Data-Driven Scholarly Intelligence: A Survey and Outlook

Chengzhe Yuan[1,5], Yong Tang[2*,3], Ronghua Lin[3,5], Feiyi Tang[4,5], Guohua Chen[3]

[1]School of Electronics and Information, Guangdong Polytechnic Normal University, Guangzhou 510665, China
[2]Institute of Data Intelligence, Guangdong University of Science and Technology, Dongguan 523083, China
[3]School of Computer Science, South China Normal University, Guangzhou 510631, China
[4]School of Information Engineering, Guangzhou Polytechnic University, Guangzhou 511483, China
[5]Pazhou Lab, Guangzhou 510330, Guangdong, China
[*]Corresponding author: ytang@m.scnu.edu.cn

Abstract: Digital scholarly platforms are reshaping scholarly communication by generating heterogeneous data beyond tra- ditional publication and citation records. This survey reviews data-driven scholarly intelligence as a framework for organizing such data and transforming them into human-centered academic and knowledge services. We first characterize the scholarly data ecosystem and discuss how heterogeneous data can be structured through data organization and scholarly knowledge modeling. We then synthesize representative methods from graph-based modeling, recommendation, language understanding, knowledge grounding, and large-language-model-based agents. The survey further examines how scholarly intelligence can support academic interaction, knowledge matching, decision-oriented services, and research workflow assistance. Finally, we identify challenges related to data governance, fairness, explainability, evaluation, reproducibility, and trustworthy service design. By connecting scholarly data mining, academic social network analysis, and AI- enabled research services, this survey provides a compact view of how scholarly data can support evidence-grounded, explainable, and responsible scholarly intelligence.

Keywords: scholarly intelligence; scholarly data mining; academic social networks; knowledge services; artificial intelli- gence.

I Introduction

The rapid growth of digital scholarly platforms has changed how researchers produce, communicate, evaluate, and use academic knowledge [1]–[3]. Scholarly activities are no longer represented only by formal publications and citation rela- tionships; instead, they are increasingly supported by digital libraries, academic social networks, institutional repositories, research platforms, and knowledge service systems [2]–[4]. These platforms generate heterogeneous scholarly data that capture not only research outputs, but also scholarly identities, social interactions, institutional contexts, and knowledge dis- semination processes [2]–[4]. As a result, scholarly activities can now be observed, modeled, and supported through data science and artificial intelligence methods.

Earlier studies on scholarly data mining have shown that large-scale scholarly data can be used for citation analysis, document analysis, conference analysis, trend analysis, litera- ture analysis, recommendation, and scientific text summariza- tion [2], [3]. These studies provide an important foundation for

understanding the structure and dynamics of science. However, the current research landscape has moved beyond conventional mining tasks. With the development of knowledge graphs, representation learning, retrieval-augmented generation, and large language models, scholarly data are increasingly used to build intelligent services that support the broader research lifecycle.

This shift motivates the concept of data-driven scholarly intelligence, which we define as the use of data science and artificial intelligence methods to transform heterogeneous scholarly data into analytical insights and intelligent research services [2], [3], [5]. Unlike traditional bibliometric analysis, which mainly focuses on publications and citation relationships, scholarly intelligence extends the analytical scope to the contextual, relational, and behavioral signals generated by digital scholarly platforms [1], [3], [4]. It therefore emphasizes not only the extraction of patterns from scholarly data, but also the development of human-centered services for knowledge discovery, research support, and decision-oriented knowledge use [4]–[6].

Academic social networks constitute an important appli- cation scenario for scholarly intelligence because they com- bine scholarly outputs with platform-mediated social inter- actions [3], [4]. Platforms such as ResearchGate1[1], AMiner2[2], and SCHOLAT3[3], among others, connect researcher profiles, publications, research interests, social relations, and academic communication records [4], [7]–[9]. Such data make it possible to study collaboration patterns, community structures, infor- mation diffusion, scholarly influence, and research interest evolution [3], [4], [10], [11]. In this survey, academic social networks are not treated as the sole focus of the paper, but as representative environments in which data-driven scholarly in- telligence methods can be developed, evaluated, and deployed. Beyond academic social networking, scholarly intelligence can also support think tanks and policy-oriented knowledge services. In these scenarios, scholarly data are organized for expert matching, evidence retrieval, topic monitoring, and decision support [3], [5]. Academic social networks emphasize scholarly interaction and communication, whereas think tank systems focus more on organizing expertise and evidence for knowledge services. Both scenarios require heterogeneous scholarly data, AI-based modeling methods, and human- centered service design.

Recent surveys on AI for research and LLM-based scientific agents further indicate a shift from isolated AI-assisted tasks toward integrated, evidence-grounded, and verifiable research workflows [5], [6]. This trend suggests that scholarly intelli- gence should not be limited to static data mining, but should move toward interactive, explainable, and trustworthy services built upon heterogeneous scholarly data. This survey therefore bridges three related research streams: scholarly data mining, academic social network analysis, and AI-enabled research services.

The contributions of this survey are threefold. First, it pro- vides a data-centric synthesis of scholarly intelligence by clar- ifying how heterogeneous scholarly data can be organized to support intelligent academic and knowledge services. Second, it develops a

1 https://www.researchgate.net/
2 https://www.aminer.cn/
3 https://www.scholat.com/

method-oriented taxonomy that connects graph-based modeling, language-based understanding, knowledge- grounded generation, and AI-enabled service construction. Third, it summarizes representative application scenarios and open challenges, highlighting the transition from scholarly data analysis toward human-centered, trustworthy, and intelligent scholarly services.

The remainder of this paper is organized as follows. Sec- tion II introduces the data ecosystem of scholarly intelligence. Section III reviews representative data-driven methods. Sec- tion IV discusses major applications. Section V presents open challenges and future directions. Section VI concludes the survey.

II Scholarly data ecosystems

Data-driven scholarly intelligence is positioned at the in- tersection of scholarly data mining and AI-enabled research services. While scholarly data mining provides methods for extracting patterns from large-scale scholarly data [3], recent AI-enabled research systems further emphasize interactive, evidence-grounded, and service-oriented workflows [5], [6]. From this perspective, scholarly intelligence focuses on how heterogeneous scholarly data can be organized, modeled, and transformed into intelligent, human-centered academic ser- vices. Accordingly, this section first summarizes the major data sources that provide the empirical foundation for scholarly intelligence.

Figure 1 illustrates a data-centric framework for scholarly intelligence, organized into four connected layers: the schol- arly data ecosystem, scholarly data organization, scholarly knowledge modeling, and scholarly intelligence services. Het- erogeneous scholarly data are first integrated and structured to generate reliable entity, relation, content, and behavioral representations [2]–[4], [12], [13]. These structured repre- sentations support scholarly knowledge modeling, which in turn enables downstream services such as research recom- mendation, expert discovery, communication analysis, impact assessment, and decision-oriented knowledge support. Addi- tionally, feedback from service use can enrich the scholarly data ecosystem, while governance requirements—including data quality, privacy, fairness, explainability, reproducibility, and evaluation—remain critical across the framework [5], [6], [14]–[16].

A. Researcher and Profile Data

Researcher and profile data describe scholarly identities and provide anchors for linking scholars with publications, insti- tutions, topics, and platform interactions. These data support scholar profiling, expert discovery, reviewer recommendation, institutional analysis, and talent search [13], [17], [18]. In academic social networks, researcher profiles are usually con- nected with publications, research interests, social relations, and user-generated content, making them more informative than static bibliographic records [4], [7].

B. Publication and Citation Data

Publication and citation data remain the foundation of scholarly analysis. They support bibliometric analysis, paper recommendation, research trend detection, knowledge evo- lution analysis, and academic impact assessment [1], [3]. Large-scale scholarly infrastructures such as DBLP, S2ORC, Semantic Scholar, and OpenAlex further demonstrate how bibliographic metadata, full-text corpora, citation links, and scholarly entities can be organized for computational analy- sis [12], [13], [18], [19]. However, publication and citation data mainly reflect formal scholarly outputs and cannot fully capture informal communication, social interaction, reading behavior, or platform-mediated knowledge diffusion.

C. Social Interaction and Behavioral Data

Social interaction and behavioral data capture how scholars connect, communicate, and disseminate knowledge beyond formal publication channels. These data support community detection, information diffusion modeling, collaboration pat- tern discovery, user behavior analysis, and social influence estimation [3], [4]. Academic social networking platforms pro- vide typical examples of such data. For instance, SCHOLAT- related studies have explored scholar recommendation, in- fluence analysis, scholar knowledge graph construction, and LLM-based academic services using platform-level scholarly data [9]–[11].

D. Institutional and Policy-Oriented Knowledge Data

Institutional and policy-oriented knowledge data connect scholarly expertise with organizational resources and decision- oriented knowledge services. They support expert match- ing, policy knowledge retrieval, institutional profiling, topic monitoring, and decision support [3], [5]. Compared with academic social networks, policy-oriented knowledge systems place stronger emphasis on evidence organization, domain expertise, and decision support. Nevertheless, both scenarios require structured modeling of scholars, documents, topics, institutions, and relationships.

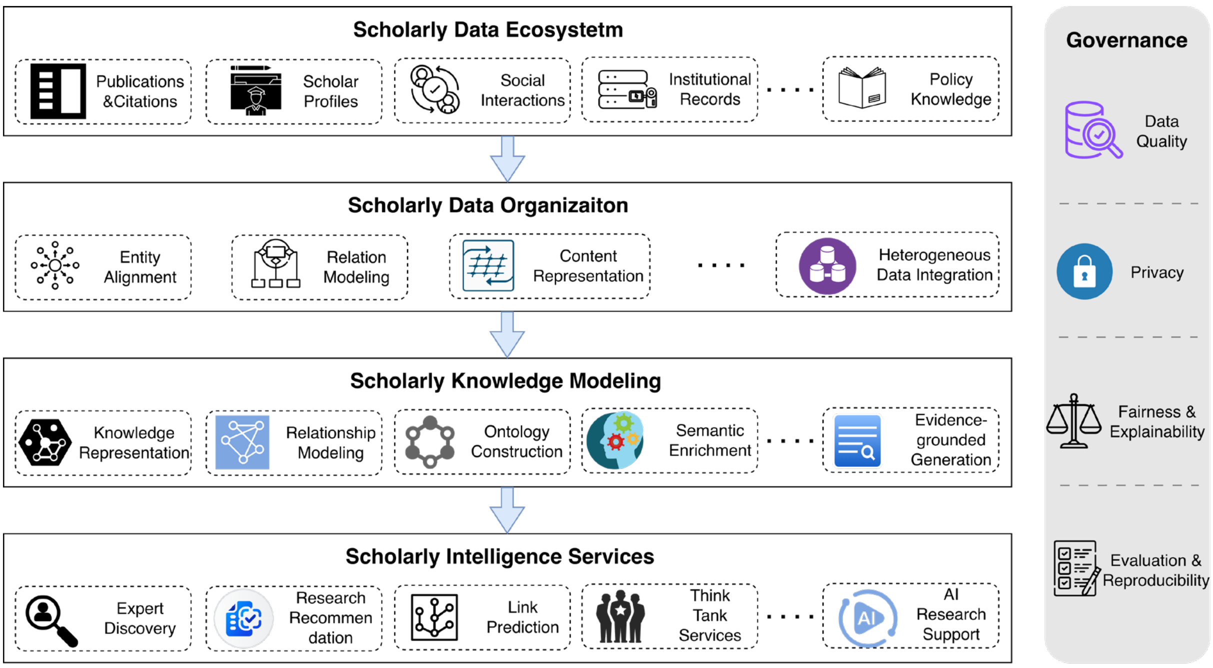

Figure 1. A data-centric framework for scholarly intelligence. Heterogeneous scholarly data are organized into structured representations and further transformed into scholarly knowledge models to support intelligent academic and knowledge services. Governance requirements, including data quality, privacy, fairness, explainability, evaluation, and reproducibility, operate across the framework.

Overall, scholarly intelligence depends on the integration of multi-source, heterogeneous, dynamic, and socially embedded scholarly data. Researcher profiles provide identity anchors, publications and citations provide formal scholarly evidence, social interaction data reveal communication and collaboration patterns, and institutional and policy-oriented knowledge data connect academic expertise with decision-making needs.

III Data-driven methods for scholarly intelligence

Data-driven scholarly intelligence relies on methods that can model heterogeneous scholarly entities, extract knowledge from text, infer relationships, recommend resources, and support interactive services. We summarize these methods into five categories.

A. Graph Mining and Network Analysis

Scholarly activities naturally form graph structures, includ- ing co-authorship networks, citation networks, academic social networks, institutional networks, and knowledge diffusion networks [4], [18], [20]. Graph mining methods, including centrality analysis, community detection, link prediction, net- work embedding, temporal network analysis, and graph neural networks, can be used to identify influential scholars, de- tect research communities, predict collaborations, and analyze knowledge diffusion [1], [21].

Earlier scholarly data mining studies have emphasized social network analysis as an important method for understanding academic networks formed by citations, collaborations, and academic communications [3]. Recent studies further extend this direction through graph representation learning and graph neural networks. In academic social networking contexts, graph-based methods can integrate social relations, research interests, publications, and behavioral traces. Platform-based examples, such as scholar recommendation and influence max- imization studies on SCHOLAT, illustrate how graph mining can support practical scholarly intelligence services [10], [11].

B. Recommender Systems

Recommender systems are important mechanisms for reduc- ing scholarly information overload and matching researchers with relevant academic resources. Existing surveys show that research-paper and scholarly recommendation have developed into mature directions covering papers, experts, collaborators, reviewers, venues, datasets, and funding opportunities [22]– [24]. In scholarly intelligence, these tasks require the joint modeling of user interests, publication records, textual content, citation relations, social networks, institutional contexts, and behavioral signals.

Compared with general recommendation scenarios, schol- arly recommendation often involves knowledge-intensive ob- jects and stronger requirements for transparency. Researchers need not only relevant recommendations, but also interpretable evidence explaining why a paper, expert, collaborator, or topic is suitable. Therefore, graph-based, content-aware, and knowledge-aware recommendation methods are particularly useful for scholarly intelligence [17], [22], [23], [25], [26]. In academic social networking scenarios, scholar recommendation studies further show that social relations, research inter- ests, publication records, and auxiliary academic information can be jointly used to improve recommendation relevance [10], [27].

C. Natural Language Processing and Scholarly Text Mining

NLP provides key techniques for extracting structured knowledge from unstructured scholarly texts. Applicable data include paper titles, abstracts, full texts, keywords, academic posts, expert profiles, policy reports, and think tank documents. Representative tasks include topic modeling, keyword ex- traction, text classification, named entity recognition, relation extraction, document summarization, citation context analysis, and trend detection.

In scholarly data mining, NLP has been used for document analysis, scientific text summarization, topic classification, and trend analysis [3], [28]. In scholarly intelligence, these methods can be further connected with platform services. For example, NLP can classify academic news, extract research topics from posts, summarize policy reports, identify expert domains, and detect emerging research trends. When combined with LLMs, NLP-based scholarly intelligence shifts from shallow text processing toward deeper semantic understanding and interactive knowledge services.

D. Knowledge Graphs and Representation Learning

Scholarly knowledge graphs organize entities such as schol- ars, papers, institutions, venues, topics, projects, reports, and their relationships. They provide structured representations for entity linking, relation extraction, semantic retrieval, knowl- edge reasoning, and academic question answering. Heteroge- neous graph representation learning can further capture the semantics of diverse scholarly entities and relations [29].

Knowledge graphs are particularly useful for expert discov- ery and decision-support services because they provide explicit relational evidence. For example, a scholar can be linked to publications, projects, collaborators, institutions, topics, and citation contexts. Such relational evidence can improve the transparency of recommendations and expert matching. Academic social network applications have shown that scholar knowledge graphs can support scholar profiling, representative work identification, research field classification, and academic information recommendation [9].

E. RAG, LLMs, and Scholarly Agents

Large language models provide new opportunities for schol- arly intelligence by supporting literature understanding, schol- arly question answering, research assistance, and knowledge service generation. However, standalone LLMs are limited by hallucination, outdated parametric knowledge, weak source attribution, and insufficient domain grounding. Retrieval-augmented generation (RAG) mitigates these limitations by grounding model outputs in external scholarly resources, while knowledge graphs can provide structured relational evidence for entity-centric reasoning [30]–[33]. Recent graph-enhanced RAG studies further suggest that scholarly evidence should be organized not only as retrieved text, but also through relations among scholars, papers, institutions, topics, and reports. For example, KGCRAG uses community detection over knowl- edge graphs to construct focused contexts for LLM-based question answering, illustrating a shift toward relation-aware evidence grounding in scholarly QA [34].

Beyond question answering, LLM-based scholarly systems are evolving toward agentic services that combine retrieval, tool use, memory, and verification. Recent surveys on AI for research and scientific agents indicate a broader transition from isolated AI-assisted tasks to integrated and verifiable research workflows [5], [6]. ScholatGPT further illustrates how scholar knowledge graphs, RAG, and domain adaptation can support academic question answering, scholar profiling, scholar think tanks, and academic information recommendation [9]. Overall, RAG, LLMs, and scholarly agents expand scholarly intelli- gence from static data analysis toward interactive, evidence- grounded, and trustworthy scholarly services.

IV Applications of data-driven scholarly intelligence

After scholarly data are organized and modeled, the next question is how these representations can be used in concrete scholarly settings. This section reviews representative applica- tions according to the primary scholarly functions they sup- port: connecting researchers, matching knowledge resources, tracing scholarly communication, and organizing expertise for knowledge services. Table I summarizes these application scenarios, together with their typical data sources, methods, and service forms.

A. Academic Social Networking

Academic social networks provide a platform-level envi- ronment for observing how scholars connect, communicate, collaborate, and disseminate knowledge. Beyond formal pub- lications and citation links, these platforms capture social re- lations, academic posts, reposts, comments, reading behavior, and institutional communication records. Such data support community detection, collaboration analysis, information dif- fusion modeling, and scholarly communication analysis [1], [3], [4].

From the perspective of scholarly intelligence, academic social networks are not only communication platforms but also empirical environments for studying how scholarly infor- mation circulates across researchers, teams, institutions, and communities. Platforms such as AMiner and SCHOLAT illus- trate how researcher profiles, publications, research interests, social links, and communication traces can be integrated to support scholar recommendation, influence analysis, academic information diffusion, and knowledge service construction [8]– [11]. This expands scholarly analysis from citation-centered impact assessment to broader socio-technical processes of academic communication and knowledge dissemination.

Table 1 Representative applications of data-driven scholarly intelligence

Scenario	Data Sources	Methods	Services
Academic social networking	Profiles, publications, social links, posts	Graph mining, community detection,diffusion modeling	Community analysis, col- laboration analysis, influence analysis
Research recommendation and expert discovery	Publications, citations, inter- ests, profiles	Recommender systems, ex- pertise retrieval, graph-based matching	Paper recommendation, expert discovery, collaborator matching
Scholar think tanks and knowledge services	Expert profiles,reports,insti- tutions, topics	Knowledge graphs, NLP, RAG, recommender systems	Expert matching, policy retrieval, decision support
AI-assisted research lifecy- cle services	Literature, citations,manu- scripts, reviews	LLMs, scholarly agents, retrieval, verification	Survey support, writing assis- tance, review support

Source: Guodian Electric Power's 2022-2024 Annual Report

B. Research Recommendation and Expert Discovery

Research recommendation and expert discovery aim to match researchers with relevant scholarly resources, collab- orators, and domain experts. These services are important for reducing information overload and improving the efficiency of academic resource allocation. Existing studies have examined a wide range of scholarly recommendation tasks, including paper recommendation, collaborator recommendation, reviewer recommendation, venue recommendation, dataset recommendation, and funding recommendation [22]–[24]. Expertise retrieval further provides a methodological foundation for identifying suitable experts based on publications, profiles, topics, and other evidence [17].

In scholarly intelligence, recommendation and expert discovery require the integration of publication records, textual content, citation relations, research interests, social connections, and behavioral signals. Academic social networking platforms provide useful contexts for these tasks because they combine scholarly profiles with social and interaction data. Existing studies on collaboration recommendation and scholar recommendation show that social relations, research interests, publication records, and auxiliary academic infor- mation can be jointly modeled to improve recommendation relevance [10], [27]. These applications demonstrate how scholarly intelligence can move beyond passive information retrieval toward personalized, explainable, and context-aware academic services.

C.Scholar Think Tanks and Knowledge Services

Scholar think tanks and knowledge service systems organize experts, evidence, topics, institutions, and reports for decision support. In scholarly intelligence, these services connect academic expertise with policy-oriented knowledge needs through expert databases, policy report analysis, topic monitoring, institutional knowledge management, and decision-support re- trieval.

This scenario links academic social networks with broader knowledge service systems. Academic social networks provide scholar-centered interaction and expertise data, while think tanks require structured expert knowledge and evidence-based decision support. Knowledge graphs, NLP, recommender sys- tems, and RAG can jointly support expert matching, report summarization, policy topic analysis, and decision-oriented knowledge generation. Applications developed on top of aca- demic social network data, such as scholar profiling and scholar think tank services, illustrate this transition from platform data to actionable scholarly intelligence [9].

D. AI-Assisted Research Lifecycle Services

AI-assisted research lifecycle services extend scholarly in- telligence from platform-level analysis to direct support for research activities. Recent studies on AI for research and scientific agents show that AI systems are increasingly used to support literature understanding, related-work retrieval, re- search planning, writing assistance, and review-oriented eval- uation [5], [6]. In scholarly intelligence, these services should be grounded in reliable scholarly data rather than relying only on the parametric knowledge of large language models.

A key distinction of scholarly intelligence is that research assistance is connected with structured scholarly evidence. Literature search and survey support require publication meta- data, citation relations, topic structures, and evidence retrieval; reviewer or expert matching requires scholar profiles, expertise representations, and conflict-aware matching; writing or report generation requires source attribution and factual verifica- tion. AI-assisted research services can therefore be viewed as downstream applications of scholarly data organization and knowledge modeling. Their practical value depends on evidence grounding, transparent provenance, and human over- sight, especially in academic evaluation, research decision- making, and policy-oriented knowledge services.

V Challenges and outlook

Although data-driven scholarly intelligence provides new opportunities for organizing scholarly data and supporting intelligent academic services, several challenges remain. These challenges are not isolated technical issues; they are closely related to data governance, responsible modeling, evaluation standards, and the long-term development of trustworthy scholarly service systems.

A. Data Governance

Scholarly data are often incomplete, noisy, and fragmented across platforms. Author name ambiguity, institution name inconsistency, duplicate entities, missing profiles, outdated records, and inconsistent metadata can affect downstream analysis [2], [3]. These problems become more complex when publication data are integrated with social relations, behav- ioral traces, academic posts, institutional records, and policy- oriented knowledge resources.

Robust entity alignment, data cleaning, relation extraction, and cross-platform integration are therefore essential for building reliable scholarly intelligence systems.

Data governance is also critical because scholarly intelli- gence may involve sensitive information, such as social re- lationships, reading behavior, browsing records, recommenda- tion logs, and user-generated content. Even when such data are generated in academic contexts, their use raises concerns about user consent, data ownership, privacy protection, and platform governance. Future systems should incorporate privacy-aware and ethically governed design, including anonymization, ac- cess control, transparent data-use policies, and human over- sight.

B. Fairness and Explainability

Data-driven methods may amplify existing inequalities in scholarly ecosystems. Recommendation and ranking algo- rithms can increase exposure for already prominent scholars, institutions, or publications, while reducing the visibility of early-career researchers, small institutions, interdisciplinary work, non-English scholarship, or underrepresented commu- nities. This effect may become stronger when algorithms rely heavily on citation counts, publication volume, social centrality, or institutional prestige [1], [15], [35].

Fairness and explainability are therefore central require- ments for scholarly intelligence. In expert discovery, academic evaluation, scholar recommendation, and policy-oriented deci- sion support, users need to understand why a scholar, paper, topic, or report is recommended. Explainable systems should provide transparent evidence, such as relevant publications, collaboration paths, research topics, citation evidence, social connections, or knowledge graph relations [14]. For LLM-based scholarly services, trustworthiness further requires re- trieval grounding, source attribution, hallucination detection, and human verification.

C. Evaluation and Reproducibility

The lack of open benchmark datasets and unified evaluation protocols makes it difficult to compare methods across schol- arly intelligence tasks. Different studies may define scholarly impact, expert relevance, recommendation quality, communi- cation influence, or service usefulness in inconsistent ways. Platform-specific data restrictions further limit reproducibil- ity, especially for tasks involving academic social networks, behavioral traces, and user interaction records.

Future work should construct benchmark datasets, standard tasks, and reproducible evaluation frameworks for scholarly intelligence. Potential tasks include scholar recommendation, collaborator prediction, expert search, influence maximiza- tion, academic information diffusion, trend detection, litera- ture retrieval, and knowledge-grounded question answering. In addition to accuracy-oriented metrics, evaluation should consider explainability, fairness, robustness, source grounding,and user-centered utility. Reproducible workflows are particu- larly important when scholarly intelligence systems are used in academic evaluation, research decision-making, or policy- oriented knowledge services [5], [6], [16].

D. Future Directions

Several trends point toward the future development of data- driven scholarly intelligence. First, scholarly data ecosys- tems will become more integrated, connecting publications, citations, profiles, social interactions, institutional records, and policy-oriented knowledge resources. Second, knowledge graphs and retrieval-augmented generation will play a larger role in evidence-grounded scholarly services by linking textual evidence with structured relations among scholars, papers, institutions, topics, and reports [32], [33]. Third, personal- ized research assistants and scholarly agents will become more feasible by combining user profiles, research interests, literature collections, interaction histories, and verification mechanisms [5], [6].

Across these directions, the goal of scholarly intelligence should not be to replace human scholarly judgment. Instead, future systems should help researchers, institutions, platforms, and decision makers better organize, discover, evaluate, and use scholarly knowledge. Advancing this field therefore re- quires not only stronger models, but also reliable data in- tegration, transparent evidence grounding, fair recommenda- tion, privacy-aware governance, reproducible evaluation, and human-centered system design.

VI Conclusion

This survey reviewed data-driven scholarly intelligence as a framework for transforming heterogeneous scholarly data into intelligent academic and knowledge services. Rather than treating scholarly data as isolated bibliographic records, this perspective emphasizes the integration of researcher profiles, publications, citations, social interactions, behavioral traces, institutional records, and policy-oriented knowledge resources. By organizing these data through graph-based modeling, rec- ommendation, natural language processing, knowledge graphs, retrieval-augmented generation, and large language models, scholarly intelligence can support a range of services, includ- ing academic social networking, research recommendation, expert discovery, scholar think tanks, and AI-assisted research lifecycle support.

The survey also highlighted that the development of schol- arly intelligence should not be evaluated only by the perfor- mance of individual models. Its practical value depends on the reliability of data integration, the interpretability of gen- erated services, the fairness of recommendation and ranking mechanisms, and the reproducibility of evaluation protocols. Academic social networks and related platform-based applica- tions, including examples such as SCHOLAT, illustrate how multi-source scholarly data can be connected with intelligent services, but they also expose challenges related to privacy, governance, bias, and evidence grounding. Therefore, data- driven scholarly intelligence should be understood as both a technical direction and a socio-technical research agenda for building trustworthy, human-centered scholarly service systems.

References

[1] S. Fortunato, C. T. Bergstrom, K. B”orner, J. A. Evans, D. Helbing,

S. Milojevic´, A. M. Petersen, F. Radicchi, R. Sinatra, B. Uzzi, A. Vespig- nani, L. Waltman, D. Wang, and A.-L. Baraba´si, “Science of science,” Science, vol. 359, no. 6379, p. eaao0185, 2018.

[2] F. Xia, W. Wang, T. M. Bekele, and H. Liu, “Big scholarly data: A survey,” IEEE Transactions on Big Data, vol. 3, no. 1, pp. 18–35, 2017.

[3] A. Dridi, M. M. Gaber, R. M. A. Azad, and J. Bhogal, “Scholarly data mining: A systematic review of its applications,” WIREs Data Mining and Knowledge Discovery, vol. 11, no. 2, p. e1395, 2021.

[4] X. Kong, Y. Shi, S. Yu, J. Liu, and F. Xia, “Academic social networks: Modeling, analysis, mining and applications,” Journal of Network and Computer Applications, vol. 132, pp. 86–103, 2019.

[5] Q. Chen, M. Yang, L. Qin, J. Liu, Z. Yan, J. Guan, D. Peng, Y. Ji,

H. Li, M. Hu et al., “Ai4research: A survey of artificial intelligence for scientific research,” arXiv preprint arXiv:2507.01903, 2025.

[6] S. Ren, C. Xie, P. Jian, Z. Ren, C. Leng, and J. Zhang, “Towards scientific intelligence: A survey of llm-

based scientific agents," arXiv preprint arXiv:2503.24047, 2025.

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

[8] J. Tang, "Aminer: Toward understanding big scholar data," in Proceed- ings of the 9th ACM International Conference on Web Search and Data Mining, 2016, p. 467.

[9] C. Yuan, G. Chen, D. Li, Y. Zhu, R. Lin, H. Zhong, and Y. Tang, "Scholatgpt: A large language model for academic social networks and its intelligent applications," Journal of Computer Applications, vol. 45, no. 3, pp. 755–764, 2025.

[10] C. Yuan, Y. He, R. Lin, W. Wei, Y. Tang, and W. Zhang, "Graph embedding for scholar recommendation in academic social networks," Frontiers in Physics, vol. 9, p. 768006, 2021.

[11] Y. Wang, J. Lin, Y. Tang, C. Yuan, and L. Zhang, "Adaptive deep- walk and prior-enhanced graph neural network for scholar influence maximization in social networks," in CCF Conference on Computer Supported Cooperative Work and Social Computing. Springer, 2024, pp. 235–250.

[12] K. Lo, L. L. Wang, M. Neumann, R. Kinney, and D. S. Weld, "S2orc: The semantic scholar open research corpus," in Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020, pp. 4969–4983.

[13] R. Kinney, C. Anastasiades, R. Authur, I. Beltagy, J. Bragg, A. Bu- raczynski, I. Cachola, S. Candra, Y. Chandrasekhar, A. Cohan et al., "The semantic scholar open data platform," arXiv preprint arXiv:2301.10140, 2023.

[14] A. B. Arrieta, N. D´ıaz-Rodr´ıguez, J. Del Ser, A. Bennetot, S. Tabik,

A. Barbado, S. Garc´ıa, S. Gil-Lo´pez, D. Molina, R. Benjamins,

R. Chatila, and F. Herrera, "Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai," Information Fusion, vol. 58, pp. 82–115, 2020.

[15] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, "A survey on bias and fairness in machine learning," ACM Computing Surveys, vol. 54, no. 6, pp. 115:1–115:35, 2021.

[16] M. Baker, "1,500 scientists lift the lid on reproducibility," Nature, vol.

533, no. 7604, pp. 452–454, 2016.

[17] K. Balog, Y. Fang, M. de Rijke, P. Serdyukov, and L. Si, "Expertise retrieval," Foundations and Trends in Information Retrieval, vol. 6, no. 2–3, pp. 127–256, 2012.

[18] J. Priem, H. Piwowar, and R. Orr, "Openalex: A fully-open index of scholarly works, authors, venues, institutions, and concepts," arXiv preprint arXiv:2205.01833, 2022.

[19] M. Ley, "The dblp computer science bibliography: Evolution, research issues, perspectives," in String Processing and Information Retrieval, ser. Lecture Notes in Computer Science, vol. 2476. Springer, 2002, pp. 1–10.

[20] W. Ammar, D. Groeneveld, C. Bhagavatula, I. Beltagy, M. Craw- ford, D. Downey, J. Dunkelberger, A. Elgohary, S. Feldman, V. Ha et al., "Construction of the literature graph in semantic scholar," 2018, arXiv:1805.02262.

[21] M. Fa¨rber, D. Lamprecht, J. Krause, L. Aung, and P. Haase, "Semope- nalex: The scientific landscape in 26 billion rdf triples," in The Semantic Web – ISWC 2023. Springer Nature Switzerland, 2023, pp. 94–112.

[22] J. Beel, B. Gipp, S. Langer, and C. Breitinger, "Research-paper recom- mender systems: a literature survey," Int. J. Digit. Libr., vol. 17, no. 4, p. 305–338, Nov. 2016.

[23] C. K. Kreutz and R. Schenkel, "Scientific paper recommendation sys- tems: A literature review of recent publications," International Journal on Digital Libraries, vol. 23, pp. 335–369, 2022.

[24] Z. Zhang, B. G. Patra, A. Yaseen, J. Zhu, R. Sabharwal, K. Roberts,

T. Cao, and H. Wu, "Scholarly recommendation systems: A literature survey," Knowledge and Information Systems, vol. 65, no. 11, pp. 4433– 4478, 2023.

[25] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," IEEE Transactions on Knowledge and Data Engineering, vol. 17, no. 6, pp. 734–749, 2005.

[26] S. Zhang, L. Yao, A. Sun, and Y. Tay, "Deep learning based rec- ommender system: A survey and new perspectives," ACM Computing Surveys, vol. 52, no. 1, pp. 5:1–5:38, 2019.

[27] G. R. Lopes, M. M. Moro, L. K. Wives, and J. P. M. de Oliveira, "Col- laboration recommendation on academic social networks," in Advances in Conceptual Modeling – Applications and Challenges, ser. Lecture Notes in Computer Science, vol. 6413. Springer, 2010, pp. 190–199.

[28] B. Min, H. Ross, E. Sulem, A. P. B. Veyseh, T. H. Nguyen, O. Sainz,

E. Agirre, I. Heinz, and D. Roth, "Recent advances in natural language processing via large pre-trained language models: A survey," ACM Computing Surveys, vol. 56, no. 2, pp. 30:1–30:40, 2024.

[29] S. Ji, S. Pan, E. Cambria, P. Marttinen, and P. S. Yu, "A survey on knowledge graphs: Representation, acquisition, and applications," IEEE Transactions on Neural Networks and Learning Systems, vol. 33, no. 2, pp. 494–514, 2022.

[30] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal,

H. K"uttler, M. Lewis, W.-t. Yih, T. Rockt"aschel, S. Riedel, and

D. Kiela, "Retrieval-augmented generation for knowledge-intensive nlp tasks," in Advances in Neural Information Processing Systems, vol. 33, 2020, pp. 9459–9474.

[31] Y. Gao, Y. Xiong, X. Gao, K. Jia, J. Pan, Y. Bi, Y. Dai, J. Sun,

M. Wang, and H. Wang, "Retrieval-augmented generation for large language models: A survey," 2023, arXiv:2312.10997.

[32] W. Fan, Y. Ding, L. Ning, S. Wang, H. Li, D. Yin, T.-S. Chua, and

Q. Li, "A survey on rag meeting llms: Towards retrieval-augmented large language models," in Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2024, pp. 6491–6501.

[33] S. Pan, L. Luo, Y. Wang, C. Chen, J. Wang, and X. Wu, "Unifying large language models and knowledge graphs: A roadmap," IEEE Transactions on Knowledge and Data Engineering, vol. 36, no. 7, pp. 3580–3599, 2024.

[34] W. Fang, G. Chen, C. Yuan, R. Lin, S. Tang, and Y. Tang, "KGCRAG: An adaptive community detection framework for robust graph-enhanced RAG," in Database Systems for Advanced Applications, ser. Lecture Notes in Computer Science, vol. 16537. Singapore: Springer, 2026, pp. 574–588.

[35] Y. Wang, W. Ma, M. Zhang, Y. Liu, and S. Ma, "A survey on the fairness of recommender systems," ACM Trans. Inf. Syst., vol. 41, no. 3, Feb. 2023.