Journal of Network and Computer Applications 132 (2019) 86–103

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

journal homepage: www.elsevier.com/locate/jnca

Review

Academic social networks: Modeling, analysis, mining and applications

Xiangjie Kong, Yajie Shi, Shuo Yu, Jiaying Liu, Feng Xia *

Key Laboratory for Ubiquitous Network and Service Software of Liaoning Province, School of Software, Dalian University of Technology, Dalian, 116620, China

ARTICLE INFO

Keywords:
Academic social networks
Science of science
Scholarly data
Academic applications

ABSTRACT

In the fast-growing scholarly big data background, social network technologies have recently aroused widespread attention in academia and industry. The concept of academic social networks is created precisely in the context of scholarly big data, which refers to the complicated academic network formed by academic entities and their relationships. There are a wealth of scholarly big data processing methods to analyze the rich structural types and related information about academic social networks. Nowadays, various academic data can be easily obtained, which makes it easier for us to analyze and study academic social networks. This study investigates the background, the current status, and trends of academic social networks. We first elaborate on the concept of academic social networks and related research background. Secondly, we analyze models based on nodes' types and timeliness. Thirdly, we review analytical methods, including relevant metrics, network properties, and available academic analysis tools. Furthermore, we sort out some key mining technologies for academic social networks. Finally, we systematically review representative research tasks in this domain from three levels: actor, relationship, and network. In addition, some academic social networking sites are presented. This survey concludes with the current challenges and open issues.

1. Introduction

In the context of Web 2.0, a great deal of research has been carried out in the academia and industry, resulting in a great deal of academic information (Wu et al., 2014). Academic inputs and outputs have created unprecedented opportunities for studying the structure and evolution of science (Fortunato et al., 2018). With the rapid popularization and development of science and technology, the data are gradually shifting from the traditional storage mode to the digital one. Academic information is generated basically in the form of scientific documents, technical reports, project proposals, papers and other types of resources (Khan et al., 2016). In addition, academics and researchers from around the world can not only produce a large volume of academic documents but also share their research results through educational materials (Xia et al., 2017) such as patents and slides. The term of Scholarly Big Data (SBD) is generated by rapidly growing academic resources.

1.1. Scholarly big data

Due to the rapid growth of academic entities and their relationships, academic data has reached the "5V" characteristic of "Big Data", namely Volume, Velocity, Variety, Value, and Veracity (Wu et al., 2014), which is called Scholarly Big Data (SBD). It includes conference papers, journal articles, books, patents, slides and experimental data, etc (Williams et al., 2014b). Effective use of SBD is not only significant for scholars to understand scientific development and academic interactions, for policymakers to better resolve resource sharing issues, but also for enterprises to guide the development directions. Therefore, how to excavate valuable information from millions of SBD is a pressing issue.

The purpose of SBD analysis is to solve academic problems under the background of Science of Science (Light et al., 2014). The in-depth analysis of SBD can not only enable researchers to make more effective use of available resources but also contribute to the development of academia and industry. However, systematic research on this subject is insufficient. Previously, it is hard for researchers to achieve valid academic information because existing tools and technologies did not satisfy SBD analysis requirement. In addition, the high dimensions and large sizes of SBD pose certain challenges for data analysis (Fan et al., 2014). However, with the increasing popularity of the Internet and the development of relevant analytical techniques, we can now take full advantages of this valid information. A series of online digital libraries

* Corresponding author.
E-mail address: f.xia@ieee.org (F. Xia).

https://doi.org/10.1016/j.jnca.2019.01.029
Received 29 April 2018; Received in revised form 11 December 2018; Accepted 25 January 2019
Available online 5 February 2019
1084-8045/

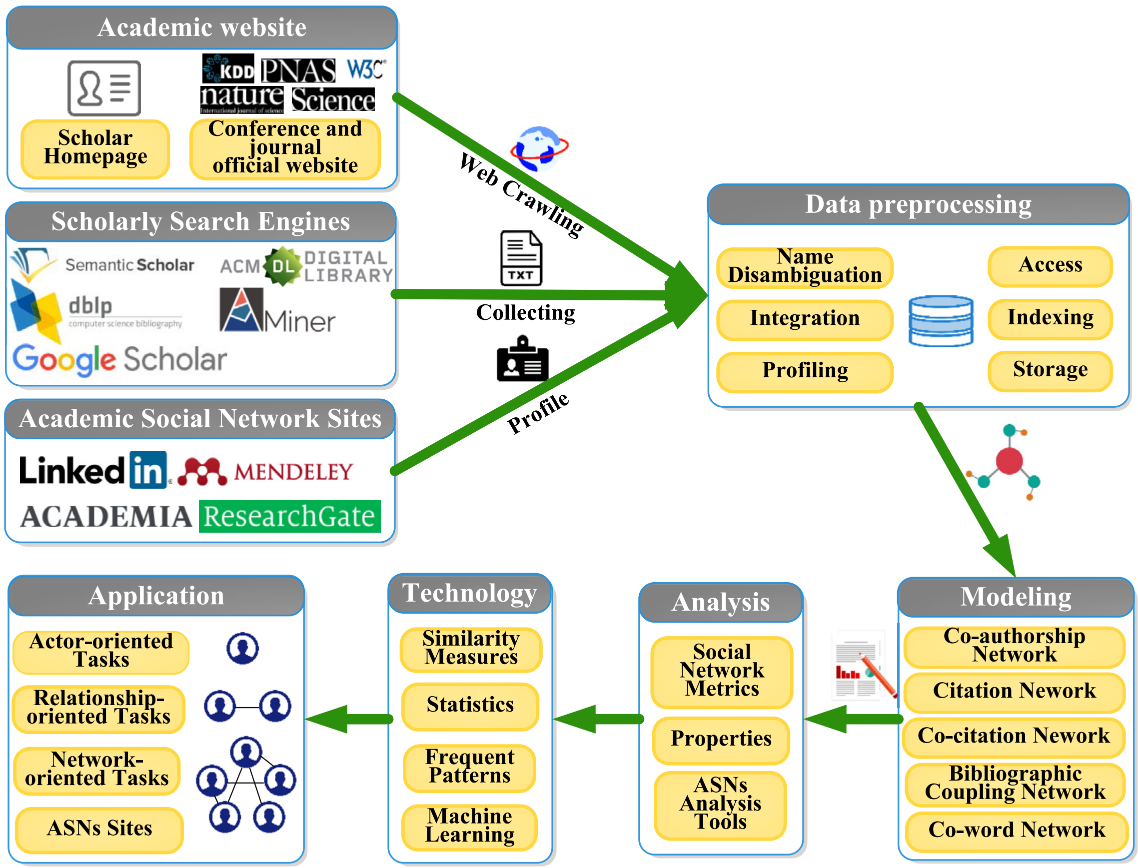

Fig. 1. Framework of academic social network survey.

and academic service platforms, for example, AMiner, Microsoft Academic Search (MAS), DBLP, Google Scholar (GS), and CiteSeerX, store millions of data on authors, publications, citations and other relevant information (Arif, 2015). SBD analysis can be divided into collaborator search, research management, expert discovery systems and recommendation systems (Khan et al., 2016).

1.2. Social networks

Social network analysis is a popular technique in recent years, exerting an increasingly important role in many fields, such as social media networks, transportation networks (e.g., traffic control), epidemiological networks (e.g., epidemics' spread modeling) and web networks (e.g., building the structure of the World Wide Web). It is used not only to analyze online social media applications such as Twitter and Facebook but also to provide integrated services in the area of scientific research. Social Networks (SNs) are collections of individuals or organizations that are interrelated in a particular situation like collaboration and socialization. In SNs, nodes and edges are used to represent entities and their interactions, respectively, to help us analyze and mine information. The analysis of SNs can identify the network relationship formed in the process of information dissemination.

The analysis method of SNs is an effective way to study SBD. In academic networks, researchers establish relationships through a variety of academic activities (Fu et al., 2014). At present, the research on different patterns of communication among various entities of SBD has attracted the great interests of researchers (Luo and Hsu, 2009). In addition, technological advances in data analysis, and recent developments in SNs' visualization software facilitate the research of these relationships as well as dynamic display (Luo and Hsu, 2009).

1.3. Social networks in scholarly data

Science of Science (SciSci) characterizes science as a complex, self-organizing and evolving network of academic information (Fortunato et al., 2018). In SBD, social networks formed through academic activities and information are called Academic Social Networks (ASNs). This expression can study ASNs from diverse geographical and temporal scales to characterize patterns of new scientific fields and accelerate the potential of science. There are many ways to establish ASNs, where co-authors are the most formal form of academic activities (Fu et al., 2014). By studying the citation networks, we can reveal the choices and trade-offs of researchers in their careers, and this is also one of the research topics in SciSci. In addition, some works have shown that well-connected academic social networks tend to be more prolific (Lopes et al., 2011), so they are imperative to study for us.

Currently, there are many surveys which use SNs in many fields, for example, Anomaly Detection (Kaur and Singh, 2016), Signed Network Mining in Social Media (Tang et al., 2016), Mobile Social Networks (Hu et al., 2015), Vehicular Social Networks (Rahim et al., 2017) and Social Influence in Social Networks (Peng et al., 2018) but there is no overview of SNs related to SBD. Meanwhile, there have been some surveys on SBD. Khan et al. (2017b) investigated the current research trends of scholarly data, identified the challenges for the development of academic data platforms and mapped future research directions to different phases of the life cycle of big data. Xia et al. (2017) conducted a comprehensive review of Big Scholarly Data from several aspects: scholarly data management, scholarly data analysis methods and representative research issues. At present, there is no study to review ASNs comprehensively.

In this work, we present a survey of the popular emerging ASNs field. To the best of our knowledge, this paper is the first to pro-

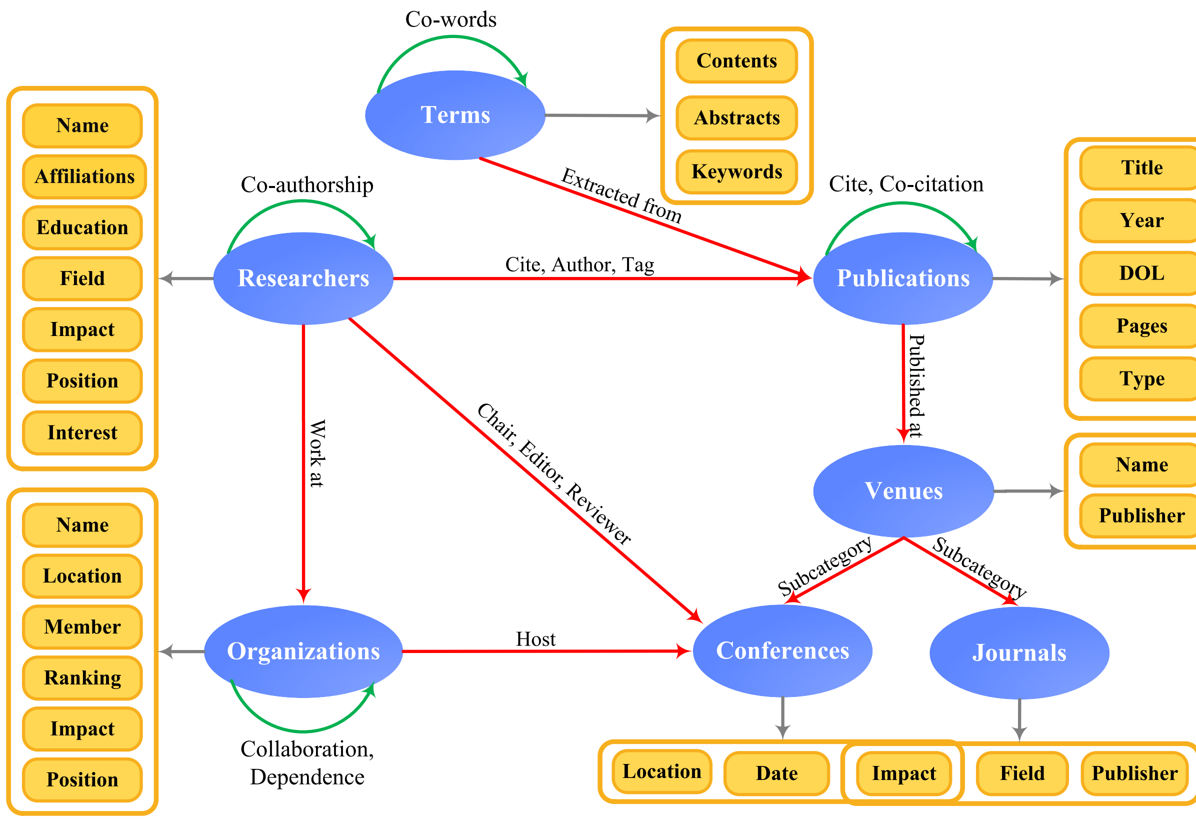

Fig. 2. Typical entities and their relationships.

vide a comprehensive review of SBD using SNs analysis. We systematically summarize the topics in ASNs from four perspectives: modeling, analysis, mining technologies and applications. In addition, we briefly present some useful ASNs tools and popular websites. Our aims are to fully interpret the current state of research in ASNs and to understand the opportunities and challenges of future research.

The framework of this paper has been illustrated in Fig. 1. Section 2 elaborates on the definition and properties of ASNs. Section 3 presents ways of modeling ASNs. Section 4 elaborates on ASNs analysis, and Section 5 covers some key mining technologies in ASNs. Section 6 describes some promising research applications and useful ASNs sites. Finally, Section 7 discusses critical open issues and challenging problems.

2. Academic social networks

In this section, we elaborate on the concept of Academic Social Networks, typical entities and their relationships, and available scholarly datasets.

2.1. Definition

Academic Social Networks (ASNs) are complex heterogeneous networks formed by a large number of entities (publications, scholars, etc.) and their relationships (citations, co-authorships, etc.) (Tang et al., 2008; Wu et al., 2014). Scholars have carried out plenty of research topics and data mining tasks. Here are some examples, author ranking (Amjad et al., 2015, 2017), author interests finding (Daud, 2012), rising star finding (Daud et al., 2013, 2015), academic recommendations (Guns and Rousseau, 2014) and community detection (Khan et al., 2017a). Attention over ASNs has led to many ASNs sites to provide SBD collection and analysis. For example, Microsoft Academic and Google Scholar provide paper searching, and CiteULike focuses on citation relationship services. Based on a variety of websites, we can easily get SBD information online.

2.2. Academic entities and relationships

Fig. 2 offers typical entities and relationships in ASNs. Nodes typically represent academic entities, including authors, publications, venues, institutions, and the terms (extracted from contents, abstracts or keywords of papers). Different types of entities have different attributes or labels that can help us analyze them more richly. Links between entities generally represent relationships, including co-authors, citations, co-citations, bibliographic couplings and co-words. Each type of relationship can form a different network, bringing a series of perspectives for research interaction and scholarly communications. Co-authors focus on finding communication patterns, bibliographic coupling, co-citation and co-word relationships which emphasize identifying research topics, whereas citation relationships pay more attention to the transfer of knowledge flows.

2.3. Academic semantic ontologies

Semantic publishing is a kind of journal publishing form with enhanced semantics (Shotton, 2009). It enriches the expression form and knowledge content of publications through Web and Semantic Web technology. It can also improve the operability, relevance and interaction of publication information, and ultimately achieve intelligent publishing. Ontology is a formal and detailed description of the shared conceptual system (Peroni and Shotton, 2012). Therefore, researchers can use ontology technologies to achieve the semantic description of document objects and their knowledge content, and then carry out rich research work. Table 1 briefly describes some commonly used ontologies.

Table 1
Basic characteristics of academic semantic ontologies.

Ontology	Description	Available Link
Bibliographic Ontology (BIBO)	Description of bibliographic resources metadata on the Semantic Web.	http://purl.org/ontology/bibo/
Semantic Web Applications in Neuroscience (SWAN)	The citation ontology module to define bibliographic resources.	http://swan.mindinformatics.org/spec/1.2/citations.html
FRBR-aligned Bibliographic Ontology (FaBiO)	Description of academic endeavors and references.	http://purl.org/spar/fabio/
Citation Typing Ontology (CiTO)	Characterization of citation relationships.	http://purl.org/spar/cito/
Bibliographic Reference Ontology (BiRO)	Defining bibliographic records, references, and compiling them into collections and lists, respectively.	http://purl.org/spar/biro/
Citation Counting and Context Characterization Ontology (C4O)	Allowing the number of text citations to the cited source to be recorded.	http://purl.org/spar/c4o/
Publishing Roles Ontology (PRO)	Describing the role of authors, editors, reviewers, publishers, etc. in the publication process.	http://purl.org/spar/pro/
Document Components Ontology (DoCO)	Providing a generic structured vocabulary for document elements, describing structural and rhetorical document components.	http://info.deepcarbon.net/schema/

Table 2
Basic characteristics of available academic datasets.

Dataset	Discipline	Size	Description	Available Link
Aminer	Computer Science	Over 2 million articles	The dataset contains the links between researchers, conferences and publications.	https://aminer.org/billboard/AMinerNetwork
APS	Physics	Over 450 thousand articles	This database contains the corpus of Physical Review Letters, Physical Review and Modern Physical Review.	http://journals.aps.org/datasets
DBLP	Computer Science	Over 2.3 million articles	All important journals on Computer Science are tracked.	https://dblp.uni-trier.de/
Microsoft Academic Graph (MAG)	Multidisci-plinary	167 million articles	The journal is classified according to SciMAGO Journal Classification into 27 different disciplines.	http://research.microsoft.com/en-us/projects/mag/
Open Academic Graph	Computer Science	64 million matching links	It is generated by linking MAG and AMiner.	https://www.openacademic.ai/oag/
Open Research Corpus	Computer Science and Neuroscience	Over 7 million articles	It contains information such as external citations, essay links, abstracts, titles, years, and more.	http://www.anc.org/

2.4. Available academic datasets

Currently, there are many search engines and digital libraries that provide their datasets to help researchers studying ASNs. Academic datasets are integrated academic documents that contain many types of general data. Many of them are freely downloadable, such as AMiner, American Physical Society (APS), DBLP, Microsoft Academic Graph (MAG), Open Academic Graph and Open Research Corpus. We list some basic characteristics and available URLs for these datasets in Table 2. We can obtain these entities from the bibliographic databases which contain metadata about the publications (e.g., authors, affiliations, pages, year), their citing publications (e.g., cited references, citation counts). Fig. 2 shows typical entities and their relationships.

3. Academic social networks modeling

Academic social networks can be constructed in various topological structures. The academic social behavior of scholars may change over time. In static networks, nodes never crash and edges maintain operational status. Scholars found that static networks can lead to a stable high level of collaboration (Rand et al., 2014). With the increasing scale of networked data, the structure of a network becomes more complex. Thus the computing time and complexity increase at the same time. Hence, Benson et al. (2016) used graphlet based on subnetworks and developed a generalized framework of higher-order connectivity patterns. Mostly, real-world networks are dynamic. In dynamic networks, nodes or edges may appear or disappear so that dynamic network topology changes over time. Dynamic networks are extensively used because they can describe both compositions and interactions (Rand et al., 2011). Another vital reason is that the ASN itself is dynamic. Plentiful researchers have gained significant results by exploring dynamic network structure. It is found that repeated positive interactions can promote collaboration between both individuals and within groups. However, dynamic ASNs are difficult in modeling since the topological structures are hard to be described.

Different kinds of networks are suitable for modeling different relationships. According to the differences of nodes in the network, ASNs can be classified into homogenous academic social networks and heterogeneous academic social networks.

3.1. Homogenous academic social networks

Homogenous ASNs refer to those networks whose nodes represent same entities. For example, in Fig. 3, the center of the figure is an example of the toy model of paper relationships and around it is several typical ASNs extracted from it. Fig. 3a is the co-authorship network in which X and Y co-author paper A and paper E, author Y and author Z co-author paper C. Fig. 3b is the citation network in which papers are connected by direct citation links. Papers published earlier are cited by papers published later, that is, arrows are drawn from earlier papers

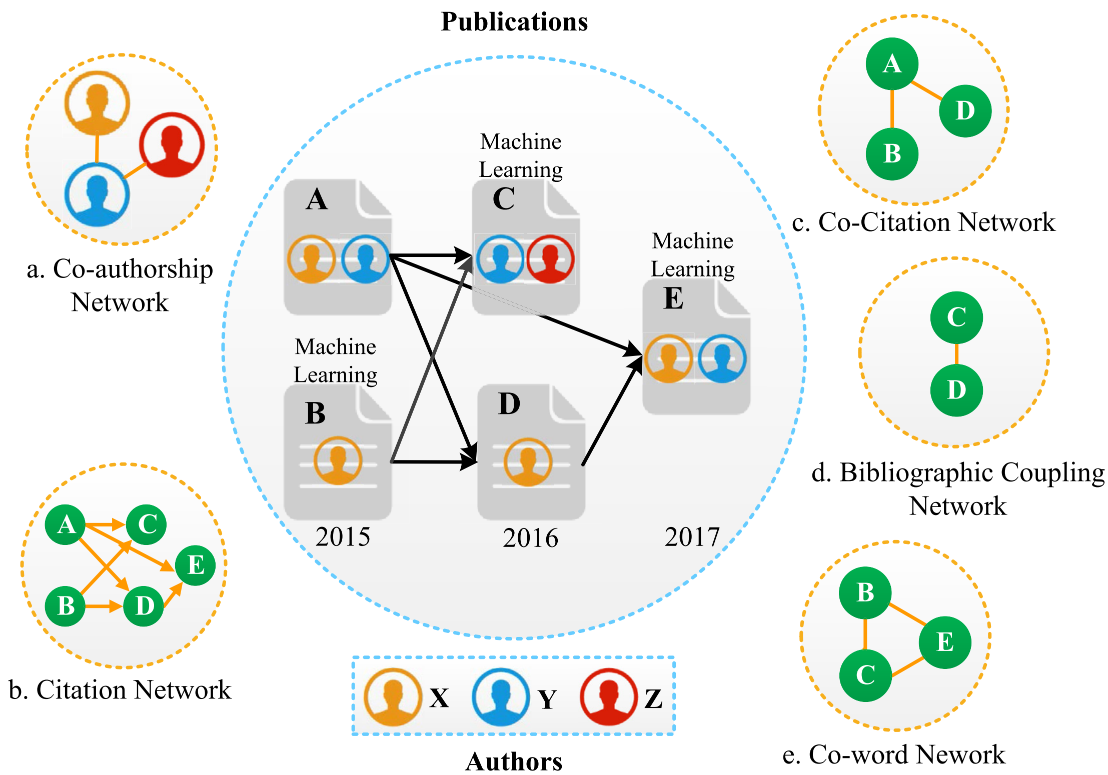

Fig. 3. Typical kinds of scholarly homogenous networks.

to latter ones. Fig. 3c is the co-citation network of the toy model in which A and B are co-citations by C and D, A and D are co-cited by E. Fig. 3d is the bibliographic coupling network. We can see C and D are bibliographically coupled as they both cite A and B. Fig. 3e is the co-word network while B, C, E all belong to the field of machine learning.

Co-authorship Networks. Co-authorship networks are one of the most widely used ASNs. In co-authorship networks of Fig. 3a, each node in the co-authorship network refers to an author. Edges in the co-authorship network refer to co-authored relationship. Scholars study co-author networks from various perspectives. It has been proved that collaboration continues to influence both the practice of research and the production of knowledge, becoming an increasing popularity among diverse disciplines (Uddin et al., 2013). Collaboration has become more and more common in nearly all disciplines. Besides, along with the development in information technology, transportation, and communication, scientists are no longer required to be physically co-located, and scientific collaboration may be conducted crossing university boundaries (Jones et al., 2008), even country boundaries (Wilsdon et al., 2011). Scholars study collaboration behaviors according to co-author networks. Furthermore, collaboration teamwork has been found to be a new research pattern.

Co-citation Networks. Co-citation is defined as two publications which are cited together in one article. Co-citation networks are constructed based on articles' citation relationships. Apparently, co-citation networks are directed networks since the two papers cannot cite each other at the same time. Scholars generate co-citation networks from publications and study scholars' behaviors from co-citation networks. Bai et al. (2016) studied co-citation networks and identified anomalous citation relationships. Actually, some academic social relationships may be not discovered through co-author networks, but can be discovered by co-citation networks. Co-citation analysis is one of the most commonly used bibliometric analysis methods. When two publications are frequently co-cited by the other articles, it is possible that the two references have something in common. As an advanced bibliographic technique, co-citation analysis is commonly used to discover the clusters of co-citation pairs, which enables scholars to obtain new insights for research trend. Although co-citation analysis has been claimed to be superior in displaying disciplinary structures to other bibliometric methods, it is still tough to provide a content profiling of the research topics dealing with the literature.

Co-word Networks. Co-word analysis has developed to address this kind of analytical problem (Leung et al., 2017). Co-word analysis is implemented based on the co-word network, which reflects the co-word frequency. The keyword co-occurrence frequency refers to the number of papers in which two keywords appear at the same time. By measuring the strengths of the keyword co-occurrence links, the co-word analysis reveals and visualizes the interactions between keywords. Since keywords are the terms used to verbalize the core of a research article, the co-word analysis is often used to explore the concept network of research topics and trends in a specific discipline. However, the co-word analysis also has its weakness instability due to term changes over time.

3.2. Heterogeneous academic social networks

Heterogeneous ASNs refer to the network whose nodes represent different entities. Fig. 4 shows an example of the heterogeneous network. In Fig. 4, nodes represent institutions, authors, publications, and venues, respectively. All of these entities are nodes within one network, which construct this heterogeneous network. Heterogeneous ASNs are widely used to analyze complex social connections between different academic entities. In most current research on network science, social and information networks are usually assumed to be homogeneous, where nodes are objects of the same entity type (e.g., scholar) and the links are relationships of the same relation type (e.g., co-authorship). Interesting results have been generated from such studies with numerous influential applications like community detection methods. How-

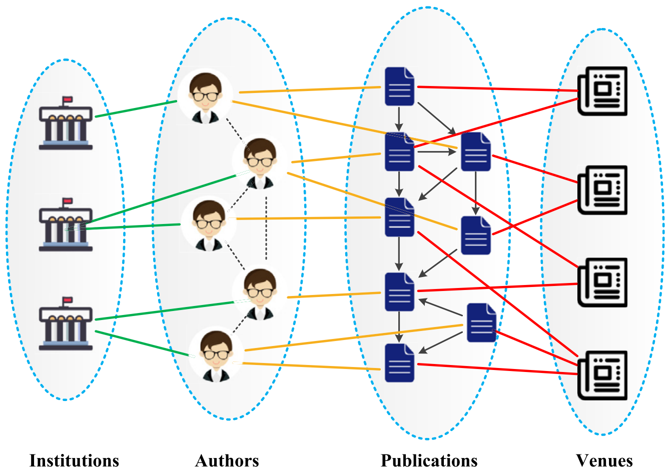

Fig. 4. An example of the heterogeneous network.

ever, most ASNs are heterogeneous, where nodes and relations are of different> types.

Paper-author Networks. Paper-author networks are constructed in order to analyze the relationships between papers and scholars (Sun and Han, 2013). Many paper-author networks are constructed in order to recommend proper papers for target scholars. Generally, scholars extract information from the bibliographic database and construct paper-author networks. These networks contain heterogeneous information, including articles, authors, co-citations, etc. By analyzing paper-author networks, both explicit and implicit citation relationships can be explored.

Bibliographic Coupling Networks. Bibliographic coupling is a widely applied approach, which is used for grouping technical and scientific papers. When two articles cite one same reference, it is defined as a basic unit of coupling between two papers. The coupling strength between two articles is measured based on the number of coupling units, which means that when two articles cite the same references, these two articles are related to some extent. The strength of this association is determined by the frequency of coupling. Studies on citation network are mainly focusing on citation of academic bibliographies. In order to explore the development and changes in the field of information science, Huang et al. (2003) measured the association between citations according to bibliographic coupling and then cluster these citations. Börner et al. (2003) made efforts in the same area by exploring the development and trends through co-citation analysis. Scholars can use the co-citation relationships to group cited literatures into clusters to study the bibliographic citations and the relationship among these clusters.

Hybrid Networks. Hybrid approaches are widely used in identifying research topics. Liu et al. (2010) presented a framework of hybrid clustering in order to combine lexical and citation data for journal analysis. Zitt et al. (2011) examined the convergence of two thematic mapping approaches, i.e., citation-based and word-based. Boyack and Klavans (2010) examined several types of scholarly networks, including a co-citation network, a bibliographic coupling network, and a citation network, which aimed at selecting the network that can best represent the research trend in biomedicine. Janssens et al. (2009) proposed a novel hybrid approach that integrates two types of information, which are citation (in the form of a term-by-document matrix) and text (in the form of a cited-references-by-document matrix), respectively.

4. Academic social networks analysis

The model for ASNs is used to represent the network, while metrics are mainly used to analyze them. In this section, we sort out some of the network metrics and popular metrics used in SNs analysis triggered by ASNs. In addition, using the common properties of social networks also helps us to understand more about academic social networks.

4.1. Social network metrics

In this part, we briefly hackle some general social network metrics. These metrics give us insights into the network structure without having to know its graphical representation. Exploring the structure of these networks aims to understand the behaviors of social systems that generate these academic networks, which is often the ultimate goal of such analysis.

4.1.1. Global metrics

There are many metrics to explore entire networks' attributes.

Diameter. In the network, the distance d_{ij} between node i and node j denotes the number of edges that connect the shortest path between these two nodes. The diameter D refers to the maximum eccentricity of the network, which describes the maximum distance (Yan et al., 2010). It is expressed by Eq. (1):

$$D = \max_{i,j} d_{ij} \tag{1}$$

Density. The density refers to measuring the connectivity of a global network, which is calculated by dividing the total number of connections present by the total number of possible connections with the same number of nodes, as defined by Eq. (2):

$$\rho = \frac{E}{E_{\max}} \tag{2}$$

where E is the number of the network's edges and $E_{\max}$ refers to the number of possible edges with the same nodes. $E_{\max}$ is $n(n-1)$ for directed networks and $n(n-1)/2$ for undirected networks.

Average Shortest Path Length. The average shortest path length L of the network is the mean length of the shortest path between any two nodes (Yan et al., 2010) and is expressed as Eq. (3):

$$L = \frac{2}{N(N-1)} \sum_{i \neq j} d_{ij} \tag{3}$$

Harmonic Average Shortest Path Length. When a network has multiple connected components, the previous formula does not hold since the metric is usually defined as infinity when there is no path connecting two nodes. In this case, we can use harmonic average shortest path length, as shown in Eq. (4), which counteracts their influence on the sum as soon as it turns an infinite distance into zero:

$$L^{-1} = \frac{2}{N(N-1)} \sum_{i \neq j} \frac{1}{d_{ij}} \tag{4}$$

Average Degree. The average degree $\langle k \rangle$ of the network is the average of the degrees k_i for all nodes, as shown in Eq. (5):

$$\langle k \rangle = \sum_{i \epsilon V} \frac{k_i}{N} \tag{5}$$

It is used to reflect the global connectivity of networks.

4.1.2. Community metrics

There are many methods for researchers to identify the communities in the network. Here we present the two most classic metrics.

Core. The metric of core can identify the groups which are closely interconnected in the whole network. K-Core is the largest entity group and its all nodes are connected to at least the other k nodes in this group. K-Core helps identify smaller interconnected core areas. The linked nodes are independent of the other nodes that they may connect to outside the group. The value of k is sometimes called the core of

Journal of Network and Computer Applications 132 (2019) 86–103

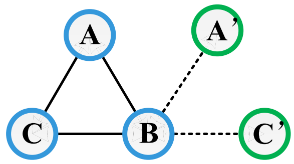

Fig. 5. An example of closed and open triplets.

a group. For example, if all nodes are associated with at least two other nodes in the group, this group is a 2-Core group.

Clique. The metric of clique is defined as the largest set of nodes that all nodes are directly adjacent to others. In ASNs, we can know that a clique is a group of authors, all of whom write papers with all the other authors. Thus, the largest cliques will be identified as those who write special articles with many co-authors.

4.1.3. Node degree

The degree of a node is defined as the number of its neighboring nodes and the formula is shown in Eq. (6):

$$k_i = \sum_{i \epsilon V} a_{ij} \tag{6}$$

4.1.4. Clustering coefficient

A cluster means that the collaboration exists between any two scholars and the construction of a toy triplet is shown in Fig. 5. Thus, if node A is connected to node B and node B is connected to node C, the probability that node A will also be connected to node C is increased. The clustering coefficient is divided into local values and global values.

Local Clustering Coefficient. The local clustering coefficient indicates the level of cohesion in the neighborhood of a node. A commonly used method for calculating local clustering coefficients is shown in Eq. (7):

$$C_i = \frac{2E_i}{k_i(k_i - 1)} \tag{7}$$

where E_i represents the number of edges for node i.

Global Clustering Coefficient. The global clustering coefficient measures the clusters of the whole network and can be used in both directed and undirected networks, but not in weighted networks. The formula is shown in Eq. (8):

$$C = \frac{\sum_i C_i}{n} \tag{8}$$

4.1.5. Centrality

Degree, closeness, betweenness and eigenvector centrality are basic approaches for calculating nodes' centrality (Wasserman and Faust, 1994). A toy model is shown in Fig. 6, where the most important nodes are marked according to these centralities. The metric of PageRank is transplanted from the Web page rankings. We will describe these five different metrics in detail.

Degree Centrality. It is the simplest of all centralities, which is calculated corresponding to the number of neighbors of a node. It represents the interconnectedness of network nodes, reflecting the nodes' communication activities. It is often calculated by dividing the degree of a node by n - 1, limiting the value in the range of [0, 1]. It is calculated by Eq. (9):

$$C_D(i) = \frac{k_i}{N - 1} \tag{9}$$

Closeness Centrality. It is used to measure the average length of the shortest path from one node to all other nodes, presented as Eq. (10):

$$C_C(i) = \frac{N - 1}{\sum_{j \neq i}^{N} d_{ij}} \tag{10}$$

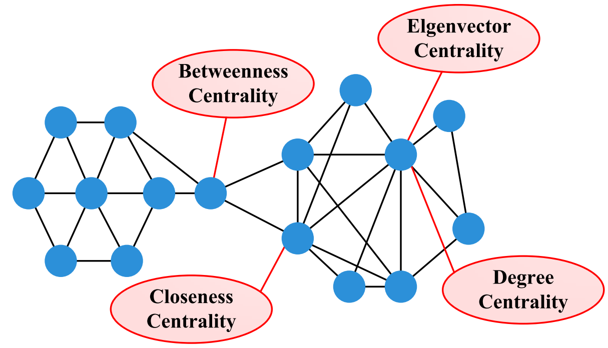

Fig. 6. An example of the co-authorship network.

It is generally used in the largest component of the network and considers all the other nodes in connected networks. In ASNs, closeness centrality is an indicator of reachability that reflects the time from one node to another.

Betweenness Centrality. Betweenness centrality is used to describe the extent of nodes that must be gone through in order to reach other nodes. In ASNs, nodes with high betweenness centrality play a key role because they act as a bridge between scholars and control the flow of information in the network to some extent. It is expressed as Eq. (11):

$$C_B(i) = \sum_{s \neq i \neq t \epsilon V, s < t} \frac{\sigma_{st}(i)}{\sigma_{st}} \tag{11}$$

where $\sigma_{st}(i)$ represents the number of shortest paths between nodes s and t through node i, and σ_{st} is the number of shortest paths between nodes s and t.

Eigenvector Centrality. Eigenvector centrality is another measure of reflecting the node's importance. It gives the relative scores of all the nodes according to the principle that the nodes connected to the high scores contribute more to the scores of the nodes rather than the ones of the low scores. It is calculated by using the adjacency matrix in Eq. (12):

$$C_E(i) = \frac{1}{\lambda} \sum_j A_{ij} C_E(j) \tag{12}$$

where A_{ij} denotes ith eigenvector of the adjacency matrix in the network.

PageRank. PageRank was originally used for Google's web pages core ranking mechanism (Page et al., 1999; Brin and Page, 2012). It finds important nodes by calculating the weight of the nodes that use out-degree links, which means that other nodes linked by these nodes also have higher page rankings. PageRank is calculated by Eq. (13):

$$PR(r) = \frac{1 - \lambda}{N} + \lambda \sum_{i=1}^{k} \frac{PR(r_i)}{K_{out}(r_i)} \tag{13}$$

where N represents the total number of nodes in the network, K_{out} is the out-degree of the node r, r_i denotes the in-degree of node r and λ is the damping factor.

4.2. Properties

For ASNs, there are some common properties of these networks.

4.2.1. Power-law degree distribution

Degree distribution refers to the probability distribution of nodes degree in the whole network, which is denoted by $P(k)$. In random networks, degree distribution is highly homogeneous since the existence of each edge is equiprobable. Unlike random networks, Barabási and Albert (1999) found that the degree distribution of nodes in the real network is heterogeneous, such as citation networks, in which most

Table 3
Basic information of major academic social network analysis tools.

Tool	Platform	Language	Access	Description
CiteSpace	Windows/iOS/Linux	Java	Free	Analyzing and visualizing the patterns and trends of scholarly publications.
CitNetExplorer	Windows/Others	Java	Free	Specifically analyzing and visualizing citation networks.
Gephi	Windows/iOS/Linux	Java	Free	An open source network tool for analysis and visualization.
HistCite	Windows	java	Free	Literature analysis and information visualization.
iGraph	Windows/iOS/Linux	C/R/Python	Free	A network analysis and visualization tool.
NetworkX	Windows/iOS	Python	Free	A highly portable tool for large real-world networks.
NodeXL	Windows	C#/.NET	Free	A network analysis package for Microsoft Excel.
Pajek	Windows/iOS/Linux	C/R	Free	A large, complex network analysis tool.
Sci2	Windows/iOS/Linux	Built on CIShell	Free	Used for scientific research, supporting temporal, geospatial, subject and network analysis.
UCINet	Windows/iOS	Java	Purchase	Analyzing social networks.
VOSviewer	Windows/Others	Java	Free	Analyzing and visualizing bibliographic networks.

nodes have lower degrees and very few nodes have higher degrees. They pointed out that the degree distribution of real networks follows a power law distribution approximately. Sometimes, we call these networks as the scale-free networks (Barabási and Bonabeau, 2003). Furthermore, they groundbreakingly designed a network growth model called the Barabási-Albert model, which is the process we know about preferential attachment.

4.2.2. Small-world property

Travers and Milgram (1977) pointed out the small-world properties of the real social networks through the well-known Milgram experiment and explained the concept of the "Six Degrees of Separation" explicitly. It has now been proven that the small-world properties are universal in real worlds. In SNs, the small-world property signifies that the average shortest path length between nodes is proportional to a fixed average network size (Newman, 2003b). In ASNs, this property shows that two uncooperative researchers can be contacted through a series of their collaborators.

4.2.3. Mixing patterns

There are different types of nodes in many networks. In these networks, node-to-node links tend to be selective and highly correlated with the type of nodes (e.g., papers, authors and venues). The real networks tend to show a higher tendency to mix. Newman (2003a) proposed a metric called assortative coefficient to measure the extent of confusion in a network.

4.2.4. Community structure

Due to the heterogeneity of edge distribution in the network, the property of the community structure is generated. The community can be defined as a similar group of the nodes(Newman and Girvan, 2004). Therefore, we usually find a high density of edges in certain areas of a network and lower density of edges between those areas. Most real networks exhibit the characteristics of community structures. In ASNs, we can classify scholars into communities according to their similarities and analyze the relationship between communities.

4.3. Academic social network analysis tools

Due to the large volume of SBD and various types of ASNs, manual processing takes too much time and effort. We can use analysis tools to build, analyze, and visualize ASNs by using network features. The functions of ASNs analysis tools are versatile, for example, network characterization, relational mining, community detection, and visualization. Now, there are plenty of tools, for example, CiteSpace (Chen, 2004), CitNetExplorer (Van Eck and Waltman, 2014), Gephi (Bastian et al., 2009), HistCite (Garfield and Pudovkin, 2004), iGraph (Csardi and Nepusz, 2006), NetworkX (Hagberg et al., 2008), NodeXL (Smith et al., 2009), Pajek (Batagelj and Mrvar, 2004), Sci2 (Team, 2009), UCINet (Borgatti et al., 2002), VOSviewer (Van Eck and Waltman, 2011). Here, we briefly describe the information about these tools in Table 3.

5. Key mining technologies

Academic social network mining is an important sub-task of academic social network analysis. It aims to mine academic social relations, which can be regarded as the association between entities. Recent trends have focused on mining interrelationships with the advent of both available data and developing technologies. From data mining and SNs mining perspectives, we classify measures of ASNs mining into the following four categories:

- Similarity measure
- Statistical relational learning
- Graph mining
- Machine learning

These techniques usually work on the relation extraction in social networks and could be implemented in academic social networks depending on the network affiliation. More specifically, they can easily be implemented in community detection, relation mining, and link prediction. Fig. 7 provides the main classification and their applications in academic social networks of these methods.

5.1. Similarity measures

Similarity measures are used to measure the extent of similarity between entities in networks. The concerns over current similarity models in multi-relational social networks fall into two areas: content-based algorithms and linkage-based and structural algorithms.

5.1.1. Linkage-based and structural methods

Most existing studies on network mining are based on classical algorithms, PageRank (Page et al., 1999) and SimRank (Jeh and Widom, 2002), which use random walk techniques in the ranking process. The aim of the random walk approach is to estimate the probability of passing each node. Therefore, a well-connected structure will have a higher PageRank score in networks. These algorithms and their applications are widely used in collaboration patterns mining and recommendation (Strohman et al., 2007; Lü and Zhou, 2011; Li et al., 2014), rising star finding (Daud et al., 2013; Li et al., 2009; Zhang et al., 2016), and scientific evaluation (Chen et al., 2007; Zhou et al., 2012; Wang et al., 2013b) in ASNs.

5.1.2. Content-based methods

In the case of content-based analysis, the most well-known methods are distance-based algorithms, cosine-based algorithms, correlation-based algorithms, and Jaccard coefficient. Minkowski distance, Chebyshev distance, Manhattan distance, and Euclidean distance are widely

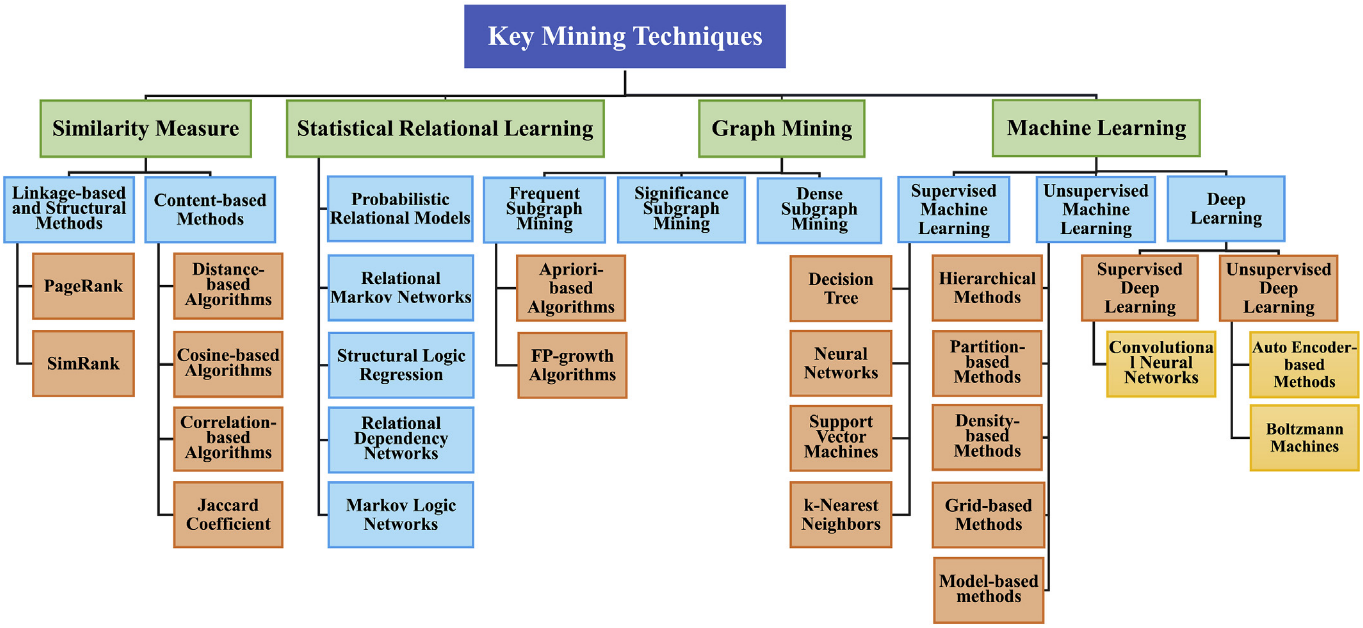

Fig. 7. Key mining technologies for ASNs mining.

used to calculate distance similarities and cluster entities with common characteristics into various groups. These methods have emerged as an appealing way to use information about an item itself and make suggestions in the scholarly recommendation as well as analysis (Ding et al., 2014).

5.2. *Statistical relational learning*

Travers and Milgram (1977) have laid the foundation for extensive analysis of structural properties in large-scale networks. Statistical properties show differently in various typical social networks. By exploring characteristics of these networks on a large scale, we can deeply study the examination of connectivity behavior among nodes and how the structure varies with the evolution of the network.

Statistical relational learning combines theories of statistical methods with data representation. It concentrates on the joint probability distribution of the data. Studies on statistical relational learning at this stage mainly focus on the challenges posed by learning probabilistic models in relational data. In particular, researchers have proposed three features including concentrated linkage, degree disparity, and relational autocorrelation, to build up models for statistical relational learning. Concentrated linage can represent significant inconsistencies in connectivity across different types of objects. Degree disparity is produced by the different degree of different types of objects. Relational autocorrelation is the relevant value of the same attribute in different affiliates.

Founded on above mentioned features, typical models include Probabilistic Relational Models (PRM) (Koller, 1999), Relational Markov Networks (RMN) (Taskar et al., 2007), Structural Logic Regression (SLR) (Popescul and Ungar, 2003), Relational Dependency Networks (RDM) (Neville and Jensen, 2007), and Markov Logic Networks (MLN) (Richardson and Domingos, 2006). These statistical learning models are established for relational data, which can be very useful for describing the network and accomplishing analytical tasks.

5.3. *Graph mining*

Graph mining, which is an important aspect of network mining, has become increasingly attractive recently. It refers to the process of using graph models to discover useful knowledge and information from massive data. In ASNs, it can be applied to various aspects, such as link analysis, group detection, metadata mining.

Frequent subgraph mining is an active research topic in graph mining. Related methods include Apriori-based algorithms (i.e., AGM (Abramowitz et al., 1966), ACGM, path-join (Li et al., 2001). etc.) and FP-growth algorithms (i.e., gSpan (Yan and Han, 2002), CloseGraph (Yan and Han, 2003), FFSM (Huan et al., 2003). etc.). They gradually expand their frequency to get frequent subgraphs with slightly different extensions of edges. Furthermore, Wang et al. (2005) proposed an index-based frequent subgraph mining algorithm GraphMiner for mining frequent patterns from large disk-based graph databases. For dynamic graph mining, Borgwardt et al. (2006) designed DynamicGREW for extant pattern mining on static graphs for time series of graphs. Other graph mining algorithms involve significant subgraph mining (Sugiyama et al., 2015) and dense subgraph mining (Gunnemann et al., 2010).

These algorithms are usually used to extract and analyze relationships among entities in ASNs, for example, community mining, scientific patterns unveiling (He et al., 2012), and impact prediction (Pobiedina and Ichise, 2014) in detail.

5.4. *Machine learning*

The major tasks involved in machine learning approaches of academic social network mining are supervised techniques and unsupervised measurements. In the supervised domain, the relation task aims to extract a set of known relations based on mentions of the entity pair. It requires a large amount of training data for learning. On the contrary, unsupervised tasks focus on predicting which relationship class of a given entity is beyond the given labels. Deep learning is also an important branch of machine learning. Thus, we classify the key mining techniques into the following categories: supervised learning, unsupervised learning, and deep learning.

5.4.1. *Supervised learning*

Traditional supervised learning methods can be divided into two categories: feature-based and kernel-based methods. Based on related theories, mining tasks in ASNs are generally identified as a classification problem. There are too many classification models for supervised learning, i.e., Decision Tree, Neural Networks, Support Vector Machines, and K-Nearest Neighbors. Regression models (Herlocker et al., 1999) can also be used for relational classification. The performance of these methods relies on feature selection and parameter settings. For example, Akritidis and Bozanis (2013) solved the problem of the automatic paper classification by introducing machine learning algorithms. They associated authors, co-authors, keywords and published journals with many labels of the taxonomy. XGboost (Chen and Guestrin, 2016) is a machine learning model proposed in recent years. Its prediction ability is better than that of support vector machine and neural network in some respects. Its disadvantage is that it needs a lot of adjustment parameters. Bai et al. (2017b) used the XGboost model to synthesize the features of individual ability, institutional location and national GDP to predict the impact of institutions.

5.4.2. *Unsupervised learning*

In practical applications, the process of supervised learning often requires accurate labels, which will take huge time and effort to generate datasets. Learning without labeling a dataset is called unsupervised learning, which is often considered as a clustering problem in the relation extraction. A number of techniques have been developed to identify similar entities. Up to now, there are five general clustering algorithms, including hierarchical methods, partition-based methods, density-based methods, grid-based methods, and model-based methods, in which the most commonly used are hierarchical methods and partition-based methods (Han et al., 2011). Partition-based algorithms divide the dataset into k parts by optimizing the evaluation function, which needs researchers to decide the value of k as input. Typical algorithms in partition-based methods are K-means, K-medoids (Park and Jun, 2009), and CLARANS (Ng and Han, 2002).

Hierarchical methods consist of different levels of segmentation clustering and the segmentation between levels have a nested relationship. It does not require input parameters, which is one of the obvious advantages compared with Partition-based methods. However, researchers need to specify the termination condition. Typical hierarchical algorithms include BIRCH (Zhang et al., 1996), DBSCAN (Birant and Kut, 2007) and CURE (Guha et al., 1998).

Although the performance of unsupervised learning is less effective than supervised learning, it can pick representative samples out of the large dataset for classifier training, and help in classification in terms of dividing samples into different categories which can be labeled by manual annotation.

5.4.3. *Deep learning*

Deep learning is one of the most important machine learning methods based on data representation. Same as machine learning methods, it also consists of supervised learning and unsupervised learning. Typical algorithm under supervised learning is Convolutional Neural Networks (CNN) (Krizhevsky et al., 2012). Unsupervised deep learning is mainly divided into two categories. One is based on the autoencoder (Rifai et al., 2011), of which the main goal is to restore the original data from the abstract data in one piece. Others are on the basis of Boltzmann machines (Aarts and Korst, 1988). The main goal of these algorithms is to reproduce the original data when the machine reaches a steady state.

6. Applications

ASNs have extensive applications in the related research of academic interaction and communication. We do not focus on the algorithm, but emphasize some of the existing new ideas and methods. In this paper, we briefly present these applications from three aspects: actor-oriented applications, relationship-oriented applications, network-oriented applications, as shown in Fig. 8. Finally, we summarize some helpful ASNs sites.

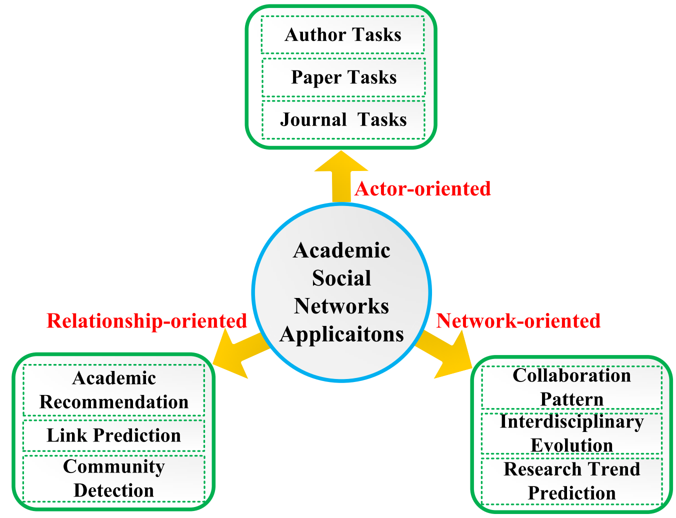

Fig. 8. Applications of academic social networks.

6.1. *Actor-oriented applications*

Actor-oriented applications include many aspects and we elaborate on them from the author, paper, and journal three aspects.

6.1.1. *Author-level tasks*

The author is an important entity in ASNs and the relationships of authors are complex. We review some of the popular studies in recent years.

Author Name Disambiguation. Author name disambiguation is still a difficult problem in literature retrieval and data processing (Tang and Walsh, 2010). A single name may signify several distinct authors, and an individual author may have multiple names searched. Some studies have shown that author name ambiguity may result in strongly bias for network attributes like the clustering coefficient (Kim and Diesner, 2016). When we extract author information, we need to pre-process to eliminate the inherent ambiguity which is associated with authors' names. Basically, there are three ways to eliminate ambiguous names: algorithm-based, first-initial, and all-initial methods (Khan et al., 2016). Treeratpituk and Giles (2009) used the Random Forest Model to disambiguate names by considering the author's name, affiliation, collaborator, and related factors. Based on the hypothesis that one author can be identified by his co-authors, Kang et al. (2009) proposed a way to eliminate ambiguity by implicit co-authors of target authors. In addition, Kim et al. (2014) used the three above-mentioned methods for the DBLP dataset to eliminate ambiguity and they found that author name disambiguation have a significant effect on analyzing data effectively. Cota et al. (2010) proposed a method of disambiguating names based on the similarity of citation information (e.g., titles) in the co-authorship network.

Author Ranking. The methods of author ranking based on network link structure can be divided into two categories: iterative method and non-iterative method. Firstly, the iterative method executes instructions iteratively until the algorithm converges. PageRank (Brin and Page, 2012; Page et al., 1999) and HITS (Kleinberg, 1999) algorithms are two basic iterative algorithms, and a lot of author ranking methods are based on these two algorithmic ideas. Fiala et al. (2008) used the modified PageRank to achieve scholar rankings in the bibliographic network. Ding et al. (2009); Yan and Ding (2011) considered the weight of edges in the network and proposed a weighted PageRank method for studying citation networks and co-authorship networks. Li and Tang (2008) considered time statistics and combined social networks with the random walk model. Radicchi et al. (2009) proposed a weighted PageRank approach by considering the credit diffusion of authors. Ding (2011) introduced the topic weights to present a topic-sensitive extension of the PageRank algorithm. Amjad et al. (2015) used a topic-based model in heterogeneous academic networks. In addition, researchers continue to enrich the author ranking methods by considering topological features. It is generally believed that if a scholar is situated at a certain key position in ASNs, he/she can be recognized to be important. Chiang et al. (2013) used social relationships and local information to find the top-k authors in the co-authorship network based on the probabilistic model of the random walk.

Expert Finding. The information retrieval subject has received more and more attention, and the search for professionals in a particular field is called Expert Finding (Serdyukov, 2009). Expert Finding is a concept that focuses on the organization and is targeted at identifying people with relevant expertise or rich experience. Basically, there are two kinds of methods: content-based method (Chen et al., 2013), which focuses on assessing a scholar's expertise by measuring the correlation between the related documents and the query, and SNs-based method (Kardan et al., 2012), which is more concerned with scholars' social interaction in ASNs, such as co-authorships and citation relationships. We highlight the latter method that is relevant to the subject of our article. Noll et al. (2009) proposed a HIT-based approach that links users and documents by assigning different weights. To integrate heterogeneous information, Nie et al. (2005) proposed the PopRank model, which showed that the author's score comes from the combined score of different types of objects. Sun et al. (2009) introduced a ranking-based clustering method in the heterogeneous network, which targets clustering and sorting objects in the clusters simultaneously. Deng et al. (2012) proposed a joint regularization framework that enhances expert finding by modeling heterogeneous networks as document-centric models of regularization constraints. Yang et al. (2013) provided scholars with information on individual social networks, research relevance, and organizational connectivity for expert recommending, in order to expand the scope of research as well as improve the specificity of the recommendation.

Rising Star Finding. Rising stars are scholars with specialized knowledge and abilities, and may gain a high reputation in their related fields in the near future (Daud et al., 2013). Searching for a rising star in a particular field is a new research direction in recent years which may make it possible for research teams to emphasize valuable and potential researchers. This idea was originally calculated by PubRank (Li et al., 2009) which only contains the static ranking of venues and the interacted influence between author and papers. The StarRank (Daud et al., 2013) method enhances the reliability of the PubRank method by considering the dynamic publication ranking. Later, Daud et al. (2015) proposed a machine learning method based on a combination of publications, co-authors, and venues to find rising stars. Wijegunawardana et al. (2016) found the rising star from multiple data sources using a combination of multi-target methods and rank-aggregation methods. Panagopoulos et al. (2017) proposed a method based on unsupervised clustering machine learning that considers all key performance indicators (KPIs) to identify rising stars. In addition, Zhang et al. (2016) proposed the CocaRank method by integrating the newly defined indicators called collaboration caliber, citation counts, and hybrid results, to find rising stars in ASNs. Their method can find more top rising stars with the higher average number of citations.

6.1.2. Paper-level tasks

For the applications of the paper, scholars are interested in the paper impact evaluation and prediction. The paper impact is significant to assess the effect of scholars, journals, institutions and even countries (Bai et al., 2017a). It is also vital for the assistance of rewards, promotions, and recruitment. Therefore, the research on evaluating and predicting paper impact has not slowed down during the past few decades.

Paper Impact Evaluation. Traditionally, the number of papers' citations (Lehmann et al., 2006), impact factor (Timilsina et al., 2016) and h-index (Hirsch, 2005) are widely used to measure scientific impact of individual papers. Most of these studies assess paper impact based on the PageRank algorithm and the HITS algorithm (Chen et al., 2007; Zhou et al., 2012; Wang et al., 2013b). For example, Ding et al. (2009) firstly applied the PageRank algorithm on the Co-citation network to rank paper impact. Network-based ranking methods sprung up to evaluate the impact of the paper in recent years. Zhou et al. (2012) proposed the MutualRank algorithm to evaluate the impact by jointly ranking papers, authors, and venue information. Wang et al. (2013b) proposed the CAJTRank approach by mining authors, references, periodicals, and time information. In order to further improve the validity of the assessment, Yao et al. (2014) proposed a non-linear PageRank method in which the high citing articles are favored and the low-citing articles are penalized.

Paper Impact Prediction. Paper Impact Prediction is as important as Paper Impact Evaluation. Much work has already been done to predict the number of future academic citations. Yan et al. (2011) studied a series of important features for future citations. Wang et al. (2013a); Shen et al. (2014) revealed the fundamental mechanisms that dominate scientific impact, which can further quantify and predict the number of future citations. However, citation distributions can seriously affect the validity of predictions. Sayyadi and Getoor (2009) proposed the FutureRank method to predict the future impact of papers, by considering the author's relationship, citation and publication information comprehensively. Yao et al. (2014) proposed a MRFRank algorithm to predict the future impact of the papers based on the co-authorship network, the weighted time-aware citation network as well as the textual features. In addition, early citations of papers have a significant impact for predicting the long-term citations (Bruns and Stern, 2016). To avoid over-reliance on historical citation data, social media activities are used to reflect the potential impact of papers. Eysenbach (2011) used Tweets to predict whether a paper can be cited frequently during the first 30 days after publication. Timilsina et al. (2016) predicted the paper impact by combining bibliometric data with social media data, demonstrating that graph-based approaches can effectively predict the scholarly impact.

6.1.3. Journal-level tasks

Compared to the previous two tasks, there are few studies on the journal-level tasks. The researches of journal-level tasks mainly focus on the journal impact evaluation.

Journal Impact Evaluation. High-quality journals tend to guide research and development in a particular field. However, there are a large number of journals, and it is important to evaluate journals reasonably. There are many metrics for evaluating journal impact, such as the Thomson Reuters Influential Factors (IF) (Stegmann, 1997), the Eigenfactor (Ei) (Bergstrom et al., 2008), SCImago Journal Rank (Jamali et al., 2014) and Source Normalized Impact per Paper (SNIP) (Waltman et al., 2013).

At present, there are many studies that integrate multiple metrics to evaluate journals. Su et al. (2013) used the distance measures to integrate multiple indicators in a multidimensional space. Bartolucci et al. (2015) proposed a kind of latent class model to cluster and rank journals. Yu et al. (2017) proposed a mutually reinforced journal ranking model (MLMRJR) for ranking journals by considering the multiple links among authors, papers, and journals. Su et al. (2017) used the ordered weighted averaging (OWA) operators to integrate the multiple journal impact metrics and proposed the fuzzy clustering method based on linguistic terms to rank journals. Beliakov and James (2011) predicted the ranking of journals based on a citation network using a Choquet integral classifier to integrate different indicators.

6.2. Relationship-oriented applications

The applications of relationship are divided into author relationship prediction, academic recommendation, and community detection.

6.2.1. Author relationship prediction

In ASNs, the excavation of deeper scholar relationships can help the scholars to establish potential co-authorship or citation relationships. At present, the methods of identifying potential collaborations mainly focus on machine learning, link prediction, and SNs analysis.

Sun et al. (2011) analyzed co-authorship and citation relationships in heterogeneous networks of DBLP to predict co-authorship. Yang et al. (2012) proposed a MRIP method, which improved the prediction accuracy of collaboration by considering the mutual flow of information between two authors. Zhang and Yu (2014) presented a supervised machine learning method to predict collaborations in the field of biomedicine by using author's network features and semantic features. Chen and Fang (2014) established a latent collaboration index model to predict the collaboration probability between patent assignees by combining network metrics (e.g., degree and distance) with complementary metrics (e.g., assignees types and topic similarities). Guns and Rousseau (2014) combined machine learning technologies with link prediction to predict highly potential collaborators. Wang et al. (2010) mined advisor-advisee relationships in the collaboration network by constructing the time-constrained probability factor graph model. On this basis, Wang et al. (2017a) further explored advisor-advisee relationships by additionally considering the local properties like academic age and the number of co-authored papers.

6.2.2. Academic recommendation

In the era of big data and information explosion, it is not the most effective way to retrieve relevant results through manual searches and browsers. Various academic recommendation systems have been introduced to mitigate the information overload problem to filter large amounts of data.

Collaboration Recommendation. In the academic field, collaboration is beneficial to the productivity of researchers (Abramo et al., 2009). Therefore, recommending collaborators to researchers is an urgent problem to be solved. The methods of collaboration recommendation can generally be divided into four categories (Kong et al., 2017), that is, content-based recommendation, collaborative filtering-based recommendation (Pham et al., 2011), social network-based recommendation (Li et al., 2014) and hybrid-based recommendation (Kong et al., 2017).

Sugiyama and Kan (2010) introduced a generic model to recommend scholarly papers related to researchers' interest. Katz and Martin (1997) considered the different levels of collaboration and discover that researchers with higher levels of collaboration tend to be more efficient and productive. Yang et al. (2015) proposed a recommendation method in heterogeneous networks that considers the social proximity and institutional connectivity. The recommendation methods above are static, but researchers' research interests sometimes change over time. Liang et al. (2012) proposed a time-aware topic recommendation that considered the dynamics of topics. Daud (2012) proposed a time topic modeling and found that researchers' interests and relationships have changed over time. Kong et al. (2017) proposed a novel BCR model that considers the distribution of interest topics, temporal dynamics of interest and the level of collaborators to recommend helpful collaborators. Chaiwanarom and Lursinsap (2015) made recommendations based on their research interest of collaboration, seniority and evolution, and discover the interdisciplinary nature of the research questions. Based on the relevant recommendations, Chen et al. (2011) developed a system to recommend potential collaborators. Link prediction is also a method for the collaborator recommendation (Lü and Zhou, 2011). Benchettara et al. (2010) implemented collaborators' recommendation through the binary topological supervised learning method. Wang and Sukthankar (2013) proposed a new link prediction method to avoid the uniform treatment of all links. In order to predict links for heterogeneous networks, Dong et al. (2012) proposed a factor-based ranking graph model with better results.

Paper Recommendation. Due to the information overload of publications, the paper recommendation can effectively help researchers find relevant papers in their particular areas. There are a variety of paper recommendation methods based on the similarities between two papers. Since the discovery of potential papers can be considered as the process of link prediction, it can be divided into citation-based link prediction and content-based link prediction.

Strohman et al. (2007) used commonly Katz index of link prediction problems for citation recommendation. The Katz index calculates the number of paths by choosing a shorter length in the citation network. Many studies use the restarted random walk for citation analysis (Lao and Cohen, 2010), which is an efficient technique that has similarities with the PageRank algorithm. Based on this, Lao and Cohen (2010) proposed a learnable measure of proximity using machine learning techniques to weight edges. Nassiri et al. (2013) proposed a normalized similarity index (NSI) to calculate the similarity of papers based on citations. On the other hand, content-based link prediction considers semantic information. Meng et al. (2013) established a network-based model to recommend papers by considering a variety of information such as authorship, content and collaboration networks. Huang et al. (2015) combined a novel neurological probability model with semantic texts to recommend papers. Kong et al. (2018) proposed a method called VOPRec which used the node embedding technology to comprehensively consider text information and network structure information to recommend papers.

Venue Recommendation. It may be a dilemma for the researcher to choose the appropriate venue to submit the paper before writing the paper or after completing the paper. Therefore, we need a system that can recommend possible publishing venues to help researchers. In recent years, as researchers face more and more information overload problems when looking for new venues, there has been a resurgence of research and development around academic activity recommendations (Huynh and Hoang, 2012).

Luong et al. (2012) used network analysis techniques to weight venues in the network by the number of co-authored articles between both authors. However, the authors only evaluated their recommendations on a small scale, with the dataset covering only 16 venues and fewer than 1000 papers. Boukhris and Ayachi (2014) proposed a hybrid recommendation approach that used collaborators, co-citers and researchers from common academic institutions to recommend conferences. Silva et al. (2015) identified journals published by similar researchers through analyzing researchers' social networks, and recommended journals considering the quality and similarity of manuscripts. Yang and Davison (2012) used the ratings of the paper's topic and writing style, while Medvet et al. (2014) used the title and the abstract to recommend venues. In addition, Xia et al. (2013) proposed a social-awareness conference recommendation system for recommending events. Alhoori and Furuta (2017) proposed a method for evaluating venues based on user-centered altmetrics and exploring authors' reading interests.

6.2.3. Community/group detection

In SNs, network community/group detection is significant since community/group structure can be used to study human behavior. Network community structures are closely related to graph partitions in ASNs. There are various approaches to detect communities/groups. Here, we present two main detection approaches, i.e., modularity-based approaches and other community partitioning approaches.

Modularity-based approaches. Modularity is first proposed by Newman and Girvan (2004), which is a commonly used criterion for determining the quality of network partitions. Modularity is a measure that can be used to reflect the level or degree of how a network's communities may be separated and recombined. However, searching for maximum modularity of a large-scale network is a NP-hard issue. Thus, many fast approximate algorithms are developed, for example, greedy-type algorithms, simulated annealing methods, extremal optimization and spectral optimization algorithms. Among them, CNM is a classical greedy-type algorithm proposed by Clauset, Newman, and Moore with the computational complexity of $O(n\log^2 n)$ (Clauset et al., 2008). In many real large-scale academic networks, there exist hierarchical structures. Moreover, a large community may also contain many smaller communities. To detect the hierarchical structure of a given large-scale academic network, Expert et al. (2011) proposed the algorithm called BGLL. BGLL supposes every node in the given network is a community in the initialization step and then consider each node's neighboring nodes and compute the increment of the modularity value of neighboring node's community.

Modularity can be also applied in time-dependent networks, multiplex networks and multi-scale networks (Chaturvedi et al., 2012). According to slice order, a multi-slice network can be divided into mainly two kinds. One is the multi-slice network in which slices have been ordered. The multi-slice community detection method can be used to uncover some refined details over time (Mucha and Onnela, 2010). Other multi-slice networks are that slices have no order. A multiplex network formed 1640 college students in an American university that had several relationships, i.e., Facebook friendships, picture friendships, roommates, and housing-group preferences (Mucha and Onnela, 2010). In this generated multiplex network, each slice is constructed based on each kind of relationship. In academia, scholars construct the multiplex network to explore relationships between scholars' different disciplines. Sometimes, a multiplex network is called a multi-layer network.

There also exist many real networks that are embedded into Euclidean space, such as the Internet and various online social networks and transportation networks. To study the influence of spatial node-edge distributions on network topological and dynamical properties, Barthélemy (2011) modified modularity by taking consideration of the effect of space. In addition, the problem of information diffusion (Myers et al., 2012; Haralabopoulos and Anagnostopoulos, 2014a) in online social networks can be captured by multi-layer information flow (Haralabopoulos and Anagnostopoulos, 2014b). Suny et al. (2018) proposed a multiple diffusion model (*MDM*) that combines the multi-labeled Hawkes process with the topic model to infer social networks' multiple structures. They validated their model to be more effective in revealing the structure of multiplex networks through experiments on real datasets.

Other approaches. There are real datasets showing that some networks with prominent community structures cannot be recognized by modularity, while some networks without prominent community structures are commonly recognized as having prominent community structures. Besides, the resolution of modularity makes it difficult or even impossible to identify small size communities in large-scale academic social networks (Fortunato and Barthelemy, 2007). Besides, the above mentioned approaches consider that every node can be clearly classified into one community. However, communities often have overlapping nodes, which may belong to more than one community. In order to deal with this situation, Palla et al. (2005) proposed clique percolation (CP) algorithm to recognize communities with overlap with software CFinder. The key process in the CP algorithm is to find all k-cliques starting to form an initial node. After finding all cliques, a clique-clique overlapping matrix can be generated, which is similar to the adjacency matrix.

Ahn et al. (2010) proposed an edge-based community detection approach since many communities are connected through massive edges where the role of nodes is not as significant. Edge-based network partition approaches can be classified through proper thresholds to reduce confusions and communities.

6.3. Network-oriented applications

The network-oriented applications study ASNs by considering relationships and characteristics of the entire network.

6.3.1. Collaboration pattern

Since the last century, collaboration has increasingly become a mainstream scientific knowledge pattern in many fields (Wuchty et al., 2007). Scientific collaboration is also a hot spot in ASNs (Wang et al., 2017b).

Barabâsi et al. (2002) and Newman (2001, 2004), respectively, made groundbreaking research contributions to exploring collaboration networks. Subsequently, the tremendous growth of literary publications explained the structure and evolution of academic co-authorship networks, especially in the properties of “scale-free” and “small world” (Yan et al., 2010). In addition, the basic evolution mechanism of academic collaboration networks and the models of evolutionary dynamics are also widely studied (Evans et al., 2011). Still, other researchers focus on the characteristic of “social cohesion” (White and Harary, 2001) in many fields. Powell et al. (2005) confirmed the existence of “cohesive core” attribute in the field of Life Sciences. By tracking the evolution of the collaboration networks in the field of complex network research, Lee et al. (2010) discovered three primary processes in the evolution of networks: small isolated components, giant tree-shaped components with powerful cores, and large-scale recycling components. Wei et al. (2017) studied the productivity patterns and international collaboration by classifying the network into individual, institutional and international levels.

6.3.2. Interdisciplinary research

Interdisciplinary research is generally considered the best way to solve complex problems in current scientific research. Smajgl and Ward (2013) pointed out that the development of interdisciplinary research promoted methodological innovations. Porter and Rafols (2009) evaluated interdisciplinary evolution in six domains and found that the number of cited disciplines, the number of references per paper, and collaborators per articles changed significantly. They also discovered citations of one paper focusing mainly on adjacent subject areas. Cronin and Sugimoto (2014) also found an upward trend in interdisciplinarity after the decline between 1945 and 1975. By studying the evolution of collaboration network in interdisciplinary fields, Liu and Xia (2015) found that the network gradually evolved from small local clusters into the structure of “chained communities” and then to a small world structure. Chang and Huang (2012) analyzed the direct citation, bibliographic coupling and the co-authorship network to study the changes of interdisciplinary in the field of Library and Information Science. Chen et al. (2015) studied interdisciplinary changes in Biochemistry and Molecular Biology over 100 years, demonstrating that interdisciplinarity has evolved primarily from neighboring areas to distant cognitive domains. Karunan et al. (2017) constructed an interdisciplinary assessment framework to demonstrate the interdisciplinary at the paper level.

6.3.3. Research trend prediction

The research subject is dynamic as breakthrough research can promote certain areas as well as emerging new research topics. Therefore, it is imperative to effectively find hot topics in academia for researchers that can help them understand the latest concepts, technologies, and trends in their concerned fields.

Earlier, this problem was solved by manually extracting the topic for a single feature like citation relationships and co-authorship. For example, Katz et al. (2001) obtained some short-term and long-term forecasts of emerging trends through co-citation analysis, which spent a lot of time and efforts to carry out. Upham and Small (2010) used co-citation clusters to find the first 20 emerging topics. Chen et al. (2012)

combined co-citation analysis with burst detection to describe emerging topic trends and found that key clusters were often associated with important papers which not only lead to a dramatic increase in the number of citations but also have high betweenness centrality.

Co-citation is not the only method used to identify emerging trends. Duvvuru et al. (2013) analyzed the co-occurring keyword network in the academic corpus and monitored the temporal evolution of link weights to examine research trends and emerging areas. Qian et al. (2014) adopted the idea of community division and analyzed the k-core of papers to reveal the basic process of the formation and development of academic topics. However, this method is limited to data quantity and its generality is also not enough. Salatino et al. (2017) used the lique-based and triad-based method to measure the impact of the dynamic development of existing topics on the creation of new topics.

6.4. Academic social network sites

In order to help researchers establish personal profiles that enable them to share interests and papers, a series of ASNs sites emerged. We review some of the more popular sites.

AMiner was established based on Perl CGI programming in 2006. Currently, this website includes more than 130 million scholars, 233 million publications and 754 million citations (Tang et al., 2008). The current tasks of Aminer are: (1) to create semantic-based profiles of researchers; (2) to integrate SBD from multiple resources; (3) to build heterogeneous networks; (4) to analyze interesting patterns. At present, a large number of ASNs studies have been conducted based on AMiner such as extraction (Tang et al., 2010), rankings (Tang et al., 2008) and impact analysis (Wang et al., 2010).

CiteSeerX is considered as the first academic digital library to offer an autonomous citation index. CiteSeerX is unique compared to other academic digital libraries and search engines, as all files are collected from public websites. That is why users have full access to all searchable files on this site. In addition, it automatically extracts and indexes graphical entities like figures and tables. People can use metadata and text extraction services (Williams et al., 2014a) for research. However, CiteSeerX has problems with data collection and information quality (Wu and Giles).

Microsoft Academic Search is a kind of academic search engine that extracts metadata from published data sources to automatically create a researcher's profile. The profiles of researchers include literature information (list of publications, keywords, co-authors, etc.) and bibliometric indicators (papers, citations, etc.). In addition, they are classified according to the subject of papers. At the same time, it provides some visualizations, including publishing trends, co-authorship information, and co-authorship paths (Osborne et al., 2013). However, there are some problems with Microsoft Academic Search. One of them is the problem of information duplication (Ortega and Aguillo, 2014), which leads to difficult data preprocessing before using the data. In addition, the process of author disambiguation can be hard to deal with (Pitts et al., 2014). Another problem is that data is updated slowly (once a year). The final problem lies in the fact that many of these profiles belong to remote periods, which causes them to be inactive profiles.

Google Scholar is an academic retrieval system that contains basic information about publications (titles, co-author, year of publications, etc.) and indicators (citation count, i10-index, h-index, etc.). Unlike Microsoft Academic Search, its researcher profiles are created and edited by researchers themselves, so the information of each researcher is optional. But this led to the shortcomings of information standardization (De Winter et al., 2014). In addition, it can update the data frequently. Then, it can be freely accessed and extracted information by creating a web parser program (Bar-Ilan, 2008). What' s more, it is the first bibliographic search to retrieve documents not limited to libraries and traditional literature databases.

ResearchGate is an academic social networking site that allows users to upload papers, participate in discussions, and follow other researchers. It is designed to help academics establish their own profiles, share their publications and raise questions with their peers (Thelwall and Kousha, 2015). In addition, it also provides users with altmetric measures such as profile view counts and document download counts. ResearchGate provides a comprehensive evaluation index for each author based on the user's profile details, their contribution to the content, and their engagement in the site. Further, the site also has a Q and A platform that allows scholars to discuss and find answers to various topics. In the meantime, there are many topics on the site where scholars can follow these topics to view updated information on these topics in real time (Ovadia, 2014).

Academia.edu is a web-hosted platform for academic papers, allowing users to create their own profiles and upload the list of documents to Academia.edu. It has an analytic dashboard that lets users see the impact and spread of their research in real time (Meishar-Tal and Pieterse, 2017). In addition, it provides the service to send emails to account holders whenever their interested researchers release new research publications that allow readers to tag papers and remind anyone concerned with a particular topic.

7. Looking ahead

Due to the accelerating growth of SBD in recent years, researchers have realized the importance of using SNs to analyze ASNs. In order to facilitate researchers to understand ASNs, our survey work has systematically reviewed emerging areas of ASNs. We review the background, modeling methods, analysis methods, key mining techniques and popular applications in this area. Besides this, we also discuss some popular tools and sites.

SBD brings the problems for storage, processing, information extraction, analysis of data and other issues. There are some problems in the process of analyzing and mining ASNs. Firstly, due to the sheer volume of data, it is challenging for researchers to mine useful and effective information. These data also bring more cooperation opportunities for researchers. However, it is troublesome for researchers to find potential collaborators (Brandao and Moro, 2012). Secondly, ASNs are complex: citations in the papers form citation networks, while co-authorships between scholars form collaboration networks. A large number of SBD generated by different agencies and recorded by various platforms. In addition, the heterogeneity of SBD leads to different variants of entity names (Williams et al., 2014b). Thirdly, sharing data is also a challenge for ASNs. For example, issues related to intellectual property and copyright may limit the copying and sharing of data between different communities (Williams et al., 2014b). Moreover, the lack of data is also a problem that needs to be dealt with, however, when the data is large, it can be recovered or complemented by various relationships existing between the data. Many of the existing academic data platforms are designed for a certain subject, most of which are limited to the field of Computer Science. This can create limitations for interdisciplinary research. Further, it is still a challenge for scholars to study research impact evaluation. Although there are many kinds of research about journals, conferences, and institutional rankings, there is no universal framework or system to integrate them.

The future research on this subject can be started from (1) building heterogeneous academic networks, (2) establishing a unified way for academic impact assessment, (3) integrating multidisciplinary academic data resources, and (4) mining implicit indicators to explore ASNs. Studying ASNs can promote the development of related technologies, promote the popularization of related platforms and systems, and make policy design for institutions and governments more reasonable.

Acknowledgments

This work was partially supported by the National Natural Science Foundation of China (61872054), the Fund for Promoting the Reform of Higher Education by Using Big Data Technology, Energizing Teachers and Students to Explore the Future (2017A01002), and the Fundamental Research Funds for the Central Universities (DUT18JC09).

References

Aarts, E., Korst, J., 1988. Simulated Annealing and Boltzmann Machines.
Abramo, G., D'Angelo, C.A., Di Costa, F., 2009. Research collaboration and productivity: is there correlation? High. Educ. 57, 155–171.
Abramowitz, M., Stegun, I.A., et al., 1966. Handbook of mathematical functions. Appl. Math. Ser. 55.
Academia.edu, http://www.academia.edu/. (Accessed 28 March 2018).
Ahn, Y.Y., Bagrow, J.P., Lehmann, S., 2010. Link communities reveal multiscale complexity in networks. Nature 466.
Akritidis, L., Bozanis, P., 2013. A supervised machine learning classification algorithm for research articles. In: Proceedings of the 28th Annual ACM Symposium on Applied Computing. ACM, pp. 115–120.
Alhoori, H., Furuta, R., 2017. Recommendation of scholarly venues based on dynamic user interests. J. Inf. 11, 553–563.
AMiner, Aminer: search and mining of academic social networks. https://www.aminer.cn/. (Accessed 28 March 2018).
Amjad, T., Ding, Y., Daud, A., Xu, J., Malic, V., 2015. Topic-based heterogeneous rank. Scientometrics 104, 313–334.
Amjad, T., Ding, Y., Xu, J., Zhang, C., Daud, A., Tang, J., Song, M., 2017. Standing on the shoulders of giants. J. Inf. 11, 307–323.
Arif, T., 2015. Extraction visualization and analysis of co-authorship based academic social networks. Int. J. Comput. Trends Technol. 23, 85–91.
Bai, X., Liu, H., Zhang, F., Ning, Z., Kong, X., Lee, I., Xia, F., 2017a. An overview on evaluating and predicting scholarly article impact. Information 8.
Bai, X., Xia, F., Lee, I., Zhang, J., Ning, Z., 2016. Identifying anomalous citations for objective evaluation of scholarly article impact. PLoS One 11.
Bai, X., Zhang, F., Hou, J., Xia, F., Tolba, A., Elashkar, E., 2017b. Implicit multi-feature learning for dynamic time series prediction of the impact of institutions. IEEE Access 5, 16372–16382.
Bar-Ilan, J., 2008. Which h-index?—a comparison of wos, scopus and google scholar. Scientometrics 74, 257–271.
Barabási, A.L., Albert, R., 1999. Emergence of scaling in random networks. Science 286, 509–512.
Barabâsi, A.L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., Vicsek, T., 2002. Evolution of the social network of scientific collaborations. Phys. A Stat. Mech. Appl. 311, 590–614.
Barabási, B.A.L., Bonabeau, E., 2003. Scale-free. Sci. Am. 288, 50–59.
Barthélemy, M., 2011. Spatial networks. Phys. Rep. 499, 1–101.
Bartolucci, F., Dardanoni, V., Peracchi, F., 2015. Ranking scientific journals via latent class models for polytomous item response data. J. Roy. Stat. Soc. Ser. A (Stat. Soc.) 178, 1025–1049.
Bastian, M., Heymann, S., Jacomy, M., et al., 2009. Gephi: an open source software for exploring and manipulating networks. Icwsm 8, 361–362.
Batagelj, V., Mrvar, A., 2004. Pajek: program for analysis and visualization of large networks. In: Timeshift-The World in Twenty-Five Years: Ars Electronica, pp. 242–251.
Beliakov, G., James, S., 2011. Citation-based journal ranks: the use of fuzzy measures. Fuzzy Sets Syst. 167, 101–119.
Benchettara, N., Kanawati, R., Rouveirol, C., 2010. A supervised machine learning link prediction approach for academic collaboration recommendation. In: Proceedings of the Fourth ACM Conference on Recommender Systems. ACM, pp. 253–256.
Benson, A.R., Gleich, D.F., Leskovec, J., 2016. Higher-order organization of complex networks. Science 353, 163–166.
Bergstrom, C.T., West, J.D., Wiseman, M.A., 2008. The eigenfactor? metrics. J. Neurosci. 28, 11433–11434.
Birant, D., Kut, A., 2007. St-dbscan: an algorithm for clustering spatial–temporal data. Data Knowl. Eng. 60, 208–221.
Borgatti, S.P., Everett, M.G., Freeman, L.C., 2002. Ucinet for Windows: Software for Social Network Analysis.
Borgwardt, K.M., Kriegel, H.P., Wackersreuther, P., 2006. Pattern mining in frequent dynamic subgraphs. In: Data Mining, 2006. ICDM'06. Sixth International Conference on. IEEE, pp. 818–822.
Börner, K., Chen, C., Boyack, K.W., 2003. Visualizing knowledge domains. Annu. Rev. Inf. Sci. Technol. 37, 179–255.
Boukhris, I., Ayachi, R., 2014. A novel personalized academic venue hybrid recommender. In: Computational Intelligence and Informatics (CINTI), 2014 IEEE 15th International Symposium on. IEEE, pp. 465–470.
Boyack, K.W., Klavans, R., 2010. Co-citation analysis, bibliographic coupling, and direct citation: which citation approach represents the research front most accurately? J. Assoc. Inf. Sci. Technol. 61, 2389–2404.
Brandao, M.A., Moro, M.M., 2012. Affiliation influence on recommendation in academic social networks. In: AMW, pp. 230–234.
Brin, S., Page, L., 2012. Reprint of: the anatomy of a large-scale hypertextual web search engine. Comput. Network. 56, 3825–3833.
Bruns, S.B., Stern, D.I., 2016. Research assessment using early citation information. Scientometrics 108, 917–935.
CFinder, http://angle.elte.hu/clustering. (Accessed 28 March 2018).
Chaiwanarom, P., Lursinsap, C., 2015. Collaborator recommendation in interdisciplinary computer science using degrees of collaborative forces, temporal evolution of research interest, and comparative seniority status. Knowl. Base Syst. 75, 161–172.
Chang, Y.W., Huang, M.H., 2012. A study of the evolution of interdisciplinarity in library and information science: using three bibliometric methods. J. Assoc. Inf. Sci. Technol. 63, 22–33.
Chaturvedi, P., Dhara, M., Arora, D., 2012. Community detection in complex network via bgll algorithm. Int. J. Comput. Appl. 48, 32–42.
Chen, C., 2004. Searching for intellectual turning points: progressive knowledge domain visualization. Proc. Natl. Acad. Sci. Unit. States Am. 101, 5303–5310.
Chen, C., Hu, Z., Liu, S., Tseng, H., 2012. Emerging trends in regenerative medicine: a scientometric analysis in citespace. Expert Opin. Biol. Ther. 12, 593–608.
Chen, H.H., Gou, L., Zhang, X., Giles, C.L., 2011. Collabseer: a search engine for collaboration discovery. In: Proceedings of the 11th Annual International ACM/IEEE Joint Conference on Digital Libraries. ACM, pp. 231–240.
Chen, H.H., Treeratpituk, P., Mitra, P., Giles, C.L., 2013. Csseer: an expert recommendation system based on citeseerx. In: Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries. ACM, pp. 381–382.
Chen, P., Xie, H., Maslov, S., Redner, S., 2007. Finding scientific gems with google's pagerank algorithm. J. Inf. 1, 8–15.
Chen, S., Arsenault, C., Gingras, Y., Larivière, V., 2015. Exploring the interdisciplinary evolution of a discipline: the case of biochemistry and molecular biology. Scientometrics 102, 1307–1323.
Chen, T., Guestrin, C., 2016. Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining. ACM, pp. 785–794.
Chen, Y., Fang, S., 2014. Mapping the evolving patterns of patent assignees' collaboration networks and identifying the collaboration potential. Scientometrics 101, 1215–1231.
Chiang, M.F., Liou, J.J., Wang, J.L., Peng, W.C., Shan, M.K., 2013. Exploring heterogeneous information networks and random walk with restart for academic search. Knowl. Inf. Syst. 36, 59–82.
CiteSeerX, Citeseerx: an evolving scientific literature digital library and search engine. http://citeseerx.ist.psu.edu/. (Accessed 28 March 2018).
CiteSpace, http://cluster.cis.drexel.edu/cchen/citespace/. (Accessed 28 March 2018).
CitNetExplorer, http://www.citnetexplorer.nl/. (Accessed 28 March 2018).
Clauset, A., Moore, C., Newman, M.E.J., 2008. Hierarchical structure and the prediction of missing links in networks. Nature 453.
Cota, R.G., Ferreira, A.A., Nascimento, C., Gonçalves, M.A., Laender, A.H., 2010. An unsupervised heuristic-based hierarchical method for name disambiguation in bibliographic citations. J. Assoc. Inf. Sci. Technol. 61, 1853–1870.
Cronin, B., Sugimoto, C.R., 2014. Beyond Bibliometrics: Harnessing Multidimensional Indicators of Scholarly Impact. MIT Press.
Csardi, G., Nepusz, T., 2006. The igraph software package for complex network research. InterJ., Complex Syst. 1695, 1–9.
Daud, A., 2012. Using time topic modeling for semantics-based dynamic research interest finding. Knowl. Base Syst. 26, 154–163.
Daud, A., Abbasi, R., Muhammad, F., 2013. Finding rising stars in social networks. In: International Conference on Database Systems for Advanced Applications. Springer, pp. 13–24.
Daud, A., Ahmad, M., Malik, M., Che, D., 2015. Using machine learning techniques for rising star prediction in co-author network. Scientometrics 102, 1687–1711.
DBLP, The dblp computer science bibliography. https://dblp.uni-trier.de/. (Accessed 28 March 2018).
De Winter, J.C., Zadpoor, A.A., Dodou, D., 2014. The expansion of google scholar versus web of science: a longitudinal study. Scientometrics 98, 1547–1565.
Deng, H., Han, J., Lyu, M.R., King, I., 2012. Modeling and exploiting heterogeneous bibliographic networks for expertise ranking. In: Proceedings of the 12th ACM/IEEE-CS Joint Conference on Digital Libraries. ACM, pp. 71–80.
Ding, Y., 2011. Topic-based pagerank on author cocitation networks. J. Assoc. Inf. Sci. Technol. 62, 449–466.
Ding, Y., Yan, E., Frazho, A., Caverlee, J., 2009. Pagerank for ranking authors in co-citation networks. J. Assoc. Inf. Sci. Technol. 60, 2229–2243.
Ding, Y., Zhang, G., Chambers, T., Song, M., Wang, X., Zhai, C., 2014. Content-based citation analysis: the next generation of citation analysis. J. Assoc. Inf. Sci. Technol. 65, 1820–1833.
Dong, Y., Tang, J., Wu, S., Tian, J., Chawla, N.V., Rao, J., Cao, H., 2012. Link prediction and recommendation across heterogeneous social networks. In: Data Mining (ICDM), 2012 IEEE 12th International Conference on. IEEE, pp. 181–190.
Duvvuru, A., Radhakrishnan, S., More, D., Kamarthi, S., Sultornsanee, S., 2013. Analyzing structural & temporal characteristics of keyword system in academic research articles. Procedia Comput. Sci. 20, 439–445.
Evans, T., Lambiotte, R., Panzarasa, P., 2011. Community structure and patterns of scientific collaboration in business and management. Scientometrics 89, 381–396.
Expert, P., Evans, T.S., Blondel, V.D., Lambiotte, R., 2011. Uncovering space-independent communities in spatial networks. Proceed. Natl. Acad. Sci. U. S. A. 108.
Eysenbach, G., 2011. Can tweets predict citations? metrics of social impact based on twitter and correlation with traditional metrics of scientific impact. J. Med. Internet Res. 13.
Fan, J., Han, F., Liu, H., 2014. Challenges of big data analysis. Natl. Sci. Rev. 1, 293–314.
Fiala, D., Rousselot, F., Ježek, K., 2008. Pagerank for bibliographic networks. Scientometrics 76, 135–158.

Fortunato, S., Barthelemy, M., 2007. Resolution limit in community detection. Proc. Natl. Acad. Sci. Unit. States Am. 104, 36–41.

Fortunato, S., Bergstrom, C., Börner, K., Evans, J., Helbing, D., Milojević, S., Petersen, A., Radicchi, F., Sinatra, R., Uzzi, B., et al., 2018. Science of Science. Science, New York, NY, p. 359.

Fu, T.Z., Song, Q., Chiu, D.M., 2014. The academic social network. Scientometrics 101, 203–239.

Garfield, E., Pudovkin, A.I., 2004. The histcite system for mapping and bibliometric analysis of the output of searches using the isi web of knowledge. In: Proceedings of the 67th Annual Meeting of the American Society for Information Science and Technology, pp. 12–17.

Gephi, http://gephi.org/. (Accessed 28 March 2018).

GS, Google scholar: stand on the shoulders of giants. https://scholar.google.co.in/. (Accessed 28 March 2018).

Guha, S., Rastogi, R., Shim, K., 1998. Cure: an efficient clustering algorithm for large databases. In: ACM Sigmod Record. ACM, pp. 73–84.

Gunnemann, S., Farber, I., Boden, B., Seidl, T., 2010. Subspace clustering meets dense subgraph mining: a synthesis of two paradigms. In: Data Mining (ICDM), 2010 IEEE 10th International Conference on. IEEE, pp. 845–850.

Guns, R., Rousseau, R., 2014. Recommending research collaborations using link prediction and random forest classifiers. Scientometrics 101, 1461–1473.

Hagberg, A., Swart, P., S Chult, D., 2008. Exploring Network Structure, Dynamics, and Function Using NetworkX. Technical Report. Los Alamos National Laboratory (LANL).

Han, J., Pei, J., Kamber, M., 2011. Data Mining: Concepts and Techniques. Elsevier.

Haralabopoulos, G., Anagnostopoulos, I., 2014a. On the information diffusion between web-based social networks. In: International Conference on Web Information Systems Engineering. Springer, pp. 14–26.

Haralabopoulos, G., Anagnostopoulos, I., 2014b. Real time enhanced random sampling of online social networks. J. Netw. Comput. Appl. 41, 126–134.

He, B., Ding, Y., Yan, E., 2012. Mining patterns of author orders in scientific publications. J. Inf. 6, 359–367.

Herlocker, J.L., Konstan, J.A., Borchers, A., Riedl, J., 1999. An algorithmic framework for performing collaborative filtering. In: Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, pp. 230–237.

Hirsch, J.E., 2005. An index to quantify an individual's scientific research output. Proc. Natl. Acad. Sci. U. S. A 102.

HistCite, http://garfield.library.upenn.edu/histcomp/. (Accessed March 28, 2018).

Hu, X., Chu, T.H., Leung, V.C., Ngai, E.C.H., Kruchten, P., Chan, H.C., 2015. A survey on mobile social networks: applications, platforms, system architectures, and future research directions. IEEE Commun. Surv. Tutor. 17, 1557–1581.

Huan, J., Wang, W., Prins, J., 2003. Efficient mining of frequent subgraphs in the presence of isomorphism. In: Data Mining, 2003. ICDM 2003. Third IEEE International Conference on. IEEE, pp. 549–552.

Huang, M.H., Chiang, L.Y., Chen, D.Z., 2003. Constructing a patent citation map using bibliographic coupling: a study of taiwan's high-tech companies. Scientometrics 58, 489–506.

Huang, W., Wu, Z., Chen, L., Mitra, P., Giles, C.L., 2015. A neural probabilistic model for context based citation recommendation. In: AAAI, pp. 2404–2410.

Huynh, T., Hoang, K., 2012. Modeling collaborative knowledge of publishing activities for research recommendation. Comput. Collect. Intell. Technol. Appl. 41–50.

iGraph, http://igraph.org/. (Accessed 28 March 2018).

Jamali, J., Salehi-Marzijarani, M., Ayatollahi, S.M.T., 2014. Factors affecting journal quality indicator in scopus (scimago journal rank) in obstetrics and gynecology journals: a longitudinal study (1999-2013). Acta Inf. Med. 22.

Janssens, F., Zhang, L., De Moor, B., Glänzel, W., 2009. Hybrid clustering for validation and improvement of subject-classification schemes. Inf. Process. Manag. 45, 683–702.

Jeh, G., Widom, J., 2002. Simrank: a measure of structural-context similarity. In: Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, pp. 538–543.

Jones, B.F., Wuchty, S., Uzzi, B., 2008. Multi-university research teams: shifting impact, geography, and stratification in science. Science 322, 1259–1262.

Kang, I.S., Na, S.H., Lee, S., Jung, H., Kim, P., Sung, W.K., Lee, J.H., 2009. On co-authorship for author disambiguation. Inf. Process. Manag. 45, 84–97.

Kardan, A., Omidvar, A., Behzadi, M., 2012. Context based expert finding in online communities using social network analysis. Int. J. Comput. Sci. Res. Appl. 2, 79–88.

Karunan, K., Lathabai, H.H., Prabhakaran, T., 2017. Discovering interdisciplinary interactions between two research fields using citation networks. Scientometrics 113, 335–367.

Katz, J., Stewart, S., Gow, T., Martin, B., 2001. Science Foresight Project Final Report.

Katz, J.S., Martin, B.R., 1997. What is research collaboration? Res. Pol. 26, 1–18.

Kaur, R., Singh, S., 2016. A survey of data mining and social network analysis based anomaly detection techniques. Egypt. Inf. J. 17, 199–216.

Khan, H.U., Daud, A., Ishfaq, U., Amjad, T., Aljohani, N., Abbasi, R.A., Alowibdi, J.S., 2017a. Modelling to identify influential bloggers in the blogosphere: a survey. Comput. Hum. Behav. 68, 64–82.

Khan, S., Liu, X., Shakil, K.A., Alam, M., 2017b. A survey on scholarly data: from big data perspective. Inf. Process. Manag. 53, 923–944.

Khan, S., Shakil, K.A., Alam, M., 2016. Cloud-based Big Data Management and Analytics for Scholarly Resources: Current Trends, Challenges and Scope for Future Research. arXiv:1606.01808.

Kim, J., Diesner, J., 2016. Distortive effects of initial-based name disambiguation on measurements of large-scale coauthorship networks. J. Assoc. Inf. Sci. Technol. 67, 1446–1461.

Kim, J., Diesner, J., Kim, H., Aleyasen, A., Kim, H.M., 2014. Why name ambiguity resolution matters for scholarly big data research. In: Big Data (Big Data), 2014 IEEE International Conference on. IEEE, pp. 1–6.

Kleinberg, J.M., 1999. Authoritative sources in a hyperlinked environment. J. ACM (JACM) 46, 604–632.

Koller, D., 1999. Probabilistic relational models. In: International Conference on Inductive Logic Programming. Springer, pp. 3–13.

Kong, X., Jiang, H., Wang, W., Bekele, T.M., Xu, Z., Wang, M., 2017. Exploring dynamic research interest and academic influence for scientific collaborator recommendation. Scientometrics 113, 369–385.

Kong, X., Mao, M., Wang, W., Liu, J., Xu, B., 2018. Voprec: vector representation learning of papers with text information and structural identity for recommendation. IEEE Trans. Emerg. Top. Comput. PP, 1, https://doi.org/10.1109/TETC.2018.2830698.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105.

Lao, N., Cohen, W.W., 2010. Relational retrieval using a combination of path-constrained random walks. Mach. Learn. 81, 53–67.

Lee, D., Goh, K.I., Kahng, B., Kim, D., 2010. Complete trails of coauthorship network evolution. Phys. Rev. E 82.

Lehmann, S., Jackson, A.D., Lautrup, B.E., 2006. Measures for measures. Nature 444, 1003–1004.

Leung, X.Y., Sun, J., Bai, B., 2017. Bibliometrics of social media research: a co-citation and co-word analysis. Int. J. Hosp. Manag. 66, 35–45.

Li, J., Xia, F., Wang, W., Chen, Z., Asabere, N.Y., Jiang, H., 2014. Acrec: a co-authorship based random walk model for academic collaboration recommendation. In: Proceedings of the 23rd International Conference on World Wide Web. ACM, pp. 1209–1214.

Li, Q., Moon, B., et al., 2001. Indexing and querying xml data for regular path expressions. In: VLDB, pp. 361–370.

Li, X.L., Foo, C.S., Tew, K.L., Ng, S.K., 2009. Searching for rising stars in bibliography networks. In: International Conference on Database Systems for Advanced Applications. Springer, pp. 288–292.

Li, Y., Tang, J., 2008. Expertise search in a time-varying social network. In: Web-Age Information Management, 2008. WAIM'08. The Ninth International Conference on. IEEE, pp. 293–300.

Liang, H., Xu, Y., Tjondronegoro, D., Christen, P., 2012. Time-aware topic recommendation based on micro-blogs. In: Proceedings of the 21st ACM International Conference on Information and Knowledge Management. ACM, pp. 1657–1661.

Light, R.P., Polley, D.E., Börner, K., 2014. Open data and open code for big science of science studies. Scientometrics 101, 1535–1551.

Liu, P., Xia, H., 2015. Structure and evolution of co-authorship network in an interdisciplinary research field. Scientometrics 103, 101–134.

Liu, X., Yu, S., Janssens, F., Glänzel, W., Moreau, Y., De Moor, B., 2010. Weighted hybrid clustering by combining text mining and bibliometrics on a large-scale journal database. J. Assoc. Inf. Sci. Technol. 61, 1105–1119.

Lopes, G.R., Moro, M.M., Da Silva, R., Barbosa, E.M., de Oliveira, J.P.M., 2011. Ranking strategy for graduate programs evaluation. In: Procs. of ICITA.

Lü, L., Zhou, T., 2011. Link prediction in complex networks: a survey. Phys. A Stat. Mech. Appl. 390, 1150–1170.

Luo, Y.L., Hsu, C.H., 2009. An empirical study of research collaboration using social network analysis. In: Computational Science and Engineering, 2009. CSE'09. International Conference on. IEEE, pp. 921–926.

Luong, H., Huynh, T., Gauch, S., Do, L., Hoang, K., 2012. Publication venue recommendation using author network's publication history. Intell. Inf. Database Syst. 426–435.

MAS, Microsoft academic search. http://academic.research.microsoft.com/. (Accessed 28 March 2018).

Medvet, E., Bartoli, A., Piccinin, G., 2014. Publication venue recommendation based on paper abstract. In: Tools with Artificial Intelligence (ICTAI), 2014 IEEE 26th International Conference on. IEEE, pp. 1004–1010.

Meishar-Tal, H., Pieterse, E., 2017. Why do academics use academic social networking sites? Int. Rev. Res. Open Dist. Learn. 18.

Meng, F., Gao, D., Li, W., Sun, X., Hou, Y., 2013. A unified graph model for personalized query-oriented reference paper recommendation. In: Proceedings of the 22nd ACM International Conference on Information & Knowledge Management. ACM, pp. 1509–1512.

Mucha, P.J., Onnela, J.P., 2010. Community structure in time-dependent, multiscale, and multiplex networks. Science 328.

Myers, S.A., Zhu, C., Leskovec, J., 2012. Information diffusion and external influence in networks. In: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, pp. 33–41.

Nassiri, I., Masoudi-Nejad, A., Jalili, M., Moeini, A., 2013. Normalized similarity index: an adjusted index to prioritize article citations. J. Inf. 7, 91–98.

NetworkX, https://networkx.github.io/. (Accessed 28 March 2018).

Neville, J., Jensen, D., 2007. Relational dependency networks. J. Mach. Learn. Res. 8, 653–692.

Newman, M.E., 2001. Scientific collaboration networks. i. network construction and fundamental results. Phys. Rev. E 64.

Newman, M.E., 2003a. Mixing patterns in networks. Phys. Rev. E 67.

Newman, M.E., 2003b. The structure and function of complex networks. SIAM Rev. 45, 167–256.

Newman, M.E., 2004. Coauthorship networks and patterns of scientific collaboration. Proc. Natl. Acad. Sci. 101, 5200–5205.

Newman, M.E., Girvan, M., 2004. Finding and evaluating community structure in networks. Phys. Rev. E 69.

Ng, R.T., Han, J., 2002. Clarans: a method for clustering objects for spatial data mining. IEEE Trans. Knowl. Data Eng. 14, 1003–1016.

Nie, Z., Zhang, Y., Wen, J.R., Ma, W.Y., 2005. Object-level ranking: bringing order to web objects. In: Proceedings of the 14th International Conference on World Wide Web. ACM, pp. 567–574.

NodeXL, http://nodexl.codeplex.com/. (Accessed 28 March 2018).

Noll, M.G., Au Yeung, C.m., Gibbins, N., Meinel, C., Shadbolt, N., 2009. Telling experts from spammers: expertise ranking in folksonomies. In: Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, pp. 612–619.

Ortega, J.L., Aguillo, I.F., 2014. Microsoft academic search and google scholar citations: comparative analysis of author profiles. J. Assoc. Inf. Sci. Technol. 65, 1149–1156.

Osborne, F., Motta, E., Mulholland, P., 2013. Exploring scholarly data with rexplore. In: International Semantic Web Conference. Springer, pp. 460–477.

Ovadia, S., 2014. Researchgate and academia. edu: academic social networks. Behav. Soc. Sci. Libr. 33, 165–169.

Page, L., Brin, S., Motwani, R., Winograd, T., 1999. The PageRank Citation Ranking: Bringing Order to the Web. Technical Report. Stanford InfoLab.

Pajek, http://mrvar.fdv.uni-lj.si/pajek/. (Accessed 28 March 2018).

Palla, G., Derényi, I., Farkas, I., Vicsek, T., 2005. Uncovering the overlapping community structure of complex networks in nature and society. Nature 435.

Panagopoulos, G., Tsatsaronis, G., Varlamis, I., 2017. Detecting rising stars in dynamic collaborative networks. J. Inf. 11, 198–222.

Park, H.S., Jun, C.H., 2009. A simple and fast algorithm for k-medoids clustering. Expert Syst. Appl. 36, 3336–3341.

Peng, S., Zhou, Y., Cao, L., Yu, S., Niu, J., Jia, W., 2018. Influence analysis in social networks: a survey. J. Netw. Comput. Appl.

Peroni, S., Shotton, D., 2012. Fabio and cito: ontologies for describing bibliographic resources and citations. Web Semant. Sci. Serv. Agents World Wide Web 17, 33–43.

Pham, M.C., Cao, Y., Klamma, R., Jarke, M., 2011. A clustering approach for collaborative filtering recommendation using social network analysis. J. UCS 17, 583–604.

Pitts, M., Savvana, S., Roy, S.B., Mandava, V., Prasath, D., 2014. Alias: author disambiguation in microsoft academic search engine dataset. In: EDBT, pp. 648–651.

Pobiedina, N., Ichise, R., 2014. Predicting citation counts for academic literature using graph pattern mining. In: International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems. Springer, pp. 109–119.

Popescul, A., Ungar, L.H., 2003. Structural Logistic Regression for Link Analysis.

Porter, A., Rafols, I., 2009. Is science becoming more interdisciplinary? measuring and mapping six research fields over time. Scientometrics 81, 719–745.

Powell, W.W., White, D.R., Koput, K.W., Owen-Smith, J., 2005. Network dynamics and field evolution: the growth of interorganizational collaboration in the life sciences. Am. J. Sociol. 110, 1132–1205.

Qian, T., Li, Q., Liu, B., Xiong, H., Srivastava, J., Sheu, P.C.Y., 2014. Topic formation and development: a core-group evolving process. World Wide Web 17, 1343–1373.

Radicchi, F., Fortunato, S., Markines, B., Vespignani, A., 2009. Diffusion of scientific credits and the ranking of scientists. Phys. Rev. E 80.

Rahim, A., Kong, X., Xia, F., Ning, Z., Ullah, N., Wang, J., Das, S.K., 2017. Vehicular Social Networks: A Survey. Pervasive and Mobile Computing.

Rand, D.G., Arbesman, S., Christakis, N.A., 2011. Dynamic social networks promote cooperation in experiments with humans. Proceed. Natl. Acad. Sci. U. S. A. 108, 19193–19198.

Rand, D.G., Nowak, M.A., Fowler, J.H., Christakis, N.A., 2014. Static network structure can stabilize human cooperation. Proceed. Natl. Acad. Sci. U. S. A. 111, 17093–17098.

ResearchGate, http://www.researchgate.net/. (Accessed 28 March 2018).

Richardson, M., Domingos, P., 2006. Markov logic networks. Mach. Learn. 62, 107–136.

Rifai, S., Vincent, P., Muller, X., Glorot, X., Bengio, Y., 2011. Contractive auto-encoders: explicit invariance during feature extraction. In: Proceedings of the 28th International Conference on International Conference on Machine Learning. Omnipress, pp. 833–840.

Salatino, A.A., Osborne, F., Motta, E., 2017. How are topics born? understanding the research dynamics preceding the emergence of new areas. Peer J. Comput. Sci. 3.

Sayyadi, H., Getoor, L., 2009. Futurerank: ranking scientific articles by predicting their future pagerank. In: Proceedings of the 2009 SIAM International Conference on Data Mining. SIAM, pp. 533–544.

Sci2, https://sci2.cns.iu.edu/user/welcome.php. (Accessed 28 March 2018).

Serdyukov, P., 2009. Search for Expertise: Going beyond Direct Evidence. University of Twente.

Shen, H.W., Wang, D., Song, C., Barabási, A.L., 2014. Modeling and predicting popularity dynamics via reinforced Poisson processes. In: AAAI, pp. 291–297.

Shotton, D., 2009. Semantic publishing: the coming revolution in scientific journal publishing. Learn. Publ. 22, 85–94.

Silva, T., Ma, J., Yang, C., Liang, H., 2015. A profile-boosted research analytics framework to recommend journals for manuscripts. J. Assoc. Inf. Sci. Technol. 66, 180–200.

Smajgl, A., Ward, J., 2013. A framework to bridge science and policy in complex decision making arenas. Futures 52, 52–58.

Smith, M.A., Shneiderman, B., Milic-Frayling, N., Mendes Rodrigues, E., Barash, V., Dunne, C., Capone, T., Perer, A., Gleave, E., 2009. Analyzing (social media) networks with nodexl. In: Proceedings of the Fourth International Conference on Communities and Technologies. ACM, pp. 255–264.

Stegmann, J., 1997. How to evaluate journal impact factors. Nature 390, 550.

Strohman, T., Croft, W.B., Jensen, D., 2007. Recommending citations for academic papers. In: Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, pp. 705–706.

Su, P., Shang, C., Chen, T., Shen, Q., 2017. Exploiting data reliability and fuzzy clustering for journal ranking. IEEE Trans. Fuzzy Syst. 25, 1306–1319.

Su, P., Shang, C., Shen, Q., 2013. Link-based approach for bibliometric journal ranking. Soft Comput. 17, 2399–2410.

Sugiyama, K., Kan, M.Y., 2010. Scholarly paper recommendation via user's recent research interests. In: Proceedings of the 10th Annual Joint Conference on Digital Libraries. ACM, pp. 29–38.

Sugiyama, M., López, F.L., Kasenburg, N., Borgwardt, K.M., 2015. Significant subgraph mining with multiple testing correction. In: Proceedings of the 2015 SIAM International Conference on Data Mining. SIAM, pp. 37–45.

Sun, Y., Barber, R., Gupta, M., Aggarwal, C.C., Han, J., 2011. Co-author relationship prediction in heterogeneous bibliographic networks. In: Advances in Social Networks Analysis and Mining (ASONAM), 2011 International Conference on. IEEE, pp. 121–128.

Sun, Y., Han, J., 2013. Mining Heterogeneous Information Networks: a Structural Analysis Approach. ACM.

Sun, Y., Yu, Y., Han, J., 2009. 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. ACM, pp. 797–806.

Suny, P., Li, J., Mao, Y., Zhang, R., Wang, L., 2018. Inferring Multiplex Diffusion Network via Multivariate Marked Hawkes Process. arXiv:1809.07688.

Tang, J., Chang, Y., Aggarwal, C., Liu, H., 2016. A survey of signed network mining in social media. ACM Comput. Surv. (CSUR) 49.

Tang, J., Yao, L., Zhang, D., Zhang, J., 2010. A combination approach to web user profiling. ACM Trans. Knowl. Discov. Data (TKDD) 5.

Tang, J., Zhang, J., Yao, L., Li, J., Zhang, L., Su, Z., 2008. Arnetminer: extraction and mining of academic social networks. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, pp. 990–998.

Tang, L., Walsh, J.P., 2010. Bibliometric fingerprints: name disambiguation based on approximate structure equivalence of cognitive maps. Scientometrics 84, 763–784.

Taskar, B., Abbeel, P., Wong, M.F., Koller, D., 2007. Relational Markov Networks. Introduction to Statistical Relational Learning, pp. 175–200.

Team, S., 2009. Science of Science (Sci2) Tool. Indiana University and SciTech Strategies.

Thelwall, M., Kousha, K., 2015. Researchgate: disseminating, communicating, and measuring scholarship? J. Assoc. Inf. Sci. Technol. 66, 876–889.

Timilsina, M., Davis, B., Taylor, M., Hayes, C., 2016. Towards predicting academic impact from mainstream news and weblogs: a heterogeneous graph based approach. In: Advances in Social Networks Analysis and Mining (ASONAM), 2016 IEEE/ACM International Conference on. IEEE, pp. 1388–1389.

Travers, J., Milgram, S., 1977. An experimental study of the small world problem. In: Social Networks. Elsevier, pp. 179–197.

Treeratpituk, P., Giles, C.L., 2009. Disambiguating authors in academic publications using random forests. In: Proceedings of the 9th ACM/IEEE-CS Joint Conference on Digital Libraries. ACM, pp. 39–48.

UCINet, https://sites.google.com/site/ucinetsoftware/home. (Accessed 28 March 2018).

Uddin, S., Hossain, L., Rasmussen, K., 2013. Network effects on scientific collaborations. PLoS One 8.

Upham, S.P., Small, H., 2010. Emerging research fronts in science and technology: patterns of new knowledge development. Scientometrics 83, 15–38.

Van Eck, N.J., Waltman, L., 2011. Text Mining and Visualization Using Vosviewer. arXiv:1109.2058.

Van Eck, N.J., Waltman, L., 2014. Citnetexplorer: a new software tool for analyzing and visualizing citation networks. J. Inf. 8, 802–823.

VOSviewer, http://www.vosviewer.com/. (Accessed 28 March 2018).

Waltman, L., van Eck, N.J., van Leeuwen, T.N., Visser, M.S., 2013. Some modifications to the snip journal impact indicator. J. Inf. 7, 272–285.

Wang, C., Han, J., Jia, Y., Tang, J., Zhang, D., Yu, Y., Guo, J., 2010. Mining advisor-advisee relationships from research publication networks. In: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, pp. 203–212.

Wang, D., Song, C., Barabási, A.L., 2013a. Quantifying long-term scientific impact. Science 342, 127–132.

Wang, W., Liu, J., Xia, F., King, I., Tong, H., 2017a. Shifu: deep learning based advisor-advisee relationship mining in scholarly big data. In: Proceedings of the 26th International Conference on World Wide Web Companion. International World Wide Web Conferences Steering Committee, pp. 303–310.

Wang, W., Wang, C., Zhu, Y., Shi, B., Pei, J., Yan, X., Han, J., 2005. Graphminer: a structural pattern-mining system for large disk-based graph databases and its applications. In: Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data. ACM, pp. 879–881.

Wang, W., Yu, S., Bekele, T.M., Kong, X., Xia, F., 2017b. Scientific collaboration patterns vary with scholars' academic ages. Scientometrics 112, 329–343.

Wang, X., Sukthankar, G., 2013. Link prediction in multi-relational collaboration networks. In: Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. ACM, pp. 1445–1447.

Wang, Y., Tong, Y., Zeng, M., 2013b. Ranking scientific articles by exploiting citations, authors, journals, and time information. AAAI.

Wasserman, S., Faust, K., 1994. Social Network Analysis: Methods and Applications, vol. 8. Cambridge university press.

Wei, F., Zhang, G., Feng, Y., Liu, L., Shao, Z., 2017. A co-authorship network-based method for understanding the evolution of a research area: a case of information systems research. Malays. J. Libr. Inf. Sci. 22, 1–14.
White, D.R., Harary, F., 2001. The cohesiveness of blocks in social networks: node connectivity and conditional density. Socio. Methodol. 31, 305–359.
Wijegunawardana, P., Mehrotra, K., Mohan, C., 2016. Finding rising stars in heterogeneous social networks. In: Tools with Artificial Intelligence (ICTAI), 2016 IEEE 28th International Conference on. IEEE, pp. 614–618.
Williams, K., Li, L., Khabsa, M., Wu, J., Shih, P.C., Giles, C.L., 2014a. A web service for scholarly big data information extraction. In: Web Services (ICWS), 2014 IEEE International Conference on. IEEE, pp. 105–112.
Williams, K., Wu, J., Choudhury, S.R., Khabsa, M., Giles, C.L., 2014b. Scholarly big data information extraction and integration in the citeseer χ digital library. In: Data Engineering Workshops (ICDEW), 2014 IEEE 30th International Conference on. IEEE, pp. 68–73.
Wilsdon, J., et al., 2011. Knowledge, Networks and Nations: Global Scientific Collaboration in the 21st Century. The Royal Society.
Wu, J., Giles, C.L.,. Information Extraction for Scholarly Document Big Data.
Wu, Z., Wu, J., Khabsa, M., Williams, K., Chen, H.H., Huang, W., Tuarob, S., Choudhury, S.R., Ororbia, A., Mitra, P., et al., 2014. Towards building a scholarly big data platform: challenges, lessons and opportunities. In: Digital Libraries (JCDL), 2014 IEEE/ACM Joint Conference on. IEEE, pp. 117–126.
Wuchty, S., Jones, B.F., Uzzi, B., 2007. The increasing dominance of teams in production of knowledge. Science 316, 1036–1039.
Xia, F., Asabere, N.Y., Rodrigues, J.J., Basso, F., Deonauth, N., Wang, W., 2013. Socially-aware venue recommendation for conference participants. In: Ubiquitous Intelligence and Computing, 2013 IEEE 10th International Conference on and 10th International Conference on Autonomic and Trusted Computing (UIC/ATC). IEEE, pp. 134–141.
Xia, F., Wang, W., Bekele, T.M., Liu, H., 2017. Big scholarly data: a survey. IEEE Trans. Big Data 3, 18–35.
Yan, E., Ding, Y., 2011. Discovering author impact: a pagerank perspective. Inf. Process. Manag. 47, 125–134.
Yan, E., Ding, Y., Zhu, Q., 2010. Mapping library and information science in China: a coauthorship network analysis. Scientometrics 83, 115–131.
Yan, R., Tang, J., Liu, X., Shan, D., Li, X., 2011. Citation count prediction: learning to estimate future citations for literature. In: Proceedings of the 20th ACM International Conference on Information and Knowledge Management. ACM, pp. 1247–1252.
Yan, X., Han, J., 2002. gspan: graph-based substructure pattern mining. In: Data Mining, 2002. ICDM 2003. Proceedings. 2002 IEEE International Conference on. IEEE, pp. 721–724.
Yan, X., Han, J., 2003. Closegraph: mining closed frequent graph patterns. In: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, pp. 286–295.
Yang, C., Ma, J., Silva, T., Liu, X., Hua, Z., 2013. A multilevel information mining approach for expert recommendation in online scientific communities. Comput. J. 58, 1921–1936.
Yang, C., Sun, J., Ma, J., Zhang, S., Wang, G., Hua, Z., 2015. Scientific collaborator recommendation in heterogeneous bibliographic networks. In: System Sciences (HICSS), 2015 48th Hawaii International Conference on. IEEE, pp. 552–561.
Yang, Y., Chawla, N., Sun, Y., Hani, J., 2012. Predicting links in multi-relational and heterogeneous networks. In: Data Mining (ICDM), 2012 IEEE 12th International Conference on. IEEE, pp. 755–764.
Yang, Z., Davison, B.D., 2012. Venue recommendation: submitting your paper with style. In: Machine Learning and Applications (ICMLA), 2012 11th International Conference on. IEEE, pp. 681–686.
Yao, L., Wei, T., Zeng, A., Fan, Y., Di, Z., 2014. Ranking scientific publications: the effect of nonlinearity. Sci. Rep. 4.
Yu, D., Wang, W., Zhang, S., Zhang, W., Liu, R., 2017. A multiple-link, mutually reinforced journal-ranking model to measure the prestige of journals. Scientometrics 111, 521–542.
Zhang, J., Xia, F., Wang, W., Bai, X., Yu, S., Bekele, T.M., Peng, Z., 2016. Cocarank: a collaboration caliber-based method for finding academic rising stars. In: Proceedings of the 25th International Conference Companion on World Wide Web. International World Wide Web Conferences Steering Committee, pp. 395–400.
Zhang, Q., Yu, H., 2014. Computational approaches for predicting biomedical research collaborations. PLoS One 9.
Zhang, T., Ramakrishnan, R., Livny, M., 1996. Birch: an efficient data clustering method for very large databases. In: ACM Sigmod Record. ACM, pp. 103–114.
Zhou, Y.B., Lü, L., Li, M., 2012. Quantifying the influence of scientists and their publications: distinguishing between prestige and popularity. New J. Phys. 14.
Zitt, M., Lelu, A., Bassecoulard, E., 2011. Hybrid citation-word representations in science mapping: Portolan charts of research fields? J. Assoc. Inf. Sci. Technol. 62, 19–39.

Xiangjie Kong received the BSc and PhD degrees from Zhejiang University, Hangzhou, China. He is currently an Associate Professor in School of Software, Dalian University of Technology, China. He has served as (Guest) Editor of several international journals, Workshop Chair or PC Member of a number of conferences. Dr. Kong has published over 70 scientific papers in international journals and conferences (with 50 + indexed by ISI SCIE). His research interests include human behavior, mobile computing, and computational social science. He is a Senior Member of IEEE and CCF, and a Member of ACM.

Yajie Shi received the BSc degree in process equipment and control engineering from Shaanxi University of Science and Technology, China, in 2016. She is currently working toward the master's degree in the School of Software, Dalian University of Technology, China. Her research interests include computational social science, social network analysis, and science of success.

Shuo Yu received her M.Sc. degree in Applied Mathematics from Shenyang University of Technology, Shenyang, China, in 2014. She is currently working toward the Ph.D. degree in Software Engineering in Dalian University of Technology (DUT), Dalian, China. Her research interests include big networks, big scholarly data, social network analysis, science of scientific team science, and computational social science.

Jiaying Liu received the BSc degree in software engineering from Dalian University of Technology, China, in 2016. She is currently working toward the master's degree in the School of Software, Dalian University of Technology, China. Her research interests include big scholarly data, social network analysis, and science of success.

Feng Xia received the BSc and Ph.D. degrees from Zhejiang University, Hangzhou, China. He is currently a Full Professor in School of Software, Dalian University of Technology, China. Dr. Xia has published 2 books and over 200 scientific papers in international journals and conferences. His research interests include computational social science, network science, data science, and mobile social networks. He is a Senior Member of IEEE and ACM.