Chapter 9
Future Research Directions

Abstract Although many data mining tasks have been exploited in heterogeneous information network, it is still a young and promising research field. Here, we illustrate some advanced topics, including challenging research issues and unexplored tasks, and point out some potential future research directions.

9.1 More Complex Network Construction

There is a basic assumption in contemporary researches that a heterogeneous information network to be investigated is well defined, and objects and links in the network are clean and unambiguous. However, it is not the case in real applications. In fact, constructing heterogeneous information network from real data often faces challenges.

If the networked data are structured data, like relational database, it may be easy to construct a heterogeneous information network with well-defined schema, such as DBLP network [36] and Movie network [28, 52]. However, even in this kind of heterogeneous networks, objects and links can still be noisy. (1) Objects in a network may not exactly correspond to entities in real world, such as duplication of name [47] in bibliography data. That is, one object in a network may refer to multiple entities, or different objects may refer to the same entity. We can integrate entity resolution [1] with network mining to clean objects or links beforehand. For example, Shen et al. [27] propose a probabilistic model SHINE to link named entity mentions detected from the unstructured Web text with their corresponding entities existing in a heterogeneous information network. Ren et al. [26] propose a relation phrase-based entity recognition framework, called ClusType. The framework runs data-driven phrase mining to generate entity mention candidates and relation phrases, and enforces the principle that relation phrases should be softly clustered when propagating type information in a heterogeneous network constructed by argument entities. (2) Relations among objects may not be explicitly given or not complete sometimes, e.g., the advisor–advisee relationship in the DBLP network [38]. Link prediction [18] can be employed to fill out the missing relations for comprehensive networks. (3) Objects and links may not be reliable or trustable, e.g., the inaccurate item information in an

C. Shi and P.S. Yu, *Heterogeneous Information Network Analysis and Applications*, Data Analytics, DOI 10.1007/978-3-319-56212-4_9

E-commerce Web site and conflicting information of certain objects from multiple Web sites. However, an HIN can be built to capture the dependency relations among the node entities to clean up and integrate the data, such as trustworthiness modeling [48, 59], spam detection [45], and co-ranking of questions, answers, and users in a Q&A system.

If the networked data are unstructured data, such as text data, multimedia data, and multilingual data, it becomes more challenging to construct qualified heterogeneous information networks. In order to construct high-quality HINs, information extraction, natural language processing, and many other techniques should be integrated with network construction. Mining quality phrases is a critical step to form entities of networks from text data. Kishky et al. [6] propose a computationally efficient and effective model ToPMine, which first executes a phrase mining framework to segment a document into single and multiword phrases, and then employs a new topic model that operates on the induced document partition. Furthermore, Liu et al. [21] propose an effective and scalable method SegPhrase+ that integrates quality phrases extraction with phrasal segmentation. Beyond the bag-of-word representation of text data, some researchers try to represent a document with the help of heterogeneous information network. Wang et al. [41] firstly map entities in documents into a knowledge base (e.g., Freebase), and then consider the knowledge base as an HIN to mine internal relations among entities. Furthermore, Wang et al. [40, 43] employ world knowledge as indirect supervision to improve the document clustering results. More recently, Wang et al. [42] propose the HIN-kernel concept for classification through representing a text as an HIN. Relationship extraction is another important step to form links among the objects in network. Wang et al. [38] mine hidden advisor–advisee relationships from bibliographic data, and they further infer hierarchical relationships among partially ordered objects with heterogeneous attributes and links [39]. Broadly speaking, we can also extract entity and relationship to construct heterogeneous network from multimedia data and multilingual data, as we have done on text data.

9.2 More Powerful Mining Methods

For ubiquitous heterogeneous information networks, numbers of mining methods have been proposed on many data mining tasks. As we have mentioned, heterogeneous information networks have two important characteristics: complex structure and rich semantics. According to these two characteristics, we summarize the contemporary works and point out future directions.

9.2.1 Network Structure

In heterogeneous network, objects can be organized in different forms. Bipartite graph is widely used to organize two types of objects and the relations between them [10, 23]. As an extension of bipartite graphs, K-partite graphs [22] are able to represent multiple types of objects. Recently, heterogeneous networks are usually organized as star-schema networks, such as bibliographic data [29, 34, 36] and movie data [28, 52]. To combine the heterogeneous and homogeneous information, star-schema with self loop is also proposed [46]. Different from only one hub object type existing in star-schema network, some networked data have multiple hub object types, e.g., the bioinformatics data [31]. For this kind of networks, Shi et al. [31] propose a HeProjI method which projects a general heterogeneous network into a sequence of subnetworks with bipartite or star-schema structure.

In applications, the networked data are usually more complex and irregular. Some real networks may contain attribute values on links, and these attribute values may contain important information. For example, users usually rate movies with a score from 1 to 5 in movie recommended system, where the rating scores represent users' attitudes to movies, and the "author of" relation between authors and papers in bibliographic networks can take values (e.g., 1, 2, 3) which represents the order of authors in the paper. In this kind of applications, we need to consider the effect of attribute values on the weighted heterogeneous information network [32]. There are some time series data, for example, a period of biographic data and rating information of users and movies. For this kind of data, we need to construct dynamic heterogeneous network [35] and consider the effect of the time factor. In some applications, one kind of objects may exist in multiple heterogeneous networks [12, 54]. For example, users usually co-exist in multiple social networks, such as Facebook, Google+, and Twitter. In this kind of applications, we need to align users in different networks and effectively fuse information from different networks [55–57]. More broadly, many networked data are difficult to be modeled with heterogeneous network with a simple network schema. For example, in RDF data, there are so many types of objects and relations, which cannot be described with network schema [25, 40]. Many research problems arise with this kind of schema-rich HINs [3, 44], for example, management of objects and relations with so many types and automatic generation of meta paths. As the real networked data become more complex, we need to design more powerful and flexible heterogeneous networks, which also provides more challenges for data mining.

9.2.2 Semantic Mining

As the unique characteristic, objects and links in HIN contain rich semantics. Meta path can effectively capture subtle semantics among objects, and many works have made use of the meta path-based mining tasks. For example, in similarity measure task, object pairs have different similarities under different meta paths [29, 36]; in recommendation task, different items will be recommended under different paths [32]. In addition, meta path is also widely used for feature extraction. Object similarity can be measured under different meta paths, which can be used as feature vectors for many tasks, such as clustering [37], link prediction [2], and recommendation [53].

However, some researchers have noticed the shortcomings of meta path. Since meta path fails to capture more microsemantics. In some applications, some researchers consider to refine meta path with some constraints. For example, the "Author-Paper-Author" path describes the collaboration relation among authors. However, it cannot depict the fact that Philip S. Yu and Jiawei Han have many collaborations in data mining field but they seldom collaborate in information retrieval field. In order to overcome the shortcoming existing in meta path, Shi et al. [16] propose the constrained meta path concept, which can confine some constraints on objects. Taking Fig. 1.3c in Chap. 1 as an example, the constrained meta path $APA|P.L =$ "$Data\ Mining$" represents the co-author relation of authors in data mining field through constraining the label of papers with "Data Mining." Moreover, Liu et al. [20] propose the concept "restricted meta-path" which enables in-depth knowledge mining on the heterogeneous bibliographic networks by allowing restrictions on the node set. In addition, traditional HIN and meta path do not consider the attribute values on links, while weighted links are very common in practical applications. Examples include rating scores between users and items in recommended system and the order of authors in papers in bibliographic network. Taking Fig. 5.2 in Chap. 5 as an example, the rating relation between users and movies can take scores from 1 to 5. Shi et al. [32] propose weighted meta path to consider attribute values on links and more subtly capture path semantics through distinguishing different link attribute values.

On the other hand, some researchers consider to capture more macro semantics through combining multiple-related meta paths. For example, two authors write two different papers that both mention the mining term and are published in the same venue, while another two authors also write two different papers that are published in the same venue and have not the same terms. Therefore, these two authors in the first case should have a higher relevance score than those two authors in the second case. However, the single meta path either $APVPA$ or $APTPA$ fails to discover this factor. In order to solve this shortcoming, Huang et al. [9] propose the relevance measure based on metastructure which is a combination of meta paths. Similarly, Fang et al. [7] propose the metagraph which is a subgraph defined on a graph schema and can measure the semantic proximity between objects. As an effective semantic capture tool, meta path has shown its power in semantic capture and feature selection. However, it may be coarse in some applications, so we need to extend traditional meta path for more subtle semantic capture. Broadly speaking, we can also design new and more powerful semantic capture tools.

More importantly, the meta path approach faces challenges on path selection and their weight importances. How can we select meta paths in real applications? Theoretically, there are infinite meta paths in an HIN. In contemporary works, the network schema of HIN is usually small and simple, so we can assign some short

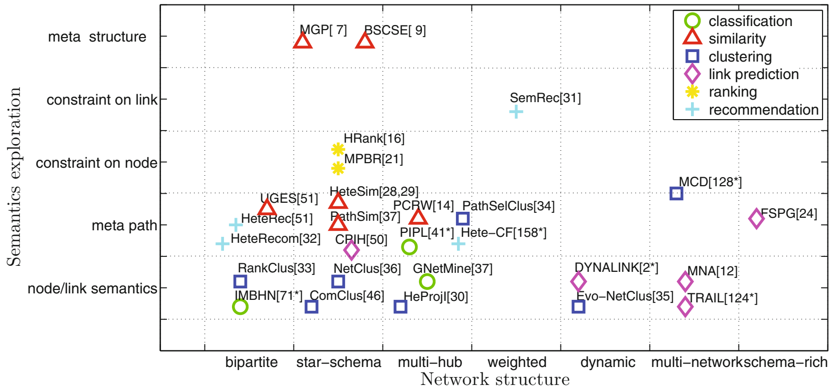

Fig. 9.1 Summarization of typical works on HIN according to network structure and semantic exploration. '∗' in a citation means this citation is from Chap. 2

and meaningful meta paths according to domain knowledge and experiences. Sun et al. [36] have validated that the long meta paths are not meaningful and they fail to produce good similarity measures. However, there is no work to study the effect of long meta paths on other mining tasks. In addition, there are so many meta paths even for short paths in some complex networks, like RDF network. It is a critical task to extract meta paths automatically in this condition. Recently, Meng et al. [25] study how to discover meta paths automatically which can best explain the relationship between node pairs. Another important issue is to determine the weights of meta paths automatically. Some methods have been proposed to explore this issue. For example, Lao et al. [14] employ a supervised method to learn weights, and Sun et al. [37] combine meta-path selection and user-guided information for clustering. In addition, Liang et al. [17] seek to find the K, most interesting path instances matching the preferred relationship type. Some interesting works are still worth doing. The ideal path weights learned should embody the importance of paths and reflect users' preferences. However, the similarity evaluations based on different paths have significant bias, which may make path weights hard to reflect path importances. So prioritized path weights are needed. In addition, if there are numerous meta paths in real applications (e.g., RDF network), the path weight learning will be more important and challenging.

In Fig. 9.1, we summarize some typical works in the HIN field from two perspectives: network structure and semantic exploration. We respectively select several typical works from six mining tasks mentioned above and put these works in a coordinate according to network structure and semantics exploration in these works. Note that we denominate those un-named methods with the first letter of keywords in the title, such as UGES [51] and CPIH [50]. Along the X-axis, the network structure becomes more complex, and semantics information becomes richer along the Y-axis. For example, RankClus [33] is designed for bi-type networks and only captures link semantics (different-typed links contain different semantics), while PathSim [36] can deal with more complex star-schema networks and use meta path to capture deeper semantics. Further, SemRec [32] adds constraints to links to explore more subtle semantic information in a weighted HIN. From the figure, we can also find that most contemporary works focus on simple network structures (e.g., bipartite or star-schema networks) and primary semantic exploration (e.g., meta path). In the future, we can exploit more complex heterogeneous networks with more powerful semantics capture tools.

9.3 Bigger Networked Data

In order to illustrate the benefits of HIN, we need to design data mining algorithms on big-networked data in wider domains. This variety is an important characteristic of big data. HIN is a powerful tool to handle the diversity of big data, since it can flexibly and effectively integrate varied objects and heterogeneous information. However, it is non-trivial work to build a real HIN-based analysis system. Besides research challenges mentioned above, such as network construction, it will face many practical technique challenges. A real HIN is huge, even dynamic, so it usually cannot be contained in memory and cannot be handled directly. We know that a user at a time could be only interested in a tiny portion of nodes, links, or subnetworks. Instead of directly mining the whole network, we can mine hidden but small networks "extracted" dynamically from some existing networks, based on user-specified constraints or expected node/link behaviors. How to discover such hidden networks and mine knowledge (e.g., clusters, behaviors, and anomalies) from such hidden but non-isolated networks could be an interesting but challenging problem.

Most of contemporary data mining tasks on HIN only work on small dataset and fail to consider the quick and parallel process on big data. Some research works have begun to consider the quick computation of mining algorithms on HIN. For example, Sun et al. [36] design a co-clustering-based pruning strategy to fasten the processing speed of PathSim. Lao et al. [13] propose the quick computation strategies of PCRW, and Shi et al. [24, 30] also consider the quick/parallel computation of HeteSim. In addition, cloud computing also provides an option to handle big-networked data. Although parallel graph mining algorithms [4] and platforms [11] have been proposed, parallel HIN analysis methods face some unique challenges. For example, the partition of HIN needs to consider the overload balances of computing nodes, as well as balances of different-typed nodes. Moreover, it is also challenging to mine integrated path semantics in partitioned subgraphs.

9.4 More Applications

Due to unique characteristics of HIN, many data mining tasks have been explored on HIN, which are summarized as above. In fact, more data mining tasks can be studied on HIN. Here, we introduce two potential applications.

The online analytical processing (OLAP) has shown its power in multidimensional analysis of structured relational data [5]. The similar analysis can also be done, when we view a heterogeneous information network from different angles and at different levels of granularity. Taking a bibliographic network as an example, we can observe the change of published papers on a conference in the time or district dimension, when we designate papers and conferences as the object types and publish relations as the link type. Some preliminary studies have been done on this issue. Zhao et al. [58] introduce graph cube to support OLAP queries effectively on large multidimensional networks; Li et al. [15] design InfoNetOLAPer to provide topic-oriented, integrated, and multidimensional organizational solutions for information networks. Yin et al. [49] have developed a novel HMGraph OLAP framework to mine multidimensional heterogeneous information networks with more dimensions and operations. These works consider link relation as a measure. However, they usually ignore semantic information in heterogeneous networks determined by multiple nodes and links. So the study of online analytical processing of heterogeneous information networks is still worth exploring.

Information diffusion is a vast research domain and has attracted research interests from many fields, such as physics and biology. Traditional information diffusion is studied on homogeneous networks [8], where information is propagated in one single channel. However, in many applications, pieces of information or diseases are propagated among different types of objects. For example, diseases could propagate among people, different kinds of animals, and food, via different channels. Few works explore this issue. Liu et al. [19] propose a generative graphical model which utilizes the heterogeneous link information and the textual content associated with each node to mine topic-level direct influence. In order to capture better spreading models that represent the real-world patterns, it is desirable to pay more attention to the study of information diffusion in heterogeneous information networks.

References

1. Bhattacharya, I., Getoor, L.: Collective entity resolution in relational data. ACM Trans. Knowl. Discov. Data **1**(1), 5 (2007)
2. Cao, B., Kong, X., Yu, P.S.: Collective prediction of multiple types of links in heterogeneous information networks. In: ICDM, pp. 50–59 (2014)
3. Cao, X., Zheng, Y., Shi, C., Li, J., Wu, B.: Link prediction in Schema-Rich heterogeneous information network. In: PAKDD, pp. 449–460 (2016)
4. Cohen, J.: Graph twiddling in a MapReduce world. Comput. Sci. Eng. **11**(4), 29–41 (2009)
5. Colliat, G.: OLAP, relational, and multidimensional database systems. ACM Sigmod Rec. **25**(3), 64–69 (1996)
6. El-Kishky, A., Song, Y., Wang, C., Voss, C.R., Han, J.: Scalable topical phrase mining from text corpora. PVLDB **8**(3), 305–316 (2014)
7. 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)
8. Gruhl, D., Guha, R., Liben-Nowell, D., Tomkins, A.: Information diffusion through blogspace. In: WWW, pp. 491–501 (2004)

9. Huang, Z., Zheng, Y., Cheng, R., Sun, Y., Mamoulis, N., Li, X.: Meta structure: Computing relevance in large heterogeneous information networks. In: KDD, pp. 1595–1604 (2016)
10. Jamali, M., Lakshmanan, L.V.S.: HeteroMF: recommendation in heterogeneous information networks using context dependent factor models. In: WWW, pp. 643–654 (2013)
11. Kang, U., Tsourakakis, C.E., Faloutsos, C.: Pegasus: A peta-scale graph mining system implementation and observations. In: ICDM, pp. 229–238 (2009)
12. Kong, X., Zhang, J., Yu, P.S.: Inferring anchor links across multiple heterogeneous social networks. In: CIKM, pp. 179–188 (2013)
13. Lao, N., Cohen, W.: Fast query execution for retrieval models based on path constrained random walks. In: KDD, pp. 881–888 (2010)
14. Lao, N., Cohen, W.W.: Relational retrieval using a combination of path-constrained random walks. Mach. Learn. **81**(2), 53–67 (2010)
15. Li, C., Yu, P.S., Zhao, L., Xie, Y., Lin, W.: InfoNetOLAPer: integrating InfoNetWarehouse and InfoNetCube with InfoNetOLAP. PVLDB **4**(12), 1422–1425 (2011)
16. 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)
17. Liang, J., Ajwani, D., Nicholson, P.K., Sala, A., Parthasarathy, S.: What links alice and bob?: Matching and ranking semantic patterns in heterogeneous networks. In: WWW, pp. 879–889 (2016)
18. Liben-Nowell, D., Kleinberg, J.: The link-prediction problem for social networks. J. Am. Soc. Inform. Sci. Technol. **58**(7), 1019–1031 (2007)
19. Liu, L., Tang, J., Han, J., Jiang, M., Yang, S.: Mining topic-level influence in heterogeneous networks. In: CIKM, pp. 199–208 (2010)
20. 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)
21. Liu, J., Shang, J., Wang, C., Ren, X., Han, J.: Mining quality phrases from massive text corpora. In: SIGMOD, pp. 1729–1744 (2015)
22. Long, B., Wu, X., Zhang, Z., Yu, P.S.: Unsupervised learning on k-partite graphs. In: KDD, pp. 317–326 (2006)
23. Long, B., Zhang, Z.M., Yu, P.S.: Co-clustering by block value decomposition. In: KDD, pp. 635–640 (2005)
24. Meng, X., Shi, C., Li, Y., Zhang, L., Wu, B.: Relevance measure in large-scale heterogeneous networks. In: APWeb, pp. 636–643 (2014)
25. Meng, C., Cheng, R., Maniu, S., Senellart, P., Zhang, W.: Discovering meta-paths in large heterogeneous information networks. In: WWW, pp. 754–764 (2015)
26. Ren, X., El-Kishky, A., Wang, C., Tao, F., Voss, C.R., Han, J.: ClusType: effective entity recognition and typing by relation phrase-based clustering. In: KDD, pp. 995–1004 (2015)
27. Shen, W., Han, J., Wang, J.: A probabilistic model for linking named entities in web text with heterogeneous information networks. In: SIGMOD, pp. 1199–1210 (2014)
28. 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)
29. 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)
30. 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)
31. 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)
32. 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)
33. 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)
34. Sun, Y., Yu, Y., Han, J.: Ranking-based clustering of heterogeneous information networks with star network schema. In: KDD, pp. 797–806 (2009)

35. Sun, Y., Tang, J., Han, J., Gupta, M., Zhao, B.: Community evolution detection in dynamic heterogeneous information networks. In: MLG, pp. 137–146 (2010)
36. 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)
37. 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)
38. Wang, C., Han, J., Jia, Y., Tang, J., Zhang, D., Yu, Y., Guo, J.: Mining advisor-advisee relationships from research publication networks. In: KDD, pp. 203–212 (2010)
39. Wang, C., Han, J., Li, Q., Li, X., Lin, W.P., Ji, H.: Learning hierarchical relationships among partially ordered objects with heterogeneous attributes and links. In: SDM, pp. 516–527 (2012)
40. 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)
41. 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)
42. Wang, C., Song, Y., Li, H., Zhang, M., Han, J.: Text classification with heterogeneous information network kernels. In: AAAI, pp. 2130–2136 (2016)
43. Wang, C., Song, Y., Roth, D., Zhang, M., Han, J.: World knowledge as indirect supervision for document clustering (2016). arXiv preprint. arXiv:1608.00104
44. 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)
45. Wang, G., Xie, S., Liu, B., Yu, P.S.: Identify online store review spammers via social review graph. TIST **3**(4), 61 (2012)
46. Wang, R., Shi, C., Yu, P.S., Wu, B.: Integrating clustering and ranking on hybrid heterogeneous information network. In: PAKDD, pp. 583–594 (2013)
47. Yin, X., Han, J., Yu, P.S.: Object distinction: distinguishing objects with identical names. In: ICDE, pp. 1242–1246 (2007)
48. Yin, X., Han, J., Yu, P.S.: Truth discovery with multiple conflicting information providers on the web. Knowl. Data Eng. **20**(6), 796–808 (2008)
49. Yin, M., Wu, B., Zeng, Z.: HMGraph OLAP: a novel framework for multi-dimensional heterogeneous network analysis. In: DOLAP, pp. 137–144 (2012)
50. Yu, X., Gu, Q., Zhou, M., Han, J.: Citation prediction in heterogeneous bibliographic networks. In: SDM, pp. 1119–1130 (2012)
51. 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)
52. 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)
53. 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)
54. Zhang, J., Kong, X., Yu, P.S.: Predicting social links for new users across aligned heterogeneous social networks. In: ICDM, pp. 1289–1294 (2013)
55. 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)
56. Zhang, J., Yu, P.S.: Integrated anchor and social link predictions across social networks. In: IJCAI, pp. 2125–2131 (2015)
57. Zhang, J., Yu, P.S.: Multiple anonymized social networks alignment. In: ICDM, pp. 599–608 (2015)
58. Zhao, P., Li, X., Xin, D., Han, J.: Graph cube: on warehousing and OLAP multidimensional networks. In: SIGMOD, pp. 853–864 (2011)
59. Zhao, B., Rubinstein, B.I.P., Gemmell, J., Han, J.: A bayesian approach to discovering truth from conflicting sources for data integration. PVLDB **5**(6), 550–561 (2012)