

Chapter 2

Survey of Current Developments

Abstract Heterogeneous information network (HIN) provides a new paradigm to manage networked data. Meanwhile, it also introduces new challenges for many data mining tasks. Here, we give a brief survey on recent developments of this field. Concretely, we have analyzed more than 100 referred papers published in the referred conferences and journals in recent years and divided them into seven categories according to their data mining tasks. In this chapter, we will summarize the developments on these seven main data mining tasks. Moreover, these data mining tasks are coarsely ordered from basic task to advanced task.

2.1 Similarity Search

Similarity measure is to evaluate the similarity of objects. It is the basis of many data mining tasks, such as Web search, clustering, and product recommendation. Similarity measure has been well studied for a long time. These studies can be roughly categorized into two types: feature-based approaches and link-based approaches. The feature-based approaches measure the similarity of objects based on their feature values, such as cosine similarity, Jaccard coefficient, and Euclidean distance. The link-based approaches measure the similarity of objects based on their link structures in a graph, such as Personalized PageRank [33] and SimRank [32].

Recently, many researchers begin to consider similarity measure on heterogeneous information networks. Different from similarity measure on homogeneous networks, similarity measure on HIN not only considers structure similarity of two objects but also takes the meta path connecting these two objects into account. As we know, there are different meta paths connecting two objects, and these meta paths contain different semantic meanings, which may lead to different similarities. And thus the similarity measure on HIN is meta path constraint.

Considering the semantics in meta paths constituted by different-typed objects, Sun et al. [88] first propose the path-based similarity measure PathSim to evaluate the similarity of same-typed objects based on symmetric paths. Following their work, some researchers [23, 24] extend PathSim by incorporating richer information, such as transitive similarity, temporal dynamics, and supportive attributes. A path-based

similarity join method [108] is proposed to return the top k similar pairs of objects based on user-specified join paths. Wang et al. [101] define a meta-path-based relation similarity measure, RelSim, to measure the similarity between relation instances in schema-rich HINs. In addition, Wang et al. [99] model a document as a heterogeneous information network and propose a novel similarity measure called KnowSim to compute the relevance of two documents. In information retrieval community, Lao and Cohen [46, 47] propose a path-constrained random walk (PCRW) model to measure the entity proximity in a labeled directed graph constructed by the rich metadata of scientific literature.

In order to evaluate the relevance of different-typed objects, Shi et al. [72, 74] propose HeteSim to measure the relevance of any object pair under arbitrary meta path. As an adaption of HeteSim, LSH-HeteSim [48] is proposed to mine the drug-target interaction in heterogeneous biological networks where drugs and targets are connected with complicated semantic paths. In order to overcome the shortcomings of HeteSim in high computation and memory demand, Meng et al. [62] propose the AvgSim measure that evaluates the similarity scores through two random walk processes along the given meta path and the reverse meta path, respectively. In addition, some methods [8, 141] combine meta path-based relevance search with user preference.

Although similarity measure based on meta path has shown the effectiveness in capturing the single relationship between source and target objects, such as the co-authorship under the meta path *APA*, it still has some shortcomings in some applications. For example, in bibliographic data, we would like to measure the relation of two authors based on the fact that their papers not only are published in the same conference but also have the same topics (i.e., the *APVPA* and *APTPA* paths). In order to measure the complex relevance between objects, Huang et al. [28] propose the relevance measure based on metastructure, which is a directed acyclic graph and can be considered as a combination of meta paths. Similarly, Fang et al. [20] identify metagraphs as a novel means to characterize the common structures for a desired class of proximity. Moreover, they propose a family of metagraph-based proximity and employ a supervised technique to automatically learn the right form of proximity within its family to suit the desired class.

More works begin to integrate the network structure and other information to measure similarity of objects in HIN. Combining the influence and similarity information, Wang et al. [102] simultaneously measure social influence and object similarity in a heterogeneous network to produce more meaningful similarity scores. Wang et al. [96] propose a model to learn relevance through analyzing the context of heterogeneous networks for online targeting. Yu et al. [116] predict the semantic meaning based on a user's query in the meta-path-based feature space and learn a ranking model to answer the similarity query. Recently, Zhang et al. [133] propose a similarity measure to compute similarity between centers in an x-star network according to the attribute similarities and the connections among centers.

2.2 Clustering

Clustering analysis is the process of partitioning a set of data objects (or observations) into a set of clusters, such that objects in a cluster are similar to one another, yet dissimilar to objects in other clusters. Conventional clustering is based on the features of objects, such as k-means and so on [30]. Recently, clustering based on networked data (e.g., community detection) has been studied a lot. This kind of methods models the data as a homogeneous network and uses the given measure (e.g., normalized cuts [78], and modularity [63]) to divide the network into a series of subgraphs. Many algorithms have been proposed to solve this NP-hard problem, such as spectral method, greedy method [93], and sampling technique [71]. Some researches also simultaneously consider objects' link structure and attribute information to increase the accuracy of clustering [110, 140].

Recently, clustering of heterogeneous networks attracts much attention. Compared with homogeneous networks, heterogeneous networks integrate multityped objects, which generates new challenges for clustering tasks. On the one hand, multiple types of objects coexisting in a network lead to new clustering paradigms. As a consequence, a cluster in HIN may include different types of objects sharing the same topic [82]. For example, in a bibliographic heterogeneous network, a cluster of the database area consists of a set of database authors, conferences, terms, and papers. In this way, clustering in HIN preserves richer information, but it also faces more challenges. On the other hand, abundant information contained in HIN makes it more convenient to integrate additional information or other learning tasks for clustering. In this section, we will review these works according to the types of integrated information or tasks.

The attribute information is widely integrated into clustering analysis on HIN. Aggarwal et al. [1] use the local succinctness property to create balanced communities across a heterogeneous network. Considering the incompleteness of objects' attributes and different types of links in heterogeneous information networks, Sun et al. [85] propose a model-based clustering algorithm to integrate the incomplete attribute information and the network structure information. Qi et al. [67] propose a clustering algorithm based on heterogeneous random fields to model the structure and content of social media networks with outlier links. Cruz et al. [14] integrate structural dimension and compositional dimension which composes an attributed graph to solve the community detection problem. Recently, a density-based clustering model TCSC [7] is proposed to detect clusters considering the connections in the network and the vertex attributes.

Text information plays an important role in many heterogeneous network studies. Deng et al. [17] introduce a topic model with biased propagation to incorporate heterogeneous information network with topic modeling in a unified way. Furthermore, they [16] propose a joint probabilistic topic model for simultaneously modeling the contents of multityped objects of a heterogeneous information network. LSA-PTM [103] is introduced to identify clusters of multityped objects by propagating the topics obtained by LSA on the HIN via the links between different objects. Incorporating

both the document content and various links in the text related heterogeneous network, Wang et al. [104] propose a unified topic model for topic mining and multiple objects clustering. Recently, CHINC [98] uses general-purpose knowledge as indirect supervision to improve the clustering results.

User guide information is also integrated into the clustering analysis. Sun et al. [87] present a semi-supervised clustering algorithm to generate different clustering results with path selection according to user guidance. Luo et al. [58] firstly introduce the concept of relation-path to measure the similarity between same-typed objects and use the labeled information to weight relation-paths and then propose SemiRPClus for semi-supervised learning in HIN.

Clustering is usually an independent data mining task. However, it can be integrated with other mining tasks to improve performances through mutual enhancing. Recently, ranking-based clustering on heterogeneous information network has emerged, which shows its advantages on the mutual promotion of clustering and ranking. RankClus [83] generates clusters for a specified type of objects in a bipartite network based on the idea that the qualities of clustering and ranking are mutually enhanced. The following work NetClus [82] is proposed to handle a network with the star-schema. Wang et al. [105] introduce ComClus to promote clustering and ranking performance by applying star-schema network with self-loop to combine the heterogeneous and homogeneous information. In addition, a general method HeProj is proposed to do ranking-based clustering in heterogeneous networks with arbitrary schema by projecting the network into a sequence of subnetworks [75]. And Chen et al. [12] propose a probabilistic generative model to achieve clustering and ranking simultaneously on a heterogeneous network with arbitrary schema. To make use of both textual information and heterogeneous linked entities, Wang et al. [97] develop a clustering and ranking algorithm to construct multityped topical hierarchies automatically. What's more, Qiu et al. [68] propose an algorithm OcdRank to combine overlapping community detection and community-member ranking together in directed heterogeneous social networks.

Outlier detection is the process of finding data objects with behaviors that are very different from expectation. Outlier detection and clustering analysis are two highly related, but different-aimed tasks. To detect outliers, Gupta et al. [21] propose an outlier-aware approach based on joint nonnegative matrix factorization to discover popular community distribution patterns. Furthermore, they propose to detect association-based clique outliers in heterogeneous networks given a conjunctive select query [22]. What's more, Zhuang et al. [142] propose an outlier detection algorithm to find subnetwork outliers according to different queries and semantics. Also based on queries, Kuck et al. [44] propose a meta-path-based outlieriness measure for mining outliers in heterogeneous networks.

In addition, some other information is also integrated. For example, a social influence-based clustering framework SI-Cluster is proposed to analyze heterogeneous information networks based on both people's connections and their social activities [138]. Besides the traditional models employed in clustering on HIN, such as topic model and spectral clustering, Alqadah et al. [4] propose a novel game theoretic framework for defining and mining clusters in heterogeneous information networks.

2.3 Classification

Classification is a data analysis task where a model or classifier is constructed to predict class (categorical) labels. Traditional machine learning has focused on the classification of identically structured objects satisfying independent identically distribution (IID). However, links exist among objects in many real-world datasets, which makes objects not satisfy IID. So link-based object classification has received considerable attention, where a data graph is composed of a set of objects connected to each other via a set of links. Many methods extend traditional classification methods to consider correlations among objects [45]. The link-based object classification usually considers that objects and links in the graph are identical, respectively. That is, the objects and links among them constitute a homogeneous network.

Different from traditional classification researches, the classification problems studied in HIN have some new characteristics. First, the objects contained in HIN are different-typed, which means we can classify multiple types of objects simultaneously. Second, label knowledge can spread through various links among different-typed objects. In the HIN condition, the label of objects is decided by the effects of different-typed objects along different-typed links.

Many works extend traditional classification to heterogeneous information networks. Some works extend transductive classification task, which is to predict labels for the given unlabeled data. For example, GNetMine [35] is proposed to model the link structure in information networks with arbitrary network schema and arbitrary number of object/link types. Wan et al. [94] propose a graph-regularized meta-path-based transductive regression model, which combines the principal philosophies of typical graph-based transductive classification methods and transductive regression models designed for homogeneous networks. Luo et al. propose HetPathMine [56] to cluster with small labeled data on HIN through a novel meta path selection model, and Jacob et al. [29] propose a method to label nodes of different types by computing a latent representation of nodes in a space where two connected nodes tend to have close latent representations. Recently, Bangcharoensap et al. [6] employ the edge betweenness centrality for the edge weight normalization and further improve the centrality to make it suitable for heterogeneous networks. Some works also extend inductive classification that is to construct a decision function in the whole data space. For example, Rossi et al. [70] use a bipartite heterogeneous network to represent textual document collections and propose IMBHN algorithm to induce a classification model assigning weights to textual terms.

Multilabel classification is prevalent in many real-world applications, where each example can be associated with a set of multiple labels simultaneously [41]. This kind of classification tasks is also extended to HIN. Angelova et al. [5] introduce a multilabel graph-based classification model for labeling heterogeneous networks by modeling the mutual influence between nodes as a random walk process. Kong et al. [41] use multiple types of relationships mined from the linkage structure of HIN to facilitate the multilabel classification process. Zhou et al. [139] propose an

edge-centric multilabel classification approach considering both the structure affinity and the label vicinity.

As a unique characteristic, meta path is widely used in classification on HIN. Meta paths are usually used for feature generation in many methods, such as GNetMine [35] and HetPathMine [56]. Moreover, Kong et al. [40] introduce the concept of meta-path-based dependencies among objects to study the collective classification problem. Recently, Wang et al. [100] develop kernel methods based on meta paths in the HIN representation of texts for text classification.

Similar to clustering problem, classification is also integrated with other data mining tasks on HIN. Ranking-based classification is to integrate classification and ranking in a simultaneous, mutually enhancing process. Ji et al. [36] propose a ranking-based classification framework, RankClass, to perform more accurate analysis. As an extension of RankClass, Chen et al. [13] propose the F-RankClass for a unified classification framework that can be applied to binary or multiclass classification of unimodal or multimodal data. Some methods also integrate classification with information propagation. For example, Jendoubi et al. [34] classify the social message based on its spreading in the network and the theory of belief functions.

2.4 Ranking

Ranking is an important data mining task in network analysis, which evaluates object importance or popularity based on some ranking functions. Many ranking methods have been proposed in homogeneous networks, such as PageRank [65] and HITS [39]. These approaches only consider the same type of objects in homogeneous networks.

Ranking in heterogeneous information networks is an important and meaningful task, but faces several challenges. First, there are different types of objects and relations in HIN, and treating all objects equally will mix different types of objects together. Second, different types of objects and relations in HIN carry different semantic meanings, which may lead to different ranking results. Taking the bibliographic heterogeneous network as an example, ranking on authors may have different results under different meta paths [50, 77], since these meta paths will construct different link structures among authors. Moreover, the rankings of different-typed objects have mutual effects. For example, reputable authors usually publish papers on top conferences.

The co-ranking problem on bipartite graphs has been widely explored in the past decades. For example, Zhou et al. [137] co-rank authors and their publications by coupling two random walk processes, and co-HITS [15] incorporates the bipartite graph with the content information and the constraints of relevance. Soulier et al. [80] propose a bipartite entity ranking algorithm to rank jointly documents and authors in a bibliographic network regarding a topical query by combining content-based and network-based features. There are also some ranking works on the multirelational network. For example, MultiRank [64] is proposed to determine the importance of

both objects and relations simultaneously for multirelational data, and HAR [49] is proposed to determine hub and authority scores of objects and relevance scores of relations in multirelational data for query search. These two methods focus on the same type of objects with multirelations. Recently, Huang et al. [26] integrate both formal genre and inferred social networks with tweet networks to rank tweets. Although this work makes use of various types of objects in heterogeneous networks, it still ranks one type of objects.

Considering the characteristics of meta path on HIN, some works propose path-based ranking methods. For example, Liu et al. [54] develop a publication ranking method with pseudorelevance feedback by leveraging a number of meta paths on the heterogeneous bibliographic graph. Applying the tensor analysis, Li et al. [50] propose HRank to evaluate the importance of multiple types of objects and meta paths simultaneously.

Ranking problem is also extended to HIN constructed by social media network. For image search in social media, Tsai et al. [91] propose SocialRank which uses social hints for image search and ranking in social networks. To identify high-quality objects (questions, answers, and users) in Q&A systems, Zhang et al. [129] devise an unsupervised heterogeneous network-based framework to co-rank multiple objects in Q&A sites. For heterogeneous cross-domain ranking problem, Wang et al. [95] propose a general regularized framework to discover a latent space for two domains and minimize two weighted ranking functions simultaneously in the latent space. Considering the dynamic nature of literature networks, a mutual reinforcement ranking framework is proposed to rank the future popularity of new publications and young researchers simultaneously [106].

2.5 Link Prediction

Link prediction is a fundamental problem in link mining that attempts to estimate the likelihood of the existence of a link between two nodes, based on observed links and the attributes of nodes. Link prediction is often viewed as a simple binary classification problem: For any two potentially linked objects, predict whether the link exists (1) or not (0). One kind of approach is to make this prediction entirely based on structural properties of the network [51], and another kind of approach is to make use of attribute information for link prediction [66].

Link prediction in an HIN has been an important research topic for recent years, which has the following characteristics. First, the links to be predicted are of different types, since objects in HIN are connected with different types of links. Second, there are dependencies existing among multiple types of links. So link prediction in an HIN needs to predict multiple types of links collectively by capturing the diverse and complex relationships among different types of links and leveraging the complementary prediction information.

Utilizing the meta path, many works employ a two-step process to solve the link prediction problem in HIN. The first step is to extract meta path-based feature vectors,

while the second step is to train a regression or classification model to compute the existence probability of a link [10, 11, 84, 86, 115]. For example, Sun et al. [84] propose PathPredict to solve the problem of co-author relationship prediction through meta path-based feature extraction and logistic regression-based model. Zhang et al. [130] use meta path-based features to predict organization chart or management hierarchy. Utilizing diverse and complex linkage information, Cao et al. [10] design a relatedness measure to construct the feature vectors of links and propose an iterative framework to predict multiple types of links collectively. In addition, Sun et al. [86] model the distribution of relationship building time with the use of the extracted topological features to predict when a certain relationship will be formed.

Probabilistic models are also widely applied for link prediction tasks in HIN. Yang et al. [112] propose a probabilistic method MRIP which models the influence propagating between heterogeneous relationships to predict links in multirelational heterogeneous networks. Also, the TFGM model [113] defines a latent topic layer to bridge multiple networks and designs a semi-supervised learning model to mine competitive relationships across heterogeneous networks. Dong et al. [19] develop a transfer-based ranking factor graph model that combines several social patterns with network structure information for link prediction and recommendation. Matrix factorization is another common tool to handle link prediction problems. For example, Huang et al. [27] develop the joint manifold factorization (JMF) method to perform trust prediction with the ancillary rating matrix via aggregating heterogeneous social networks.

The approaches mentioned above mainly focus on link prediction on one single heterogeneous network. Recently, Zhang et al. [42, 126, 128] propose the problem of link prediction across multiple aligned heterogeneous networks. A two-phase link prediction method is put forward in [42]. The first phase is to extract heterogeneous features from multiple networks, while the second phase is to infer anchor links by formulating it as a stable matching problem. In addition, Zhang et al. [126] propose SCAN-PS to solve the social link prediction problem for new users using the “anchors.” Furthermore, they propose the TRAIL [128] method to predict social links and location links simultaneously. Also aimed at the cold-start problem of new users, Liu et al. [52] propose the aligned factor graph model for user–user link prediction problem by utilizing information from another similar social network. In order to identify users from multiple heterogeneous social networks and integrate different networks, an energy-based model COSNET [134] is proposed by considering both local and global consistency among multiple networks.

Most of the available works on link prediction are designed for static networks; however, the problem of dynamic link prediction is also very important and challenging. Taking into account both the dynamic and heterogeneous nature of Web data, Zhao et al. [135] propose a general framework to characterize and predict community members from the evolution of heterogeneous Web data. In order to solve the problem of dynamic link inference in temporal and heterogeneous information networks, Aggarwal et al. [2, 3] develop a two-level scheme which makes efficient macro- and microdecisions for combining the topology and type information. Aiming

at predicting the distribution of the labels on neighbors of a given node, Ma et al. [60] propose an evolution factor model which utilizes two new structures, neighbor distribution vector and neighbor label evolution matrix.

2.6 Recommendation

Recommender system helps consumers to search products that are likely to be of interest to the user such as books, movies, and restaurants. It uses a broad range of techniques from information retrieval, statistics, and machine learning to search for similarities among items and customer preferences. Traditional recommended systems normally only utilize the user–item rating feedback information for recommendation. Collaborative filtering is one of the most popular techniques, which includes two types of approaches: memory-based methods and model-based methods. Recently, matrix factorization has shown its effectiveness and efficiency in recommended systems, which factorizes the user–item rating matrix into two low-rank user-specific and item-specific matrices and then utilizes the factorized matrices to make further predictions [81]. With the prevalence of social media, more and more researchers study social recommended system, which utilizes social relations among users [59, 111].

Recently, some researchers have begun to realize the importance of heterogeneous information for recommendations. The comprehensive information and rich semantics of HIN make it promising to generate better recommendations. For example, in an HIN extracted from movie-recommended system [76], it not only contains different types of objects (e.g., users and movies) but also illustrates all kinds of relations among objects, such as viewing information, social relations, and attribute information. Constructing heterogeneous networks for recommendation can effectively fuse all kinds of information, which can be potentially utilized for recommendation. Moreover, the objects and relations in the networks have different semantics, which can be explored to reveal subtle relations among objects.

Meta path is well used to explore the semantics and extract relations among objects. Shi et al. [73] implement a semantic-based recommended system, HeteRecom, which employs the semantics information of meta path to evaluate the similarities between movies. Furthermore, considering the attribute values, such as rating score on links, they model the recommended system as a weighted HIN and propose a semantic path-based personalized recommendation method SemRec [76]. In order to take full advantage of the relationship heterogeneity, Yu et al. [117, 118] introduce meta-path-based latent features to represent the connectivity between users and items along different types of paths and then define recommendation models at both global and personalized levels with Bayesian ranking optimization techniques. Also based on meta path, Burke et al. [9] present an approach for recommendation which incorporates multiple relations in a weighted hybrid.

A number of approaches employ heterogeneous information network to fuse various kinds of information. Utilizing different contexts information, Jamali et al. [31]

propose a context-dependent matrix factorization model which considers a general latent factor for every entity and context-dependent latent factors for every context. Using user implicit feedback data, Yu et al. [117, 118] solve the global and personalized entity recommendation problem. Based on related interest groups, Ren et al. [69] propose a cluster-based citation recommendation framework to predict each query's citations in bibliographic networks. Similarly, Wu et al. [107] exploit graph summarization and content-based clustering for media recommendation with the interest group information. Based on multiple heterogeneous network features, Yang et al. [109] model multiple features into a unified framework with a SVM-Rank-based method. And using multiple types of relations, Luo et al. [57] propose a social collaborative filtering algorithm. In addition, adopting the similarity of users and items as regularization, some works [75, 136] propose matrix factorization-based frameworks for recommendation.

2.7 Information Fusion

Information fusion denotes the process of merging information from heterogeneous sources with differing conceptual, contextual, and typographical representations. Due to the availability of various data sources, fusing these scattered distributed information sources has become an important research problem. In the past decades, dozens of papers have been published on this topic in many traditional data mining areas, e.g., data schema integration in data warehouse [61], protein-protein interaction (PPI) networks and gene regulatory networks matching in bioinformatics [79], and ontology mapping in Web semantics [18]. Nowadays, with the surge of HIN, information fusion across multiple HINs has become a novel yet important research problem. By fusing information from different HINs, we can obtain a more comprehensive and consistent knowledge about the common information entities shared in different HINs, including their structures, properties, and activities.

To fuse the information in multiple HINs, an important prerequisite will be to align the HINs via the shared common information entities, which can be users in social networks, authors in bibliographical networks, and protein molecules in biological networks. Perfect HIN alignment is a challenging problem as the underlying subgraph isomorphism problem is actually NP-complete [38]. Meanwhile, based on the structure and attribute information available in HINs, a large number of approximated HIN alignment algorithms have been proposed so far. Enlightened by the homogeneous network alignment method in [92], Koutra et al. [43] propose to align two bipartite graphs with a fast network alignment algorithm. Zafarani et al. [138] propose to match users across social networks based on various node attributes, e.g., username, typing patterns, and language patterns. Kong et al. [42] formulate the heterogeneous social network alignment problem as an anchor link prediction problem. A two-step supervised method MNA is proposed in [42] to infer potential anchor links across networks with heterogeneous information in the networks. However, social networks in the real world are actually mostly partially aligned, and lots of

users are not anchor users. Zhang et al. have proposed the partial network alignment methods based on supervised learning setting and PU learning setting in [123, 131], respectively. In addition to these pairwise social network alignment problems, multiple (more than two) social networks can be aligned simultaneously. Zhang et al. [124] discover that the inferred cross-network mapping of entities in social network alignment should meet the transitivity law and has an inherent one-to-one constraint. A new multiple social alignment framework is introduced in [124] to minimize the alignment costs and preserve the transitivity law and one-to-one constraint on the inferred mappings. Besides users, many other kinds of information entities can also be shared by multiple social sites, such as the geo-spatial locations shared by location-based social networks and products shared by e-commerce sites. To infer the corresponding mapping between these different kinds of information entities simultaneously, Zhang et al. propose the network partial co-alignment problem in [125].

By fusing multiple HINs, the heterogeneous information available in each network can be transferred to other aligned networks, and lots of application problems on HIN, e.g., link prediction and friend recommendation [90, 123, 127], community detection [122], information diffusion [119, 120, 132], and product recommendation [55], will benefit from it a lot.

Via the inferred mappings, Zhang et al. propose to transfer heterogeneous links across aligned networks to improve quality of predicted links/recommended friends [123, 127]. Tang et al. [90] propose a transfer-based factor graph model which predicts the types of social relationships in a target network by borrowing knowledge from a different source network. For new networks [128] and new users [126] with little social activity information, the transferred information can greatly overcome the cold-start problem when predicting links for them. What's more, information about the shared entities across aligned networks can provide us with a more comprehensive knowledge about the community structures formed by them. By utilizing the information across multiple aligned networks, Zhang et al. [114] propose a new model to refine the clustering results of the shared entities with information in other aligned networks mutually. Jin et al. [37] propose a scalable framework to study the synergistic partitioning of multiple aligned large-scale networks, which takes the relationships among different networks into consideration and tries to maintain the consistency on partitioning the same nodes of different networks into the same partitions. Zhang et al. [122] study the community detection in emerging networks with information transferred from other aligned networks to overcome the cold-start problem. In addition, by fusing multiple heterogeneous social networks, users in networks will be extensively connected with each other via both intra-network connections (e.g., friendship connections among users) and inter-network connections (i.e., the inferred mappings across networks). As a result, information can reach more users and achieve broader influence across the aligned social networks. Zhan et al. propose a new model to study the information diffusion process across multiple aligned networks in [119] and introduce a new problem to discover the tipping users across aligned networks in [120].

2.8 Other Applications

Besides the tasks discussed above, there are many other applications in heterogeneous networks, such as influence propagation and privacy risk problem. To quantitatively learn influence from heterogeneous networks, Liu et al. [53] first use a generative graphical model to learn the direct influence and then use propagation methods to mine indirect and global influence. Using meta paths, Zhan et al. [119] propose a model M&M to solve the influence maximization problem in multiple partially aligned heterogeneous online social networks. For privacy risk in anonymized HIN, Zhang et al. [121] present a de-anonymization attack that exploits the identified vulnerability to prey upon the risk. Aiming at the inferior performances of unsupervised text embedding methods, Tang et al. [89] propose a semi-supervised representation learning method for text data, in which labeled information and different levels of word co-occurrence information are represented as a large-scale heterogeneous text network. To improve the effectiveness of offline sales, Hu et al. [25] construct a company-to-company graph from semantics based meta-path learning and then adopt label propagation on the graph to predict promising companies.

References

1. Aggarwal, C., Xie, Y., Yu, P.: Towards community detection in locally heterogeneous networks. In: SDM, pp. 391–402 (2011)
2. Aggarwal, C.C., Xie, Y., Yu, P.S.: On dynamic link inference in heterogeneous networks. In: SDM, pp. 415–426 (2012)
3. Aggarwal, C.C., Xie, Y., Yu, P.S.: A framework for dynamic link prediction in heterogeneous networks. *Stat. Anal. Data Min. ASA Data Sci. J.* 7(1), 14–33 (2014)
4. Alqadah, F., Bhatnagar, R.: A game theoretic framework for heterogeneous information network clustering. In: KDD, pp. 795–802 (2011)
5. Angelova, R., Kasneci, G., Weikum, G.: Graffiti: graph-based classification in heterogeneous networks. In: WWW, pp. 139–170 (2012)
6. Bangcharoensap, P., Murata, T., Kobayashi, H., Shimizu, N.: Transductive classification on heterogeneous information networks with edge betweenness-based normalization. In: WSDM, pp. 437–446 (2016)
7. Boden, B., Ester, M., Seidl, T.: Density-based subspace clustering in heterogeneous networks. In: ECML/PKDD, pp. 149–164 (2014)
8. Bu, S., Hong, X., Peng, Z., Li, Q.: Integrating meta-path selection with user-preference for top-k relevant search in heterogeneous information networks. In: CSCWD, pp. 301–306 (2014)
9. Burke, R., Vahedian, F., Mobasher, B.: Hybrid recommendation in heterogeneous networks. In: UMAP, pp. 49–60 (2014)
10. Cao, B., Kong, X., Yu, P.S.: Collective prediction of multiple types of links in heterogeneous information networks. In: ICDM, pp. 50–59 (2014)
11. Chen, J., Gao, H., Wu, Z., Li, D.: Tag co-occurrence relationship prediction in heterogeneous information networks. In: ICPADS, pp. 528–533 (2013)
12. Chen, J., Dai, W., Sun, Y., Dy, J.: Clustering and ranking in heterogeneous information networks via gamma-poisson model. In: SDM, pp. 425–432 (2015)
13. Chen, S.D., Chen, Y.Y., Han, J., Moulin, P.: A feature-enhanced ranking-based classifier for multimodal data and heterogeneous information networks. In: ICDM, pp. 997–1002 (2013)

14. Cruz, J.D., Bothorel, C., Poulet, F.: Integrating heterogeneous information within a social network for detecting communities. In: *ASONAM*, pp. 1453–1454 (2013)
15. Deng, H., Lyu, M.R., King, I.: A generalized Co-HITS algorithm and its application to bipartite graphs. In: *KDD*, pp. 239–248 (2009)
16. Deng, H., Zhao, B., Han, J.: Collective topic modeling for heterogeneous networks. In: *SIGIR*, pp. 1109–1110 (2011)
17. Deng, H., Han, J., Zhao, B., Yu, Y., Lin, C.X.: Probabilistic topic models with biased propagation on heterogeneous information networks. In: *KDD*, pp. 795–802 (2011)
18. Doan, A., Madhavan, J., Domingos, P., Halevy, A.: Ontology matching: a machine learning approach. *Handbook on Ontologies*, pp. 385–403. Springer, Berlin (2004)
19. Dong, Y., Tang, J., Wu, S., Tian, J., Chawla, N.V., Rao, J., Cao, H.: Link prediction and recommendation across heterogeneous social networks. In: *ICDM*, pp. 181–190 (2012)
20. Fang, Y., Lin, W., Zheng, V.W., Wu, M., Chang, C.C., Li, X.L.: Semantic proximity search on graphs with metagraph-based learning. In: *ICDE*, pp. 277–288 (2016)
21. Gupta, M., Gao, J., Han, J.: Community distribution outlier detection in heterogeneous information networks. In: *ECML*, pp. 557–573 (2013)
22. Gupta, M., Gao, J., Yan, X., Cam, H., Han, J.: On detecting association-based clique outliers in heterogeneous information networks. In: *ASONAM*, pp. 108–115 (2013)
23. He, J., Bailey, J., Zhang, R.: Exploiting transitive similarity and temporal dynamics for similarity search in heterogeneous information networks. In: *International Conference on Database Systems for Advanced Applications*, pp. 141–155 (2014)
24. Hou U.L., Yao, K., Mak, H.: PathSimExt: revisiting PathSim in heterogeneous information networks. In: *WAIM*, pp. 38–42 (2014)
25. Hu, Q., Xie, S., Zhang, J., Zhu, Q., Guo, S., Yu, P.S.: HeteroSales: utilizing heterogeneous social networks to identify the next enterprise customer. In: *WWW*, pp. 41–50 (2016)
26. Huang, H., Zubiaga, A., Ji, H., Deng, H., Wang, D., Le, H.K., Abdelzaher, T.F., Han, J., Leung, A., Hancock, J.P., Others: Tweet ranking based on heterogeneous networks. In: *COLING*, pp. 1239–1256 (2012)
27. Huang, J., Nie, F., Huang, H., Tu, Y.C.: Trust prediction via aggregating heterogeneous social networks. In: *CIKM*, pp. 1774–1778 (2012)
28. Huang, Z., Zheng, Y., Cheng, R., Sun, Y., Mamouli, N., Li, X.: Meta structure: computing relevance in large heterogeneous information networks. In: *SIGKDD*, pp. 1595–1604 (2016)
29. Jacob, Y., Denoyer, L., Gallinari, P.: Learning latent representations of nodes for classifying in heterogeneous social networks. In: *WSDM*, pp. 373–382 (2014)
30. Jain, A.K.: Data clustering: 50 years beyond K-means. *Pattern Recognit. Lett.* **31**(8), 651–666 (2010)
31. Jamali, M., Lakshmanan, L.: HeteroMF: recommendation in heterogeneous information networks using context dependent factor models. In: *WWW*, pp. 643–654 (2013)
32. Jeh, G., Widom, J.: SimRank: a measure of structural-context similarity. In: *KDD*, pp. 538–543 (2002)
33. Jeh, G., Widom, J.: Scaling personalized web search. In: *WWW*, pp. 271–279 (2003)
34. Jendoubi, S., Martin, A., Lietard, L., Yaghlane, B.B.: Classification of message spreading in a heterogeneous social network. In: *IPMU*, pp. 66–75 (2014)
35. Ji, M., Sun, Y., Danilevsky, M., Han, J., Gao, J.: Graph regularized transductive classification on heterogeneous information networks. In: *ECML/PKDD*, pp. 570–586 (2010)
36. Ji, M., Han, J., Danilevsky, M.: Ranking-based classification of heterogeneous information networks. In: *KDD*, pp. 1298–1306 (2011)
37. Jin, S., Zhang, J., Yu, P.S., Yang, S., Li, A.: Synergistic partitioning in multiple large scale social networks. In: *IEEE BigData*, pp. 281–290 (2014)
38. Klau, G.W.: A new graph-based method for pairwise global network alignment. *BMC Bioinform.* **10**(Suppl 1), S59 (2009)
39. Kleinberg, J.M.: Authoritative sources in a hyperlinked environment. In: *SODA*, pp. 668–677 (1999)

40. Kong, X., Yu, P.S., Ding, Y., Wild, D.J.: Meta path-based collective classification in heterogeneous information networks. In: CIKM, pp. 1567–1571 (2012)
41. Kong, X., Cao, B., Yu, P.S.: Multi-label classification by mining label and instance correlations from heterogeneous information networks. In: KDD, pp. 614–622 (2013)
42. Kong, X., Zhang, J., Yu, P.S.: Inferring anchor links across multiple heterogeneous social networks. In: CIKM, pp. 179–188 (2013)
43. Koutra, D., Tong, H., Lubensky, D.: Big-align: fast bipartite graph alignment. In: ICDM, pp. 389–398 (2013)
44. Kuck, J., Zhuang, H., Yan, X., Cam, H., Han, J.: Query-based outlier detection in heterogeneous information networks. In: EDBT, pp. 325–336 (2015)
45. Lafferty, J., McCallum, A., Pereira, F.C.N.: Conditional random fields: probabilistic models for segmenting and labeling sequence data. In: ICML, pp. 282–289 (2001)
46. Lao, N., Cohen, W.: Fast query execution for retrieval models based on path constrained random walks. In: KDD, pp. 881–888 (2010)
47. Lao, N., Cohen, W.W.: Relational retrieval using a combination of path-constrained random walks. *Mach. Learn.* **81**(2), 53–67 (2010)
48. Li, C., Sun, J., Xiong, Y., Zheng, G.: An efficient drug-target interaction mining algorithm in heterogeneous biological networks. In: PAKDD, pp. 65–76 (2014)
49. Li, X., Ng, M.K., Ye, Y.: HAR: hub, authority and relevance scores in multi-relational data for query search. In: SDM, pp. 141–152 (2012)
50. Li, Y., Shi, C., Yu, P.S., Chen, Q.: HRank: a path based ranking method in heterogeneous information network. In: WAIM, pp. 553–565 (2014)
51. Liben-Nowell, D., Kleinberg, J.: The link-prediction problem for social networks. *J. Am. Soc. Inf. Sci. Tech.* **58**(7), 1019–1031 (2007)
52. Liu, F., Xia, S.: Link prediction in aligned heterogeneous networks. In: PAKDD, pp. 33–44 (2015)
53. Liu, L., Tang, J., Han, J., Yang, S.: Learning influence from heterogeneous social networks. *Data Min. Knowl. Discov.* **25**(3), 511–544 (2012)
54. Liu, X., Yu, Y., Guo, C., Sun, Y.: Meta-path-based ranking with pseudo relevance feedback on heterogeneous graph for citation recommendation. In: CIKM, pp. 121–130 (2014)
55. Lu, C.T., Xie, S., Shao, W., He, L., Yu, P.S.: Item recommendation for emerging online businesses. In: IJCAI, pp. 3797–3803 (2016)
56. Luo, C., Guan, R., Wang, Z., Lin, C.: HetPathMine: a novel transductive classification algorithm on heterogeneous information networks. In: *Advances in Information Retrieval*, vol. 8416, pp. 210–221 (2014)
57. Luo, C., Pang, W., Wang, Z.: Hete-CF: social-based collaborative filtering recommendation using heterogeneous relations. In: ICDM, pp. 917–922 (2014)
58. Luo, C., Pang, W., Wang, Z.: Semi-supervised clustering on heterogeneous information networks. In: *Advances in Knowledge Discovery and Data Mining*, vol. 8444, pp. 548–559 (2014)
59. Ma, H., King, I., Lyu, M.R.: Learning to recommend with social trust ensemble. In: SIGIR, pp. 203–210 (2009)
60. Ma, Y., Yang, N., Li, C., Zhang, L., Yu, P.S.: Predicting neighbor distribution in heterogeneous information networks. In: SDM, pp. 784–791 (2015)
61. Melnik, S., Garcia-Molina, H., Rahm, E.: Similarity flooding: a versatile graph matching algorithm and its application to schema matching. In: ICDE, pp. 117–128 (2002)
62. Meng, X., Shi, C., Li, Y., Zhang, L., Wu, B.: Relevance measure in large-scale heterogeneous networks. In: APWeb, pp. 636–643 (2014)
63. Newman, M.E.J., Girvan, M., M.E.J., Newman, M.G.: Finding and evaluating community structure in networks. *Phys. Rev. E* **69**(026113), 1757–1771 (2004)
64. Ng, M.K., Li, X., Ye, Y., Ng, M., Li, X., Ye, Y.: MultiRank: co-ranking for objects and relations in multi-relational data. In: KDD, pp. 1217–1225 (2011)
65. Page, L., Brin, S., Motwani, R., Winograd, T.: The pagerank citation ranking: bringing order to the web. In: Stanford InfoLab, pp. 1–14 (1998)

66. Popescul, A., Ungar, L.H.: Statistical relational learning for link prediction. In: IJCAI Workshop on Learning Statistical Models from Relational Data, vol. 2003 (2003)
67. Qi, G.J., Aggarwal, C.C., Huang, T.S.: On clustering heterogeneous social media objects with outlier links. In: WSDM, pp. 553–562 (2012)
68. Qiu, C., Chen, W., Wang, T., Lei, K.: Overlapping community detection in directed heterogeneous social network. In: WAIM, pp. 490–493 (2015)
69. Ren, X., Liu, J., Yu, X., Khandelwal, U., Gu, Q., Wang, L., Han, J.: ClusCite: effective citation recommendation by information network-based clustering. In: KDD, pp. 821–830 (2014)
70. Rossi, R.G., de Paulo Faleiros, T., de Andrade Lopes, A., Rezende, S.O.: Inductive model generation for text categorization using a bipartite heterogeneous network. In: ICDM, pp. 1086–1091 (2012)
71. Sales-Pardo, M., Guimera, R., Moreira, A.A., Amaral, L.A.N.: Extracting the hierarchical organization of complex systems. *Proc. Natl. Acad. Sci.* **104**(39), 15224–15229 (2007)
72. Shi, C., Kong, X., Yu, P.S., Xie, S., Wu, B.: Relevance search in heterogeneous networks. In: EDBT, pp. 180–191 (2012)
73. Shi, C., Zhou, C., Kong, X., Yu, P.S., Liu, G., Wang, B.: HeteRecom: a semantic-based recommendation system in heterogeneous networks. In: KDD, pp. 1552–1555 (2012)
74. Shi, C., Kong, X., Huang, Y., Philip, S.Y., Wu, B.: Hetesim: a general framework for relevance measure in heterogeneous networks. *IEEE Trans. Knowl. Data Eng.* **26**(10), 2479–2492 (2014)
75. Shi, C., Wang, R., Li, Y., Yu, P.S., Wu, B.: Ranking-based clustering on general heterogeneous information networks by network projection. In: CIKM, pp. 699–708 (2014)
76. Shi, C., Zhang, Z., Luo, P., Yu, P.S., Yue, Y., Wu, B.: Semantic path based personalized recommendation on weighted heterogeneous information networks. In: CIKM, pp. 453–462 (2015)
77. Shi, C., Li, Y., Philip, S.Y., Wu, B.: Constrained-meta-path-based ranking in heterogeneous information network. *Knowl. Inf. Syst.* 1–29 (2016)
78. Shi, J., Malik, J.: Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(8), 888–905 (2000)
79. Shih, Y.K., Parthasarathy, S.: Scalable global alignment for multiple biological networks. *BMC Bioinf.* **13**, 1–13 (2012)
80. Soulier, L., Jabeur, L.B., Tamine, L., Bahoun, W.: On ranking relevant entities in heterogeneous networks using a language-based model. *J. Am. Soc. Inf. Sci. Technol.* **64**(3), 500–515 (2013)
81. Srebro, N., Jaakkola, T.: Weighted low-rank approximations. In: ICML, pp. 720–727 (2003)
82. Sun, Y., Yu, Y., Han, J.: Ranking-based clustering of heterogeneous information networks with star network schema. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 797–806 (2009)
83. Sun, Y., Han, J., Zhao, P., Yin, Z., Cheng, H., Wu, T.: RankClus: integrating clustering with ranking for heterogeneous information network analysis. In: EDBT, pp. 565–576 (2009)
84. Sun, Y., Barber, R., Gupta, M., Aggarwal, C.C., Han, J.: Co-author relationship prediction in heterogeneous bibliographic networks. In: ASONAM, pp. 121–128 (2011)
85. Sun, Y., Aggarwal, C., Han, J.: Relation strength-aware clustering of heterogeneous information networks with incomplete attributes. In: VLDB, pp. 394–405 (2012)
86. Sun, Y., Han, J., Aggarwal, C.C., Chawla, N.V.: When will it happen?: relationship prediction in heterogeneous information networks. In: WSDM, pp. 663–672 (2012)
87. Sun, Y., Norick, B., Han, J., Yan, X., Yu, P.S., Yu, X.: Integrating meta-path selection with user-guided object clustering in heterogeneous information networks. In: KDD, pp. 1348–1356 (2012)
88. Sun, Y.Z., Han, J.W., Yan, X.F., Yu, P.S., Wu, T.: PathSim: meta path-based Top-K similarity search in heterogeneous information networks. In: VLDB, pp. 992–1003 (2011)
89. Tang, J., Qu, M., Mei, Q.: PTE: predictive text embedding through large-scale heterogeneous text networks. In: KDD, pp. 1165–1174 (2015)
90. Tang, J., Lou, T., Kleinberg, J., Wu, S.: Transfer learning to infer social ties across heterogeneous networks. *ACM Trans. Inf. Syst.* **34**(2), 7:1–7:43 (2016)

91. Tsai, M.H., Aggarwal, C., Huang, T.: Ranking in heterogeneous social media. In: WSDM, pp. 613–622 (2014)
92. Umeyama, S.: An eigendecomposition approach to weighted graph matching problems. *IEEE Trans. Pattern Anal. Mach. Intell.* **10**(5), 695–703 (1988)
93. Wakita, K., Tsurumi, T.: Finding community structure in mega-scale social networks. In: WWW, pp. 1275–1276 (2007)
94. Wan, M., Ouyang, Y., Kaplan, L., Han, J.: Graph regularized meta-path based transductive regression in heterogeneous information network. In: SDM, pp. 918–926 (2015)
95. Wang, B., Tang, J., Fan, W., Chen, S., Tan, C., Yang, Z.: Query-dependent cross-domain ranking in heterogeneous network. *Knowl. Inf. Syst.* **34**(1), 109–145 (2013)
96. Wang, C., Raina, R., Fong, D., Zhou, D., Han, J., Badros, G.J.: Learning relevance from heterogeneous social network and its application in online targeting. In: SIGIR, pp. 655–664 (2011)
97. Wang, C., Danilevsky, M., Liu, J., Desai, N., Ji, H., Han, J.: Constructing topical hierarchies in heterogeneous information networks. In: ICDM, pp. 767–776 (2013)
98. Wang, C., Song, Y., El-Kishky, A., Roth, D., Zhang, M., Han, J.: Incorporating world knowledge to document clustering via heterogeneous information networks. In: KDD, pp. 1215–1224 (2015)
99. Wang, C., Song, Y., Li, H., Zhang, M., Han, J.: Knowsim: a document similarity measure on structured heterogeneous information networks. In: ICDM, pp. 1015–1020 (2015)
100. Wang, C., Song, Y., Li, H., Zhang, M., Han, J.: Text classification with heterogeneous information network kernels. In: AAAI, pp. 2130–2136 (2016)
101. Wang, C., Sun, Y., Song, Y., Han, J., Song, Y., Wang, L., Zhang, M.: Relsim: relation similarity search in schema-rich heterogeneous information networks. In: Siam International Conference on Data Mining, pp. 621–629 (2016)
102. Wang, G., Hu, Q., Yu, P.S.: Influence and similarity on heterogeneous networks. In: CIKM, pp. 1462–1466 (2012)
103. Wang, Q., Peng, Z., Jiang, F., Li, Q.: LSA-PTM: a propagation-based topic model using latent semantic analysis on heterogeneous information networks. In: WAIM, pp. 13–24 (2013)
104. Wang, Q., Peng, Z., Wang, S., Yu, P.S., Li, Q., Hong, X.: cluTM: content and link integrated topic model on heterogeneous information networks. In: WAIM, pp. 207–218 (2015)
105. Wang, R., Shi, C., Yu, P.S., Wu, B.: Integrating clustering and ranking on hybrid heterogeneous information network. In: PAKDD, pp. 583–594 (2013)
106. Wang, S., Xie, S., Zhang, X., Li, Z., Yu, P.S., Shu, X.: Future influence ranking of scientific literature. In: SDM, pp. 749–757 (2014)
107. Wu, J., Chen, L., Yu, Q., Han, P., Wu, Z.: Trust-aware media recommendation in heterogeneous social networks. *WWW* **18**(1), 139–157 (2015)
108. Xiong, Y., Zhu, Y., Yu, P.S.: Top-k similarity join in heterogeneous information networks. *IEEE Trans. Knowl. Data Eng.* **27**(6), 1710–1723 (2015)
109. Yang, C., Sun, J., Ma, J., Zhang, S., Wang, G., Hua, Z.: Scientific collaborator recommendation in heterogeneous bibliographic networks. In: HICSS, pp. 552–561 (2015)
110. Yang, T., Jin, R., Chi, Y., Zhu, S.: Combining link and content for community detection: a discriminative approach. In: KDD, pp. 927–936 (2009)
111. Yang, X., Steck, H., Liu, Y.: Circle-based recommendation in online social networks. In: KDD, pp. 1267–1275 (2012)
112. Yang, Y., Chawla, N.V., Sun, Y., Han, J.: Predicting links in multi-relational and heterogeneous networks. In: ICDM, pp. 755–764 (2012)
113. Yang, Y., Tang, J., Keomany, J., Zhao, Y., Li, J., Ding, Y., Li, T., Wang, L.: Mining competitive relationships by learning across heterogeneous networks. In: CIKM, pp. 1432–1441 (2012)
114. Yu, P.S., Zhang, J.: MCD: mutual clustering across multiple social networks. In: IEEE International Congress on Big Data, pp. 762–771 (2015)
115. Yu, X., Gu, Q., Zhou, M., Han, J.: Citation prediction in heterogeneous bibliographic networks. In: SDM, pp. 1119–1130 (2012)

116. Yu, X., Sun, Y., Norick, B., Mao, T., Han, J.: User guided entity similarity search using meta-path selection in heterogeneous information networks. In: CIKM, pp. 2025–2029 (2012)
117. Yu, X., Ren, X., Sun, Y., Sturt, B., Khandelwal, U., Gu, Q., Norick, B., Han, J.: Recommendation in heterogeneous information networks with implicit user feedback. In: RecSys, pp. 347–350 (2013)
118. Yu, X., Ren, X., Sun, Y., Gu, Q., Sturt, B., Khandelwal, U., Norick, B., Han, J.: Personalized entity recommendation: a heterogeneous information network approach. In: WSDM, pp. 283–292 (2014)
119. Zhan, Q., Zhang, J., Wang, S., Yu, P.S., Xie, J.: Influence maximization across partially aligned heterogeneous social networks. In: PAKDD, pp. 58–69 (2015)
120. Zhan, Q., Zhang, J., Philip, S.Y., Emery, S., Xie, J.: Discover tipping users for cross network influencing. In: 2016 IEEE 17th International Conference on Information Reuse and Integration (IRI), pp. 67–76 (2016)
121. Zhang, A., Xie, X., Chang, K.C.C., Gunter, C.A., Han, J., Wang, X.: Privacy risk in anonymized heterogeneous information networks. In: EDBT, pp. 595–606 (2014)
122. Zhang, J., Yu, P.: Community detection for emerging networks. In: SDM, pp. 127–135 (2015)
123. Zhang, J., Yu, P.S.: Integrated anchor and social link predictions across social networks. In: IJCAI, pp. 2125–2131 (2015)
124. Zhang, J., Yu, P.S.: Multiple anonymized social networks alignment. In: ICDM, pp. 599–608 (2015)
125. Zhang, J., Yu, P.S.: PCT: partial co-alignment of social networks. In: WWW, pp. 749–759 (2016)
126. Zhang, J., Kong, X., Yu, P.S.: Predicting social links for new users across aligned heterogeneous social networks. In: ICDM, pp. 1289–1294 (2013)
127. Zhang, J., Yu, P.S., Zhou, Z.H.: Meta-path based multi-network collective link prediction. In: KDD, pp. 1286–1295 (2014)
128. Zhang, J., Kong, X., Yu, P.S.: Transferring heterogeneous links across location-based social networks. In: WSDM, pp. 303–312 (2014)
129. Zhang, J., Kong, X., Jie, L., Chang, Y., Yu, P.S.: NCR: a scalable network-based approach to co-ranking in question-and-answer sites. In: CIKM, pp. 709–718 (2014)
130. Zhang, J., Yu, P.S., Lv, Y.: Organizational chart inference. In: KDD, pp. 1435–1444 (2015)
131. Zhang, J., Shao, W., Wang, S., Kong, X., Yu, P.S.: Partial network alignment with anchor meta path and truncated generic stable matching. ArXiv e-prints (2015)
132. Zhang, J., Yu, P.S., Lv, Y., Zhan, Q.: Information diffusion at workplace. In: CIKM, pp. 1673–1682. ACM (2016)
133. Zhang, M., Hu, H., He, Z., Wang, W.: Top-k similarity search in heterogeneous information networks with x-star network schema. *Expert Syst. Appl.* **42**(2), 699–712 (2015)
134. Zhang, Y., Tang, J., Yang, Z., Pei, J., Yu, P.S.: COSNET: connecting heterogeneous social networks with local and global consistency. In: KDD, pp. 1485–1494 (2015)
135. Zhao, Q., Bhowmick, S.S., Zheng, X., Yi, K.: Characterizing and predicting community members from evolutionary and heterogeneous networks. In: CIKM, pp. 309–318 (2008)
136. Zheng, J., Liu, J., Shi, C., Zhuang, F., Li, J., Wu, B.: Dual similarity regularization for recommendation. In: PAKDD, pp. 542–554 (2016)
137. Zhou, D., Orshanskiy, S.A., Zha, H., Giles, C.L.: Co-ranking authors and documents in a heterogeneous network. In: ICDM, pp. 739–744 (2007)
138. Zhou, Y., Liu, L.: Social influence based clustering of heterogeneous information networks. In: KDD, pp. 338–346 (2013)
139. Zhou, Y., Liu, L.: Activity-edge centric multi-label classification for mining heterogeneous information networks. In: KDD, pp. 1276–1285 (2014)

140. Zhou, Y., Cheng, H., Yu, J.X.: Graph clustering based on structural/attribute similarities. In: VLDB, pp. 718–729 (2009)
141. Zhu, M., Zhu, T., Peng, Z., Yang, G., Xu, Y., Wang, S., Wang, X., Hong, X.: Relevance search on signed heterogeneous information network based on meta-path factorization. In: WAIM, pp. 181–192 (2015)
142. Zhuang, H., Zhang, J., Brova, G., Tang, J., Cam, H., Yan, X., Han, J.: Mining query-Based subnetwork outliers in heterogeneous information networks. In: ICDM, pp. 1127–1132 (2014)