Chapter 1
Introduction

Abstract In this chapter, we introduce some basic concepts and definitions in heterogeneous information network and compare the heterogeneous information network with other related concepts. Then, we give some popular examples in this field. In the end, we analyze the reason why mining heterogeneous information network is a new paradigm.

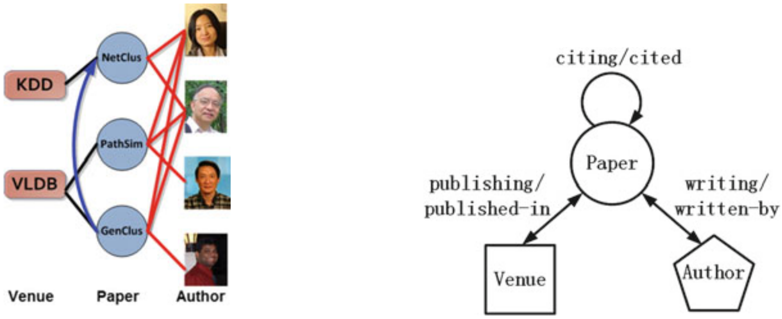
1.1 Basic Concepts and Definitions

As we know, most real systems usually consist of a large number of interacting, multityped components, such as human social activities, communication and computer systems, and biological networks. In such systems, the interacting components constitute interconnected networks, which can be called information networks without loss of generality. Clearly, information networks are ubiquitous and form a critical component of modern information infrastructure. The information network analysis has gained extremely wide attentions from researchers in many disciplines, such as computer science, social science, and physics. Particularly, the information network analysis has become a hot research topic in the fields of data mining and information retrieval in the preceding decades. The basic paradigm is to mine hidden patterns through mining link relations from networked data. The analysis of information network is related to the works in link mining and analysis [3, 4, 6], social network analysis [20, 34], hypertext and web mining [1], network science [12], as well as graph mining [2].

An information network represents an abstraction of the real world, focusing on the objects and the interactions among these objects. Formally, we define an information network as follows.

Definition 1.1 (*Information network* [27, 28]). An information network is defined as a directed graph $G = (V, E)$ with an object type mapping function $\varphi : V \to \mathtt{A}$ and a link type mapping function $\psi : E \to \mathtt{R}$. Each object $v \in V$ belongs to one particular object type in the object type set $\mathtt{A}$: $\varphi(v) \in \mathtt{A}$, and each link $e \in E$ belongs to a particular relation type in the relation type set $\mathtt{R}$: $\psi(e) \in \mathtt{R}$. If two links belong

C. Shi and P.S. Yu, *Heterogeneous Information Network Analysis and Applications*, Data Analytics, DOI 10.1007/978-3-319-56212-4_1



(a) Network instance (b) Network schema

Fig. 1.1 An example of heterogeneous information network on bibliographic data [27]

to the same relation type, the two links share the same starting object type as well as the ending object type.

Different from the traditional network definition, we explicitly distinguish object types and relation types in an information network and propose the concepts of heterogeneous/homogeneous information network. For simplicity, we also call heterogeneous information network as heterogeneous network or HIN in this book.

Definition 1.2 (*Heterogeneous/Homogeneous information network*). The information network is called **heterogeneous information network** if the types of objects $|\mathtt{A}| > 1$ or the types of relations $|\mathtt{R}| > 1$; otherwise, it is a **homogeneous information network**.

Example 1.1 Figure 1.1 shows an HIN example on bibliographic data [27]. A bibliographic information network, such as the bibliographic network involving computer science researchers derived from DBLP,[1] is a typical heterogeneous network containing three types of information entities: papers, venues, and authors. For each paper, it has links to a set of authors, and a venue, and these links belong to a set of link types.

In order to understand the object types and link types better in a complex heterogeneous information network, it is necessary to provide the meta-level (i.e., schema-level) description of the network. Therefore, the concept of network schema is proposed to describe the metastructure of a network.

Definition 1.3 (*Network schema* [27, 28]). The network schema, denoted as $T_G = (\mathtt{A}, \mathtt{R})$, is a metatemplate for an information network $G = (V, E)$ with the object type mapping $\varphi : V \to \mathtt{A}$ and the link type mapping $\psi : E \to \mathtt{R}$, which is a directed graph defined over object types $\mathtt{A}$, with edges as relations from $\mathtt{R}$.

[1] http://dblp.uni-trier.de/.

The network schema of a heterogeneous information network specifies type constraints on the sets of objects and relationships among the objects. These constraints make a heterogeneous information network semi-structured, guiding the semantics explorations of the network. An information network following a network schema is called a **network instance** of the network schema. For a link type R connecting object type S to object type T, i.e., $S \xrightarrow{R} T$, S and T are the **source object type** and **target object type** of link type R, which can be denoted as $R.S$ and $R.T$, respectively. The inverse relation R^{-1} holds naturally for $T \xrightarrow{R^{-1}} S$. Generally, R is not equal to R^{-1}, unless R is symmetric.

Example 1.2 As described above, Fig. 1.1a demonstrates the real objects and their connections on bibliographic data. Figure 1.1b illustrates its network schema which describes the object types and their relations in the HIN. Moreover, Fig. 1.1a is a network instance of the network schema Fig. 1.1b. In this example, it contains objects from three types of objects: papers (P), authors (A), and venues (V). There are links connecting different types of objects. The link types are defined by the relations between two object types. For example, links existing between authors and papers denote the writing or written-by relations, while those between venues and papers denote the publishing or published-in relations.

Different from homogeneous networks, two objects in a heterogeneous network can be connected via different paths and these paths have different physical meanings. These paths can be categorized as meta paths as follows.

Definition 1.4 (*Meta path* [29]). A meta path P is a path defined on a schema $S = (\mathtt{A}, \mathtt{R})$, and is denoted in the form of $A_1 \xrightarrow{R_1} A_2 \xrightarrow{R_2} \ldots \xrightarrow{R_l} A_{l+1}$, which defines a composite relation $R = R_1 \circ R_2 \circ \cdots \circ R_l$ between objects $A_1, A_2, \cdots, A_{l+1}$, where $\circ$ denotes the composition operator on relations.

For simplicity, we can also use object types to denote the meta path if there are no multiple relation types between the same pair of object types: $P = (A_1 A_2 \cdots A_{l+1})$. For example, in Fig. 1.1a, the relation, authors publishing papers in conferences, can be described using the length-2 meta path $A \xrightarrow{writting} P \xrightarrow{written-by} A$, or APA for short. We say a concrete path $p = (a_1 a_2 \cdots a_{l+1})$ between objects a_1 and a_{l+1} in network G is a **path instance** of the relevance path P, if for each a_i, $\phi(a_i) = A_i$ and each link $e_i = \langle a_i, a_{i+1} \rangle$ belongs to the relation R_i in P. It can be denoted as $p \in P$. A meta path P is a **symmetric path**, when the relation R defined by it is symmetric (i.e., P is equal to P^{-1}), such as APA and $APVPA$. Two meta paths $P_1 = (A_1 A_2 \cdots A_l)$ and $P_2 = (B_1 B_2 \cdots B_k)$ are **concatenable** if and only if A_l is equal to B_1, and the concatenated path is written as $P = (P_1 P_2)$, which equals to $(A_1 A_2 \cdots A_l B_2 \ldots B_k)$. A simple concatenable example is that AP and PA can be concatenated to the path APA.

Example 1.3 Consider the examples shown in Fig. 1.2, authors can be connected via meta paths such as "Author-Paper-Author" (APA) path, "Author-Paper-Venue-Paper-Author" ($APVPA$) path, and so on. Moreover, Table 1.1 shows path instances

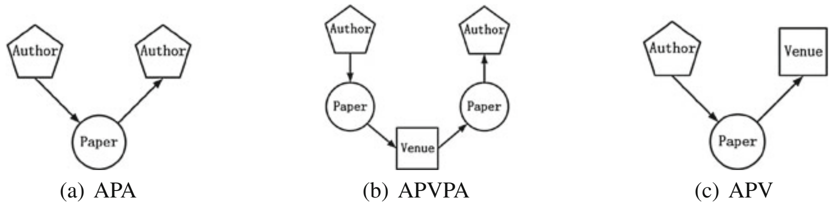


Fig. 1.2 Examples of meta paths in heterogeneous information network on bibliographic data

Table 1.1 Meta path examples and their physical meanings on bibliographic data

Path instance	Meta path	Physical meaning
Sun-NetClus-Han Sun-PathSim-Yu	Author-Paper-Author (*APA*)	Authors collaborate on the same paper
Sun-PathSim-VLDB-PathSim-Han Sun-PathSim-VLDB-GenClus-Aggarwal	Author-Paper-Venue-Paper-Author (*APVPA*)	Authors publish papers on the same venue
Sun-NetClus-KDD Sun-PathSim-VLDB	Author-Paper-Venue (*APV*)	Authors publish papers at a venue

and semantics of these meta paths. It is obvious that semantics underneath these paths are different. The APA path means authors collaborating on the same papers (i.e., co-author relation), while $APVPA$ path means authors publishing papers on the same venue. The meta paths can also connect different types of objects. For example, the authors and venues can be connected with the APV path, which means authors publishing papers on venues.

The rich semantics of meta paths is an important characteristic of HIN. Based on different meta paths, objects have different connection relations with diverse path semantics, which may have an effect on many data mining tasks. For example, the similarity scores among authors evaluated based on different meta paths are different [29]. Under the APA path, the authors who co-publish papers will be more similar, while the authors who publish papers on the same venues will be more similar under the $APVPA$ path. Another example is the importance evaluation of objects [13]. The importance of authors under APA path has a bias on the authors who write many multiauthor papers, while the importance of authors under $APVPA$ path emphasizes the authors who publish many papers on those productive conferences. As a unique characteristic and effective semantic capturing tool, meta path has been widely used in many data mining tasks in HIN, such as similarity measure [22, 29], clustering [30], and classification [10].

1.2 Comparisons with Related Concepts

With the boom of social network analysis, all kinds of networked data have emerged, and numbers of concepts to model networked data have been proposed. These concepts have similar meanings, as well as subtle differences. For example, the multitype relational data proposed by Long et al. [18] is an HIN in deed, and the multiview data [15] can also be organized as an HIN. Here, we compare the heterogeneous network concept with those most related concepts.

Heterogeneous network versus homogeneous network. Heterogeneous networks include different types of nodes or links, while homogeneous networks only have one type of objects and links. Homogeneous networks can be considered as a special case of heterogeneous networks. Moreover, a heterogeneous network can be converted into a homogeneous network through network projection or ignoring object heterogeneity, while it will make significant information loss. Traditional link mining [11, 14, 32] is usually based on the homogeneous network, and many analysis techniques on homogeneous network cannot be directly applied to heterogeneous network.

Heterogeneous network versus multirelational network [36]. Different from heterogeneous network, multirelational network has only one type of objects, but more than one kind of relationship between objects. So multirelational network can be seen as a special case of heterogeneous network.

Heterogeneous network versus multidimensional/mode network [31]. Tang et al. [31] proposed the multidimensional/mode network concept, which has the same meaning with multirelational network. That is, the network has only one type of objects and more than one kind of relationship between objects. So multidimensional/mode network is also a special case of heterogeneous network.

Heterogeneous network versus composite network [39, 40]. Qiang Yang et al. proposed the composite network concept [39, 40], where users in networks have various relationships, exhibit different behaviors in each individual network or subnetwork, and share some common latent interests across networks at the same time. So composite network is, in fact, a multirelational network, a special case of heterogeneous network.

Heterogeneous network versus complex network. A complex network is a network with non-trivial topological features and patterns of connection between its elements that are neither purely regular nor purely random [7]. Such non-trivial topological features include a heavy tail in the degree distribution, a high clustering coefficient, community structure, and hierarchical structure. The studies of complex networks have brought together researchers from many areas, including mathematics, physics, biology, computer science, sociology, and others. The studies show that many real networks are complex networks, such as social networks, information networks, technological networks, and biological networks [19]. So we can say that many real heterogeneous networks are complex networks. However, the studies on complex networks usually focus on the structures, functions, and features of networks.

1.3 Example Datasets of Heterogeneous Information Networks

Intuitively, most real systems include multityped interacting objects. For example, a social media website (e.g., Facebook) contains a set of object types, such as users, posts, and tags, and a health care system contains doctors, patients, diseases, and devices. Generally speaking, these interacting systems can all be modeled as heterogeneous information networks. Concretely, this kind of networks can be constructed from the following three types of data.

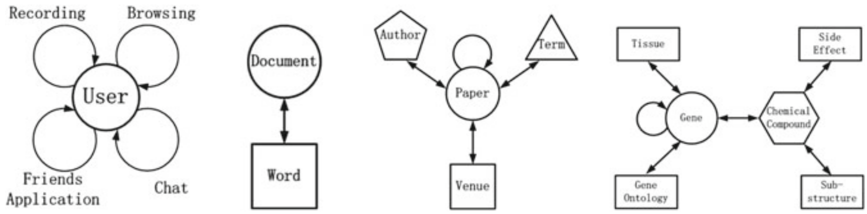
1. **Structured data**. Structured data stored in database table is organized with entity-relation model. The different-typed entities and their relations naturally construct information networks. For example, the bibliographic data (see the above example) is widely used as heterogeneous information network.
2. **Semi-structured data**. Semi-structured data is usually stored with XML format. The attributes in XML can be considered as object types, and the object instances can be determined by analyzing the contents of attributes. The connections among attributes construct object relations.
3. **Non-structured data**. For non-structured data, heterogeneous information networks can also be constructed by objects and relationship extraction. For example, for text data, entity recognition and relation extraction can form the objects and links of HIN.

Although heterogeneous information networks are ubiquitous, there are not many standard datasets for study, since these heterogeneous information usually exist in different data sources. Here, we summarize some widely used heterogeneous networks in literatures.

Multirelational network with single-typed object. Traditional multirelational network is a kind of HINs, where there is one type of object and several types of relations among objects. This kind of networks widely exists in social websites, such as Facebook and Xiaonei [40]. Figure 1.3a shows the network schema of such a network [40], where users can be extensively connected with each other through connections, such as recording, browsing, chatting, and sending friends applications.

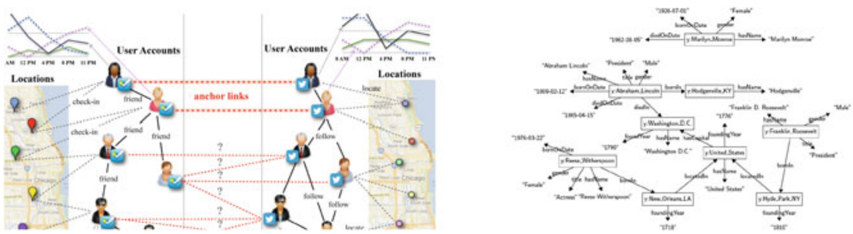
Bipartite network. As a typical HIN, bipartite network is widely used to construct interactions among two types of objects, such as user–item [5] and document–word [16]. Figure 1.3b shows the schema of a bipartite network connecting documents and words [16]. As an extension of bipartite graphs, k-partite graphs [17] contain multiple types of objects where links exist among adjacent object types. The bipartite network has been well studied for a long time. As the simplest HIN, we will not discuss this type of network in this book.

Star-schema network. Star-schema network is the most popular HIN in this field. In the database table, a target object and its attribute objects naturally construct an HIN, where the target object, as the hub node, connects different attribute objects. As an example shown in Fig. 1.3c, a bibliographic information network is a typical star-schema heterogeneous network [22, 29], containing different objects (e.g., paper,



(a) Multi-relation (b) Bipartite (c) Star-schema (d) Multiple-hub

Fig. 1.3 Network schema of heterogeneous information networks



(a) Multiple HINs [10] (b) Schema-rich HIN [42]

Fig. 1.4 Two examples of complex heterogeneous information network

venue, author, and term) and links among them. Many other datasets can also be represented as star-schema networks, such as the movie data [23, 37] from the Internet Movie Database [2] (IMDB) and the patent data [41] from US patents data.[3]

Multiple-hub network. Beyond star schema, some networks have more complex structures, which involve multiple-hub objects. This kind of networks widely exists in bioinformatics data [8, 33]. A bioinformatics example is shown in Fig. 1.3d, includes two hubs: gene and chemical compound. Another example can be found in the Douban dataset [4] [24].

Besides these widely used networks, many real systems can also be constructed as more complex heterogeneous networks. In some real applications, users may exist in multiple social networks, and each social network can be modeled as an HIN. Figure 1.4a shows an example of two heterogeneous social networks (Twitter and Foursquare) [9]. In each network, users are connected with each other through social links, and they are also connected with a set of locations, timestamps, and text contents through online activities. Moreover, some users have two accounts in two social networks separately, and they serve as anchor nodes to connect two networks. More generally, some interaction systems are too complex to be modeled as an HIN with a simple network schema. Knowledge graph [25] is such an example. We know that knowledge graph is based on resource description framework (RDF) data [21],

[2] www.imdb.com/.

[3] http://www.uspto.gov/patents/.

[4] http://www.douban.com/.

which complies with an $< Subject, Property, Object >$ model. Here, "Subject" and "Object" can be considered as objects, and "Property" can be considered as the relation between "Subject" and "Object". And thus a knowledge graph can be considered as a heterogeneous network, and such an example is shown in Fig. 1.4b. In such a semantic knowledge base, like Yago [26], there are more than 10-million entities (or nodes) of different types, and more than 120-million links among these entities. In such a schema-rich network, it is impossible to depict such network with a simple network schema.

In HIN, we distinguish the types of nodes and links, which should introduce some novel pattern discovery, compared to traditional homogeneous networks. Although many networked data can be modeled as heterogeneous networks, heterogeneous networks still have some limitations. Firstly, some real data are too complex to be modeled as meaningful HINs. For example, we can consider the RDF data as an HIN, while we cannot simply depict its network schema. Secondly, it may be difficult to analyze some networked data with an HIN perspective, even these data can be modeled as an HIN. These limitations are also the future works of HIN. We need to design more powerful mining methods in HIN to make it capable to be applied in more applications and discover more novel patterns.

1.4 Why Heterogeneous Information Network Analysis

In the past decades, link analysis has been extensively explored [4]. So many methods have been developed for information network analysis, and numerous data mining tasks have been explored in homogeneous networks, such as ranking, clustering, link prediction, and influence analysis. However, due to some unique characteristics (e.g., fusion of more information and rich semantics) of HIN, most methods in homogeneous networks cannot be directly applied in heterogeneous networks, and it is potential to discover more interesting patterns in this kind of networks.

It is a new development of data mining. Early data mining problems focused on analyzing feature vectors of objects. In the late 1990s, with the advent of WWW, more and more data mining researchers turned to studying links among objects. It is one of the main research directions to mine hidden patterns from feature and link information of objects. In these researches, homogeneous networks are usually constructed from interconnected objects. In recent years, abundant social media emerge, and many different types of objects are interconnected. It is hard to model these interacted objects as homogeneous networks, while it is natural to model different types of objects and relations among them as heterogeneous networks. Particularly, with the rapid increment of user-generated content online, big data analysis is an emergent yet important task to be studied. Variety is one significant characteristic of big data [35]. As a semi-structured representation, heterogeneous information network can be an effective way to model complex objects and their relations in big data.

It is an effective tool to fuse more information. Compared to homogeneous network, heterogeneous network is natural to fuse more objects and their interactions.

In addition, traditional homogeneous networks are usually constructed from single data source, while heterogeneous network can fuse information across multiple data sources. For example, customers use many services provided by Google, such as Google search, G-mail, maps, and Google+. So we can fuse this information with a heterogeneous information network, in which customers interact with many different types of objects, such as key words, mails, locations, and followers. Broadly speaking, heterogeneous information network can also fuse information cross multiple social network platforms [9]. We know that there are many social network platforms with different objectives, such as Facebook, Twitter, WeChat, and Weibo. Moreover, users often participate in multiple social networks. Since each social network only captures a partial or biased view of a user, we can fuse information across multiple social network platforms with multiple heterogeneous information networks, where each heterogeneous network represents information from one social network with some anchor nodes connecting these networks [38].

It contains rich semantics. In heterogeneous networks, different-typed objects and links coexist and they carry different semantic meanings. As a bibliographic example shown in Fig. 1.1, it includes author, paper, and venue object types. The relation type "Author-Paper" means authors writing papers, while the relation type "Paper-Venue" means papers published in venues. Considering the semantic information will lead to more subtle knowledge discovery. For example, in DBLP bibliographic data [29], if you find the most similar authors to "Christos Faloutsos," you will get his students, like Spiros Papadimitriou and Jimeng Sun, under the APA path, while the results are reputable researchers, like Jiawei Han and Rakesh Agrawal, under the $APVPA$ path. How to mine interesting patterns with the semantic information is a unique issue in heterogeneous networks.

References

1. Chakrabarti, S., et al.: Mining the Web: Analysis of Hypertext and Semi Structured Data. Morgan Kaufmann, San Francisco (2002)
2. Cook, D.J., Holder, L.B.: Graph-based data mining. IEEE Intell. Syst. **15**(2), 32–41 (2000)
3. Feldman, R.: Link analysis: current state of the art. In: Tutorial at the KDD-2 (2002)
4. Getoor, L., Diehl, C.P.: Link mining: a survey. SIGKDD Explor. **7**(2), 3–12 (2005)
5. Jamali, M., Lakshmanan, L.: HeteroMF: recommendation in heterogeneous information networks using context dependent factor models. In: WWW, pp. 643–654 (2013)
6. Jensen, D., Goldberg, H.: AAAI Fall Symposium on AI and Link Analysis. AAAI Press (1998)
7. Kim, J., Wilhelm, T.: What is a complex graph? Phys. A Stat. Mech. Appl. **387**(11), 2637–2652 (2008)
8. 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)
9. Kong, X., Zhang, J., Yu, P.S.: Inferring anchor links across multiple heterogeneous social networks. In: CIKM, pp. 179–188 (2013)
10. 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)
11. Konstas, I., Stathopoulo, V., Jose, J.M.: On social networks and collaborative recommendation. In: SIGIR, pp. 195–202 (2009)

12. Lewis, T.G.: Network Science: Theory and Applications. Wiley, New York (2011)
13. 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)
14. Liben-Nowell, D., Kleinberg, J.: The link prediction problem for social networks. In: CIKM, pp. 556–559 (2003)
15. Liu, J., Wang, C., Gao, J., Han, J.: Multi-view clustering via joint nonnegative matrix factorization. In: SDM, pp. 252–260 (2013)
16. Long, B., Zhang, Z.M., Yu, P.S.: Co-clustering by block value decomposition. In: KDD, pp. 635–640 (2005)
17. Long, B., Wu, X., Zhang, Z., Yu, P.S.: Unsupervised learning on k-partite graphs. In: KDD, pp. 317–326 (2006)
18. Long, B., Zhang, Z., Wu, X., Yu, P.S.: Spectral clustering for multi-type relational data. In: ICML, pp. 585–592 (2006)
19. Newman, M.E.: The structure and function of complex networks. SIAM Rev. **45**(2), 167–256 (2003)
20. Otte, E., Rousseau, R.: Social network analysis: a powerful strategy, also for the information sciences. J. Inf. Sci. **28**(6), 441–453 (2002)
21. Özsu, M.T.: A survey of RDF data management systems. Front. Comput. Sci. **10**(3), 418–432 (2016)
22. Shi, C., Kong, X., Yu, P.S., Xie, S., Wu, B.: Relevance search in heterogeneous networks. In: International Conference on Extending Database Technology, pp. 180–191 (2012)
23. 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)
24. Shi, C., Zhang, Z., Luo, P., Yu, P.S., Yue, Y., Wu, B.: Semantic path based personalized recommendation on weighted heterogeneous information networks. In: The ACM International, pp. 453–462 (2015)
25. Singhal, A.: Introducing the Knowledge Graph: things, not strings. In: Official Google Blog (2012)
26. Suchanek, F.M., Kasneci, G., Weikum, G.: Yago: A core of semantic knowledge. In: WWW, pp. 697–706 (2007)
27. Sun, Y., Han, J.: Mining heterogeneous information networks: a structural analysis approach. SIGKDD Explor. **14**(2), 20–28 (2012)
28. Sun, Y., Yu, Y., Han, J.: Ranking-based clustering of heterogeneous information networks with star network schema. In: KDD, pp. 797–806 (2009)
29. Sun, Y., Han, J., Yan, X., Yu, P., Wu, T.: PathSim: meta path-based top-k similarity search in heterogeneous information networks. In: VLDB, pp. 992–1003 (2011)
30. 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)
31. Tang, L., Liu, H., Zhang, J., Nazeri, Z.: Community evolution in dynamic multi-mode networks. In: KDD, pp. 677–685 (2008)
32. Tang, J., Gao, H., Hu, X., Liu, H.: Exploiting homophily effect for trust prediction. In: WSDM, pp. 53–62 (2013)
33. Wang, R., Shi, C., Yu, P.S., Wu, B.: Integrating clustering and ranking on hybrid heterogeneous information network. In: PAKDD, pp. 583–594 (2013)
34. Wasserman, S.: Social Network Analysis: Methods and Applications. Cambridge University Press, Cambridge (1994)
35. Wu, X., Zhu, X., Wu, G., Ding, W.: Data mining with big data. IEEE Trans. Knowl. Data Eng. **26**(1), 97–107 (2014)
36. Yang, Y., Chawla, N.V., Sun, Y., Han, J.: Predicting links in multi-relational and heterogeneous networks. In: ICDM, pp. 755–764 (2012)
37. 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)

38. Zhang, J., Yu, P.S.: Integrated anchor and social link predictions across social networks. In: Proceedings of the 24th International Conference on Artificial Intelligence, IJCAI'15, pp. 2125–2131. AAAI Press (2015)
39. Zhong, E., Fan, W., Wang, J., Xiao, L., Li, Y.: ComSoc: adaptive transfer of user behaviors over composite social network. In: KDD, pp. 696–704 (2012)
40. Zhong, E., Fan, W., Zhu, Y., Yang, Q.: Modeling the dynamics of composite social networks. In: KDD, pp. 937–945 (2013)
41. 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)
42. Zou, L., Özsu, M.T., Chen, L., Shen, X., Huang, R., Zhao, D.: gStore: a graph-based SPARQL query engine. VLDB J. **23**(4), 565–590 (2014)