification, clustering, and similarity search. In the homogeneous network setting, a natural assumption is that if two objects share lots of common neighbors their similarity is higher. When coming to HIN setting, it is not that straightforward, as the neighbors of an object come from different object types and are connected by different relation types. For example, a paper can be *cited by* other papers, *written by* authors, *published in* a venue, and *contains* keywords. Which of these relations play a more important role in determining the similarity between two papers? In addition, these relations can be composited together to form a higher-order relation, which corresponds to a path in the network schema, and thus called *meta-path*. For example, an author can be connected to another author via a meta-path author $\xrightarrow{\text{writes}}$ paper $\xrightarrow{\text{written by}}$ author, denoting the co-authorship. Symmetric meta-path such as paper $\xrightarrow{\text{cited by}}$ paper $\xrightarrow{\text{cites}}$ paper can naturally capture the similarity between two objects.

When coming to the link prediction task (e.g., recommendation), meta-paths provide a systematic way to extract features for a pair of nodes. For example, when deciding whether two authors will become co-author in the future [42], we can examine the connectivity according to meta-paths such as author $\xrightarrow{\text{writes}}$ paper $\xrightarrow{\text{written by}}$ author (i.e., two authors are co-authors), author $\xrightarrow{\text{writes}}$ paper $\xrightarrow{\text{published in}}$ venue $\xrightarrow{\text{publishes}}$ paper $\xrightarrow{\text{written by}}$ author (i.e., two authors publish in the same venues), and author $\xrightarrow{\text{writes}}$ paper $\xrightarrow{\text{contains}}$ keywords $\xrightarrow{\text{contained in}}$ paper $\xrightarrow{\text{written by}}$ author (i.e., two authors publish papers sharing common keywords), and feed proximity measures (e.g., count) defined on them as features to any classification or learning-to-rank model. Lots of mining problems on HINs then are reduced to meta-path selection problem [44, 47].

Applications and Deployment. In this line, there are a couple of quite successful applications, such as co-author prediction [42], citation prediction [30], and topic diffusion [15] in bibliographic networks, recommendation [58], drug-target interaction prediction [11], Android malware detection [57], and social network alignment [60, 61]. A summary of our earlier work can be found in [43] and a more thorough survey in this line can be found in [40].

3 HETEROGENEOUS NETWORK EMBEDDING

Feature engineering has gradually been replaced by automatic representation learning, which is also the case in the network setting. Inspired by the word2vec [32] algorithm that maps discrete words into embedding vectors, several network embedding approaches that map nodes into embedding vectors are proposed, such as DeepWalk [36], LINE [50], and node2vec [14]. These methods make the assumptions that nodes in the same neighborhood (context) should be similar to each other and thus share similar embeddings, which, again, becomes much more sophisticated in the HIN case.

Similar to the roles played by meta-paths in the traditional mining tasks, meta-paths provide a systematical way to define context in HINs and thus the embeddings can be learned accordingly. Metapath2vec [7] is the most representative work in this direction. The main idea of metapath2vec is to generate node sequences following meta-path specific random walk, and the context of a node is the nearby nodes along those sequences. Later, HIN2Vec [12] proposes an embedding method that models the probability for a given pair of nodes following a specific meta-path, which learns the embeddings not only for nodes but also for each meta-path. The idea of representing meta-path is closely related to the idea of representing relations in KG that will be discussed in Section 5.

Applications and Deployment. With the power of embedding, the applications of this line of research has been significantly enriched and some of them have been deployed in industry-level systems. Meta-path-based network embedding or in general heterogeneous network embedding have been successfully applied in author identification [4], question and answering system [28], recommendation system in Alibaba [3], recommendation in other settings [39], similarity search in Microsoft Academic Graph [7], disease diagnosis [17], malware detection [10], and time series prediction [24]. Twitter has successfully deployed industry-scale embedding for twitter HIN [9], which has significantly enhanced different downstream tasks, including personalized ads ranking, malicious content detection, and account recommendation. Several more detailed surveys can be found in [8, 55].

4 HETEROGENEOUS GRAPH NEURAL NETWORKS

Network embedding approaches discussed in Section 3 are also called shallow embedding, as it can be considered as a linear mapping from one-hot encoding vector. This causes two types of issues. First, the number of parameters is extremely large for large networks, as each node is associated with a learnable vector. Second, it cannot handle node (and edge) attributes that are typically associated with the network. For example, we usually have product description in a product recommender system. This motivates the design of Graph Neural Networks (GNNs), which applies deep learning to graphs. A representative algorithm is Graph Convolutional Network [25], which transforms each node's feature into a representation vector following a message passing framework. More concretely, in each layer of GNN, each node collects "messages" from their neighbors, aggregates these messages, and then applies a non-linear transformation to the aggregated message. The parameters are in these transformations which are shared across different nodes. In the inference stage, it can easily handle nodes that are not seen in the training stage, as long as its feature and neighborhood information is given. It is obvious that the message passing framework also heavily relies on the definition of "neighbors", which is straightforward in the homogeneous networks but much trickier in the HIN case.

This line of research is still very active [41], and we introduce several representative algorithms below. R-GCN [37] differentiates the transformation for each relation type, which is associated with a separate set of parameters and thus handles edge heterogeneity. HAN [53] extends the neighbors from 1-hop neighbors to high-order neighbors that are defined by meta-paths. Given a meta-path (e.g., author $\xrightarrow{\text{writes}}$ paper $\xrightarrow{\text{written by}}$author), HAN aggregates messages from that meta-path determined neighbors (e.g., co-authors) with a meta-path specific attention mechanism, followed by a meta-path specific non-linear transformation. Then it aggregates those embeddings again over different meta-paths via a second-level attention mechanism. BA-GNN [22] applies a two-level attention at both relation-level and node-level, and designs a node-specific attention for each relation to make the GNN more interpretable. HetGNN [59] takes special treatment for each type of nodes, and then designs a node type-level attention mechanism to aggregate embeddings from different node types. HGT [19] proposes to use *meta-relation* as the basic message passing unit, which can be considered as a length-1 meta-path that encodes both node types and relation type information. A special parameter sharing design is used for both meta-relation-based message passing and meta-relation-based attention mechanism. More interestingly, by reading out the importance score for each meta-relation, we can detect the most important meta-paths.

Applications and Deployment. Since GNN based approaches are in the big umbrella of representation learning, all of the applications mentioned in Section 3 can benefit from heterogeneous GNNs, which are more powerful in general. For example, they have been applied to cyber security [13], social event detection [2, 35], rumor detection [27], and recommender systems [38, 56]. There are also several successful industry-level deployments. For example, HGT [19] has been successfully deployed in Microsoft Office Graph; DHGAT [33] has been deployed to Taobao's search service, which is the largest e-commerce platform in China; and others have been deployed in advertising business in Alibaba [31] and product recommendation in Amazon [63]. These approaches can also be used to solve KG completion task (more details in Section 5), which aims to infer the tail entity given the head entity and the relation, such as to infer the capital city of France.

Benchamark Datasets and Code. The HIN community has accumulated several benchmark datasets for fair evaluation, such as the Open Academic Graph (OAG), which are hosted on Open Graph Benchmark Platform (OGB) [18]. Open source graph learning libraries such as PyG[1] and Amazon DGL[2] have provided Heterogeneous GNN implementation workflow, which significantly boost the research and applications in this direction. The recent book [41], the two surveys [8, 55], and the BA-GNN paper [22] have provided a comprehensive introduction to heterogeneous GNNs. A comprehensive survey on HIN-based recommender systems can be found in [29].

5 KNOWLEDGE GRAPHS, SYMBOLIC REASONING, AND HETEROGENEOUS INFORMATION NETWORKS

Knowledge Graph (KG) or Knowledge Base (KB) has been used in expert systems to conduct symbolic reasoning in early days, which is a collection of triples denoted as (h, r, t) (head entity, relation, and tail entity) as well as a graph with entities as nodes and relations as edges. An example of KG can be found in Figure 2. More recently, KGs become extremely important to provide external background knowledge for many applications, such as search engine, question and answering system, dialogue systems, and e-commerce. KG-based reasoning such as KG completion receives wide attention. Then a natural question is: *what is the relationship between KG and HIN?* In our view, KGs are special cases of HINs, but with much richer schema than most of the HINs that we have discussed so far.

[1]https://pytorch-geometric.readthedocs.io/en/latest/notes/heterogeneous.html
[2]https://docs.dgl.ai/en/0.6.x/tutorials/basics/5_hetero.html

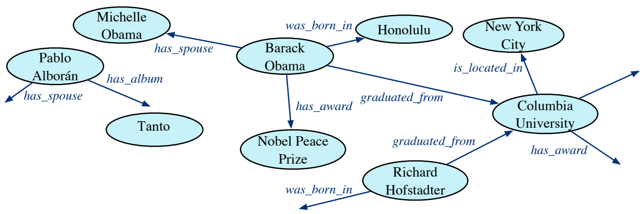

Figure 2: An example of KG, where nodes denote entities and links denotes relations. (Figure is adapted from [16])

Similar to the HIN representation learning, KG community nowadays embraces representation learning and conducts symbolic reasoning in a differentiable way. Earlier KG embedding approaches are shallow embedding methods similar to the ones mentioned in Section 3, but provide embeddings for both entities and relations. More precisely, relations are considered as algebraic operators that transform the head entity to the tail entity. For example, in TransE [1], a relation is considered as a translation operation, which is represented as a vector with the same dimensionality as the entities. In DistMult [54], a relation is considered as a linear transformation, which is represented as a matrix (simplified as a diagonal matrix in DistMult). In RotatE [49], the entities are represented as vectors in complex space, and a relation is considered as a rotation operation, which is represented by a vector of rotation angles. Then the losses are evaluated based on the projected tail entity and the ground truth tail entity. Not surprisingly, some of the heterogeneous GNNs can be applied to KG completion tasks directly, such as R-GCN [37] and BA-GNN [22].

More interestingly, from logic perspective, a triple in KG is a ground binary predicate, e.g., $(Miller, liveIn, USA)$ corresponds to *liveIn (Miller, USA)*; a path in KG is a conjunction of predicates, e.g., $\text{Thomas} \xrightarrow{\text{liveIn}} \text{USA} \xrightarrow{\text{hasOfficialLanguage}} \text{English}$ corresponds to *liveIn (Thomas, USA)* ∧ *hasOfficialLanguage (USA, English)*; and a meta-path can be considered as the body of a logical rule, e.g., the meta-path $\text{person} \xrightarrow{\text{liveIn}} \text{country} \xrightarrow{\text{hasOfficialLanguage}} \text{language}$ corresponds to the body in $\text{speakLanguage}(person, language) \leftarrow \text{liveIn}(person, country) \wedge \text{hasOfficalLanguage}(country, language)$.

Applications. With this view, we can enhance the KG embedding tasks by incorporating logical rules (e.g., UniKER [6]) and explore new applications such as mining logical rules from KGs (e.g., RLogic [5]). This direction is still very new, and will inspire us to define new techniques for HINs.

6 FUTURE DIRECTIONS AND CONCLUSION

Where are we going from here? We list several challenging and promising directions below.

- Scalability. For web-scale applications such as Microsoft Office Graph and LinkedIn Economics Graph, how to make meta-path-based approach more scalable is always a challenging question. In PathSim, we proposed partial materialization to address the issue. For heterogeneous GNNs, methods such as smart sampling will be a promising direction.
- Deeper understanding of heterogeneity. From earlier days' meta-path-based feature engineering to more recent meta-relation/meta-path-based relation projection in heterogeneous GNNs, we have deeper and deeper understanding on what heterogeneity implies. Recently, several studies propose that heterogeneity could also mean the nodes in the network belong to different geometric spaces [21, 52].
- From meta-path to meta-structure. From logic perspective, a good meta-path should correspond to a high-quality logical rule, where the rule body is a conjunction of predicates. Can we extend it to model more complicated dependency? There are some studies to extend meta-path to meta-structure [20] or meta-graph [62], and there is plenty room to explore in this direction.

In all, in the past decade we have evidenced the booming of research on HINs in different research communities, the successful applications of HINs across different domains, and the deployment of HIN techniques at industry-level systems. We are looking forward to more exciting work in this area.

ACKNOWLEDGMENTS

This paper was partially supported by NSF III-1705169, NSF 1937599, NSF 2119643, CCF-2211557 IIS-1704532, IIS-1741317, NSF IIS-1956151, IIS-1763325, IIS-1909323, IIS-2106758, Amazon Research Awards, Cisco research grant, Picsart Gifts, and Snapchat Gifts. We also thank VLDB Test of Time Award Committee for selecting PathSim for the award.

REFERENCES

[1] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *NIPS*.

[2] Yuwei Cao, Hao Peng, Jia Wu, Yingtong Dou, Jianxin Li, and Philip S. Yu. 2021. Knowledge-Preserving Incremental Social Event Detection via Heterogeneous GNNs. In *WWW*.

[3] Yukuo Cen, Xu Zou, Jianwei Zhang, Hongxia Yang, Jingren Zhou, and Jie Tang. 2019. Representation Learning for Attributed Multiplex Heterogeneous Network. In *KDD*.

[4] Ting Chen and Yizhou Sun. 2017. Task-Guided and Path-Augmented Heterogeneous Network Embedding for Author Identification. In *Proc. of 2017 ACM Int. Conf. on Web Search and Data Mining (WSDM'17)*.

[5] Kewei Cheng, Jiahao Liu, Wei Wang, and Yizhou Sun. 2022. RLogic: Recursive Logical Rule Learning from Knowledge Graphs. In *Proc. of 2022 ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD'22)*.

[6] Kewei Cheng, Ziqing Yang, Ming Zhang, and Yizhou Sun. 2021. UniKER: A Unified Framework for Combining Embedding and Definite Horn Rule Reasoning for Knowledge Graph Inference. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP'21)*.

[7] Yuxiao Dong, Nitesh V Chawla, and Ananthram Swami. 2017. metapath2vec: Scalable representation learning for heterogeneous networks. In *KDD*.

[8] Yuxiao Dong, Ziniu Hu, Kuansan Wang, Yizhou Sun, and Jie Tang. 2020. Heterogeneous Network Representation Learning. In *Proc. of 2020 Int. Joint Conf. on Artificial Intelligence and Pacific Rim Int. Conf. on Artificial Intelligence (IJCAI-PRICAI'20) (Survey Paper)*.

[9] Ahmed El-Kishky, Thomas Markovich, Serim Park, Chetan Verma, Baekjin Kim, Ramy Eskander, Yury Malkov, Frank Portman, Sofía Samaniego, Ying Xiao, and Aria Haghighi. 2022. TwHIN: Embedding the Twitter Heterogeneous Information Network for Personalized Recommendation. https://arxiv.org/abs/2202.05387

[10] Yujie Fan, Shifu Hou, Yiming Zhang, Yanfang Ye, and Melih Abdulhayoglu. 2018. Gotcha - Sly Malware! Scorpion A Metagraph2vec Based Malware Detection System. In *KDD'18*.

[11] Gang Fu, Ying Ding, Abhik Seal, Bin Chen, Yizhou Sun, and Evan Bolton. 2016. Predicting drug target interactions using meta-path-based semantic network analysis. *BMC Bioinformatics 17:160* (2016).

[12] Tao-yang Fu, Wang-Chien Lee, and Zhen Lei. 2017. HIN2Vec: Explore Meta-paths in Heterogeneous Information Networks for Representation Learning. In *CIKM*.

[13] Yali Gao, Xiaoyong Li, Hao Peng, Binxing Fang, and Philip S. Yu. 2022. HinCTI: A Cyber Threat Intelligence Modeling and Identification System Based on Heterogeneous Information Network. *IEEE Transactions on Knowledge and Data Engineering* 34, 2 (2022), 708–722.

[14] Aditya Grover and Jure Leskovec. 2016. node2vec: Scalable feature learning for networks. In *KDD*.

[15] Huan Gui, Yizhou Sun, Jiawei Han, and George Brova. 2014. Modeling Topic Diffusion in Multi-Relational Bibliographic Information Networks. In *Proc. of 2014 ACM Int. Conf. on Information and Knowledge Management (CIKM'14)*.

[16] Junheng Hao, Muhao Chen, Wenchao Yu, Yizhou Sun, and Wei Wang. 2019. Universal Representation Learning of Knowledge Bases by Jointly Embedding Instances and Ontological Concepts. In *Proc. of 25th Int. ACM SIGKDD Conf. on Knowledge Discovery and Data Mining (KDD'19)*.

[17] Anahita Hosseini, Ting Chen, Wenjun Wu, Yizhou Sun, and Majid Sarrafzadeh. 2018. HeteroMed: Heterogeneous Information Network for Medical Diagnosis. In *Proc. of 2018 ACM Int. Conf. on Information and Knowledge Management (CIKM'18)*.

[18] Weihua Hu, Matthias Fey, Marinka Zitnik, Yuxiao Dong, Hongyu Ren, Bowen Liu, Michele Catasta, and Jure Leskovec. 2020. Open Graph Benchmark: Datasets for Machine Learning on Graphs.

[19] Ziniu Hu, Yuxiao Dong, Kuansan Wang, and Yizhou Sun. 2020. Heterogeneous Graph Transformer. In *Proc. 2020 Int. World Wide Web Conf. (WWW'20)*.

[20] Zhipeng Huang, Yudian Zheng, Reynold Cheng, Yizhou Sun, Nikos Mamoulis, and Xiang Li. 2016. Meta Structure: Computing Relevance in Large Heterogeneous Information Networks. In *Proc. of 2016 ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD'16)*.

[21] Roshni G. Iyer, Yusheng Bai, Wei Wang, and Yizhou Sun. 2022. Dual-Geometric Space Embedding Model for Two-View Knowledge Graphs. In *Proc. of 2022 ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD'22)*.

[22] Roshni G. Iyer, Wei Wang, and Yizhou Sun. 2021. Bi-Level Attention Graph Neural Networks. In *2021 IEEE International Conference on Data Mining (ICDM'21)*.

[23] Glen Jeh and Jennifer Widom. 2002. SimRank: a measure of structural-context similarity. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 538–543.

[24] Song Jiang, Bernard Koch, and Yizhou Sun. 2021. HINTS: Citation Time Series Prediction for New Publications via Dynamic Heterogeneous Information Network Embedding. In *Proc. The Web Conf. 2021 (WWW'21)*.

[25] Thomas N Kipf and Max Welling. 2017. Semi-supervised classification with graph convolutional networks. In *ICLR*.

[26] Jon M. Kleinberg. 1999. Authoritative Sources in a Hyperlinked Environment. *JOURNAL OF THE ACM* 46, 5 (1999), 604–632.

[27] Chen Li, Hao Peng, Jianxin Li, Lichao Sun, Lingjuan Lyu, Lihong Wang, Philip S. Yu, and Lifang He. 2022. Joint Stance and Rumor Detection in Hierarchical Heterogeneous Graph. *IEEE Trans. Neural Networks Learn. Syst.* 33(6) (2022).

[28] Zeyu Li, Jyun-Yu Jiang, Yizhou Sun, and Wei Wang. 2019. Personalized Question Routing via Heterogeneous Network Embedding. In *Proc. of The 33rd AAAI Conference on Artificial Intelligence (AAAI '19)*.

[29] Jiawei Liu, Chuan Shi, Cheng Yang, Zhiyuan Lu, and Philip S. Yu. 2022. A survey on heterogeneous information network based recommender systems: Concepts, methods, applications and resources. *AI Open* 3 (2022), 40–57.

[30] Xiaozhong Liu, Yingying Yu, Chun Guo, and Yizhou Sun. 2014. Meta-Path-Based Ranking with Pseudo Relevance Feedback on Heterogeneous Graph for Citation Recommendation. In *Proc. of 2014 ACM Int. Conf. on Information and Knowledge Management (CIKM'14)*.

[31] Zongtao Liu, Bin Ma, Quan Liu, Jian Xu, and Bo Zheng. 2021. Heterogeneous Graph Neural Networks for Large-Scale Bid Keyword Matching. In *CIKM*.

[32] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *NIPS*.

[33] Xichuan Niu, Bofang Li, Chenliang Li, Rong Xiao, Haochuan Sun, Hongbo Deng, and Zhenzhong Chen. 2020. A Dual Heterogeneous Graph Attention Network to Improve Long-Tail Performance for Shop Search in E-Commerce. In *Proc. of 2020 ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD'20)*.

[34] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. *The PageRank Citation Ranking: Bringing Order to the Web.* Technical Report 1999-66.

[35] Hao Peng, Jianxin Li, Yangqiu Song, Renyu Yang, Rajiv Ranjan, Philip S. Yu, and Lifang He. 2021. Streaming Social Event Detection and Evolution Discovery in Heterogeneous Information Networks. *ACM Trans. Knowl. Discov. Data* 15(5) (2021).

[36] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. Deepwalk: Online learning of social representations. In *KDD*.

[37] Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne Van Den Berg, Ivan Titov, and Max Welling. 2018. Modeling relational data with graph convolutional networks. In *ESWC*.

[38] Chuan Shi, Xiaotian Han, Li Song, Xiao Wang, Senzhang Wang, Junping Du, and Philip S. Yu. 2021. Deep Collaborative Filtering with Multi-Aspect Information in Heterogeneous Networks. *IEEE Trans. Knowl. Data Eng.* 33(4) (2021).

[39] Chuan Shi, Binbin Hu, Wayne Xin Zhao, and Philip S. Yu. 2018. Heterogeneous information network embedding for recommendation. *IEEE Transactions on Knowledge and Data Engineering* 31, 2 (2018), 357–370.

[40] Chuan Shi, Yitong Li, Jiawei Zhang, Yizhou Sun, and Philip S. Yu. 2017. A Survey of Heterogeneous Information Network Analysis. *IEEE Trans. Knowl. Data Eng. 29(1): 17-37* (2017).

[41] Chuan Shi, Xiao Wang, and Philip S. Yu. 2022. *Heterogeneous Graph Representation Learning and Applications*. Springer.

[42] Yizhou Sun, Rick Barber, Manish Gupta, Charu C. Aggarwal, and Jiawei Han. 2011. Co-Author Relationship Prediction in Heterogeneous Bibliographic Networks. In *Proc. of 2011 Int. Conf. on Advances in Social Network Analysis and Mining (ASONAM'11)*.

[43] Yizhou Sun and Jiawei Han. 2012. Mining Heterogeneous Information Networks: A Structural Analysis Approach. *SIGKDD Explorations* (2012).

[44] Yizhou Sun, Jiawei Han, Charu C. Aggarwal, and Nitesh V. Chawla. 2012. When Will It Happen? — Relationship Prediction in Heterogeneous Information Networks. In *Proc. 2012 ACM Int. Conf. on Web Search and Data Mining (WSDM'12)*.

[45] Yizhou Sun, Jiawei Han, Xifeng Yan, Philip S. Yu, and Tianyi Wu. 2011. PathSim: Meta Path-Based Top-K Similarity Search in Heterogeneous Information Networks. In *PVLDB 4(11): 992-1003*.

[46] Yizhou Sun, Jiawei Han, Peixiang Zhao, Zhijun Yin, Hong Cheng, and Tianyi Wu. 2009. RankClus: Integrating Clustering with Ranking for Heterogeneous Information Network Analysis. In *Proc. 2009 Int. Conf. Extending Data Base Technology (EDBT'09)*.

[47] Yizhou Sun, Brandon Norick, Jiawei Han, Xifeng Yan, Philip S. Yu, and Xiao Yu. 2013. PathSelClus: Integrating Meta-Path Selection with User-Guided Object Clustering in Heterogeneous Information Networks. *ACM Transactions on Knowledge Discovery from Data (TKDD)* (2013).

[48] Yizhou Sun, Yintao Yu, and Jiawei Han. 2009. Ranking-Based Clustering of Heterogeneous Information Networks with Star Network Schema. In *Proc. 2009 ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD'09)*.

[49] Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. 2019. Rotate: Knowledge graph embedding by relational rotation in complex space. (2019).

[50] Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. 2015. Line: Large-scale information network embedding. In *WWW*.

[51] Chenguang Wang, Yizhou Sun, Yanglei Song, Jiawei Han, Yangqiu Song, Danli Wang, and Ming Zhang. 2016. RelSim: Relation Similarity Search in Schema-Rich Heterogeneous Information Networks. In *Proc of 2016 SIAM Int. Conf. on Data Mining (SDM'16)*.

[52] Shen Wang, Xiaokai Wei, Cícero Nogueira dos Santos, Zhiguo Wang, Ramesh Nallapati, Andrew O. Arnold, Bing Xiang, Philip S. Yu, and Isabel F. Cruz. 2021. Mixed-Curvature Multi-Relational Graph Neural Network for Knowledge Graph Completion. In *WWW*.

[53] Xiao Wang, Houye Ji, Chuan Shi, Bai Wang, Yanfang Ye, Peng Cui, and Philip S Yu. 2019. Heterogeneous Graph Attention Network. In *WWW*.

[54] Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. 2015. Embedding entities and relations for learning and inference in knowledge bases. In *ICLR*.

[55] Carl Yang, Yuxin Xiao, Yu Zhang, Yizhou Sun, and Jiawei Han. 2020. Heterogeneous Network Representation Learning: A Unified Framework with Survey and Benchmark. *IEEE Transactions on Knowledge and Data Engineering* (2020).

[56] Liangwei Yang, Zhiwei Liu, Yu Wang, Chen Wang, Ziwei Fan, and Philip S. Yu. 2022. Large-scale Personalized Video Game Recommendation via Social-aware Contextualized Graph Neural Network. In *WWW*.

[57] Yanfang Ye, Shifu Hou, and Yangqiu Song. 2017. HinDroid: An Intelligent Android Malware Detection System Based on Structured Heterogeneous Information Network. In *Proc. of 2017 ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD'17)*.

[58] Xiao Yu, Xiang Ren, Yizhou Sun, Quanquan Gu, Bradley Sturt, Urvashi Khandelwal, Brandon Norick, and Jiawei Han. 2014. Personalized Entity Recommendation: A Heterogeneous Information Network Approach. In *Proc. 2014 ACM Int. Conf. on Web Search and Data Mining (WSDM'14)*.

[59] Chuxu Zhang, Dongjin Song, Chao Huang, Ananthram Swami, and Nitesh V. Chawla. 2019. Heterogeneous Graph Neural Network. In *KDD*.

[60] Jiawei Zhang and Philip S. Yu. 2019. *Broad Learning Through Fusions - An Application on Social Networks.* Springer.

[61] Jiawei Zhang, Philip S. Yu, and Zhi-Hua Zhou. 2014. Meta-Path Based Multi-Network Collective Link Prediction. In *Proc. of 2014 ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD'14)*.

[62] Huan Zhao, Quanming Yao, Jianda Li, Yangqiu Song, and Dik Lun Lee. 2017. Meta-Graph Based Recommendation Fusion over Heterogeneous Information Networks. In *Proc. of 2017 ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD'17)*.

[63] Liyuan Zheng, Zhen Zuo, Wenbo Wang, Chaosheng Dong, Michinari Momma, and Yi Sun. 2021. Heterogeneous graph neural networks with neighbor-SIM attention mechanism for substitute product recommendation. In *DLG-AAAI 2021*.