

# Visual Exploration of Literature with ARGO SCHOLAR

Kevin Li  
Georgia Institute of  
Technology  
kevin.li@gatech.edu

Haoyang Yang  
Georgia Institute of  
Technology  
alexanderyang@gatech.edu

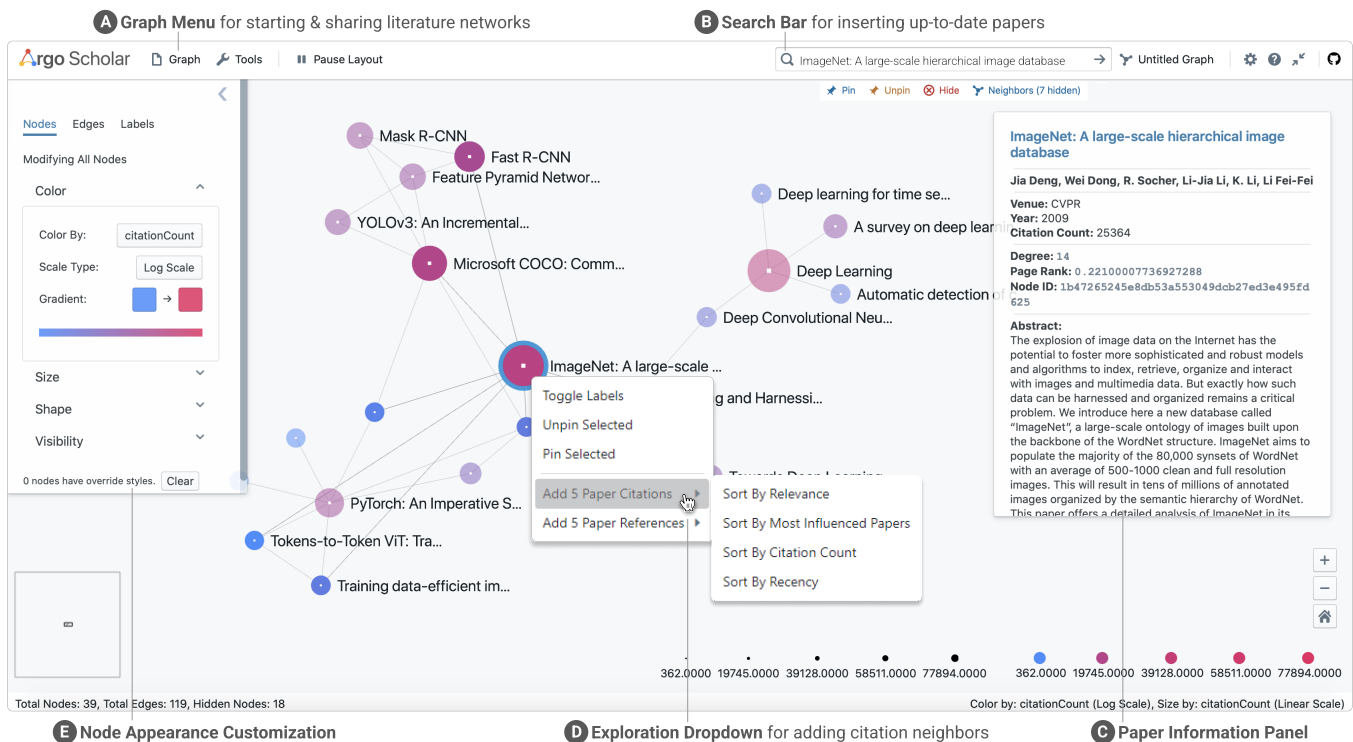
Evan Montoya  
Georgia Institute of  
Technology  
emontoya30@gatech.edu

Anish Upadhayay  
Georgia Institute of  
Technology  
aupadhayay3@gatech.edu

Zhiyan Zhou  
Georgia Institute of  
Technology  
zzhou406@gatech.edu

Jon Saad-Falcon  
Georgia Institute of  
Technology  
jonsaadfalcon@gatech.edu

Duen Horng Chau  
Georgia Institute of  
Technology  
polo@gatech.edu



**Figure 1:** ARGO SCHOLAR visualizing a literature network of deep learning papers; nodes are papers and edges are citations. (A) *Graph Menu* for starting new explorations, and saving and sharing user’s literature exploration snapshots as URLs. (B) *Search Bar* for adding papers via keyword querying. (C) *Paper Information Panel* displays key information (e.g., paper title, abstract, citation count). (D) *Literature Exploration Dropdown* for incrementally adding more citations or references. (E) *Visualization Customization Panel* for modifying visual properties of the graph of papers (e.g., color, size, labeling).

## ABSTRACT

Discovering and making sense of relevant literature is fundamental in any scientific field. Visualization can aid this process; however, existing tools’ adoption and impact are often constrained, such as by their reliance on small, curated paper datasets that quickly

become outdated or a lack of support for personalized exploration. We introduce ARGO SCHOLAR, an open-source visualization tool designed for interactive exploration of literature and easy sharing of exploration results. ARGO SCHOLAR is tightly integrated with Semantic Scholar’s live data of over 200 million papers, enabling users to generate personalized literature exploration results in real-time through flexible, incremental exploration, a common and effective method for researchers to discover relevant work. A large-scale user study of 122 participants from diverse backgrounds and experiences showed that ARGO SCHOLAR is effective at helping users



This work is licensed under a Creative Commons Attribution International 4.0 License.

CIKM '22, October 17–21, 2022, Atlanta, GA, USA  
© 2022 Copyright held by the owner/author(s).  
ACM ISBN 978-1-4503-9236-5/22/10.  
<https://doi.org/10.1145/3511808.3557177>

find related work and understand paper connections, and incremental graph-based exploration is effective across diverse disciplines. Developed using modern web technologies, our tool is widely available at <https://poloclub.github.io/argo-scholar> and allows users to easily share their literature search results as a URL.

## CCS CONCEPTS

• **Human-centered computing** → **Visualization systems and tools**; • **Information systems** → **Open source software**.

## KEYWORDS

Graph visualization, literature explorer, user interface design

### ACM Reference Format:

Kevin Li, Haoyang Yang, Evan Montoya, Anish Upadhayay, Zhiyan Zhou, Jon Saad-Falcon, and Duen Horng Chau. 2022. Visual Exploration of Literature with ARGO SCHOLAR. In *Proceedings of the 31st ACM Int'l Conference on Information and Knowledge Management (CIKM '22)*, Oct. 17–21, 2022, Atlanta, GA, USA. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3511808.3557177>

## 1 INTRODUCTION

Discovering and making sense of relevant literature is fundamental in any scientific field. Academic search engines [2, 19] are common starting point [6, 12]; however, keyword searching does not provide users with all relevant papers [26], so incremental exploration, a common and effective method where users iterate from familiar (identified via search engines) to novel works while filtering based on paper attributions [6, 21, 25, 29], is also employed [23, 24]. Visualization can aid this process [9, 11, 30], such as node-link diagram-based tools that support the visual exploration of citation networks with a bottom-up approach that allows users to grow their own networks and understanding of its papers [1, 3, 7, 15, 25].

However, existing tools' adoption and impact are often restricted by a variety of drawbacks. Some tools are constrained by their reliance on manually-curated paper datasets that were often small and subsets of the available literature that quickly became outdated [7, 15]. Others do not support incremental exploration [1]. Existing tools often also do not provide users with the ability to organize network exploration results to align with their evolving mental schemas [1, 3, 15], potentially hindering their sensemaking process [22]. Another challenge tools face is their barriers to access as many are downloaded applications whose availability are subject to computer system constraints [28], or are available only to user study participants [25]. Addressing these common literature exploration needs, we contribute:

- **ARGO SCHOLAR, an open-source, web-based visualization tool for interactive exploration of literature and easy sharing of exploration results** (Figure 1). Anyone can visualize their literature networks by accessing ARGO SCHOLAR using their web browser on desktops or laptops without the need for any software installation or additional work associated with similar tools [7, 25]. ARGO SCHOLAR is developed using modern web technologies and is open-sourced under the permissive MIT license at <https://github.com/poloclub/argo-scholar> and is available for use at the public link: <https://poloclub.github.io/argo-scholar>.

For a demo video of our tool, please visit <https://youtu.be/z2RgSloW9P4>.

- **Incremental, flexible literature exploration with up-to-date data.** ARGO SCHOLAR connects to the live data of Semantic Scholar [5, 19] via its API, gaining access to its 205 million indexed papers (as of April 2022). ARGO SCHOLAR's ability to query and visualize up-to-date data in real-time overcomes major limitations faced by prior tools proposed for literature sensemaking that relied on manually-curated paper datasets that were often small and quickly became outdated [7, 15, 28]. Users can add papers indexed by Semantic Scholar by either querying keywords related to their interests, entering a paper's unique CorpusID (Figure 1B), or exploring citations and references based on select properties such as citation count or relevance (Figure 1D).
- **Easy sharing of literature exploration results.** ARGO SCHOLAR allows users to share as many snapshots of their literature networks and visualizations as a URL or IFrame web-embedded application (Figure 1A) stored on the cloud at no cost to users. By accessing a shared link, researchers can build upon existing exploration progress to discover more relevant papers and grow their body of knowledge [6]. JSON snapshots that store the information of explorer papers and their connections can also be embedded in Jupyter Notebooks [16] or blog posts and personal websites through HTML IFrames.
- **Large-scale user study with 122 participants.** We evaluated ARGO SCHOLAR with participants from much more diverse backgrounds and research experiences and at a scale a magnitude larger than previous studies of literature exploration tools [7, 25]. Our study showed that our tool is effective at helping users find related work and understand connections among papers, and incremental graph-based exploration is effective across disciplines.

## 2 ARGO SCHOLAR SYSTEM DESIGN

ARGO SCHOLAR builds on our early prototype [17]. The following sections describe how ARGO SCHOLAR's system components support the incremental finding and understanding of relevant research.

### 2.1 Initializing and Displaying Paper Relations

ARGO SCHOLAR's main view displays current papers as a citation network in the form of a node-link diagram, where nodes represent papers and citations represent edges. The *Graph Menu* (Figure 1A) allows users to generate a new citation network, resume a previously saved snapshot with the same positioning and user customization, or build upon different samples of pre-generated citation networks. Through the *Search Bar* (Figure 1B), users can search for and add any of the 200 million plus papers indexed by Semantic Scholar by querying keywords or a paper title or inputting a specific paper's unique CorpusID [19]. Users can add new papers into their citation network at any point in their exploration process, allowing users to expand their indexed papers without restriction or even locate intersectional papers. Edge directionality (from citing to cited) can be toggled for a more detailed overview of the relationships among and within clusters of publications. ARGO SCHOLAR is able to smoothly render hundreds of papers and their relationships due to being built with WebGL to render graphs.

## 2.2 Incremental Exploration

Due to the effectiveness of incremental exploration as a technique used when discovering relevant work [7, 15, 30], ARGO SCHOLAR supports this literature sensemaking strategy through the exploration of current papers' citations and references. Users can right-click on a given paper multiple times to iteratively add five more of its citations (sorted by relevance, most influenced<sup>1</sup>, citation count, or recency) or references (by relevance, most influenced, or recency) [5] each time via the *Exploration Dropdown* (Figure 1D) menu<sup>2</sup>. To improve label readability, newly-added papers are aligned vertically, with citations on the right side of the selected paper, and references on the left. Citation edges between non-selected and newly-added papers are automatically added to the visualization. Users can also prune their networks by hiding unwanted papers.

## 2.3 Accessing Paper Information and Customizing Graph Visualizations

The *Paper Information Panel* (Figure 1C) shows key information for a highlighted or hovered-over paper, e.g., paper title, abstract, authors, citation count, publication venue, publication year, and the URL to its Semantic Scholar page (for full paper PDF, figures, etc.). Such information helps researchers in literature sensemaking [14]. A variety of visual properties of the nodes, edges, and node labels can be customized via the *Customization Panel* (Figure 1E). Nodes can be colored and sized through global rules based on paper attributes such as citation count or graph statistics such as Pagerank, and users can specify the size, color, label, and shape of individual or groups of papers to facilitate the compartmentalization of papers. The force-directed algorithm can also be toggled to help generate aesthetically-pleasing layouts of visible nodes.

## 2.4 Saving and Sharing Progress

ARGO SCHOLAR allows users to save and share their literature exploration as a *snapshot* that stores all the literature network data (e.g., explored papers, their connections, and attributes) and visualization customizations (e.g., node positioning and visibility, color and size rules, labels) via the *Graph Menu* (Figure 1A) to support collaboration [26]. Users have two options when saving and sharing their exploration snapshots: locally on their device or in the cloud. If users opt for the first option, they will download the JSON file and then upload it to ARGO SCHOLAR whenever they wish to resume their sensemaking progress. Or, users can save their snapshots to ARGO SCHOLAR's server for free, enabling them to share their literature network as a custom URL, bypassing the need to save and store JSON files. Saving the snapshot to ARGO SCHOLAR's servers also grants the ability to embed the literature network as either an HTML IFrame or a Jupyter Notebook IFrame [16].

## 2.5 Open-source Implementation

ARGO SCHOLAR is open-sourced and hosted at <https://poloclub.github.io/argo-scholar> as a Github page and runs on all modern web browsers (e.g., Chrome, Safari, Firefox, Edge), broadening its access. Developed using modern web technologies, ARGO SCHOLAR

provides users with new features critical for literature sensemaking (e.g., incremental exploration, live connection to Semantic Scholar's Corpus API endpoint [19]) while inheriting Argo Lite's core graph visualization capabilities built with React, Blueprint for UI, MobX state management, and Three.js WebGL graph rendering [18]. The Github page includes documentation on how to develop and deploy ARGO SCHOLAR, enabling tailored use-cases built on top of the existing application. Users are encouraged to open pull requests and Github issues with feature suggestions and implementations. ARGO SCHOLAR's open-sourced nature is one primary differentiating factor from other existing tools and aims to provide higher levels of accessibility and ease of use than current paper visualization tools.

## 3 EVALUATION

We conducted a large-scale user study with 122 participants to evaluate ARGO SCHOLAR's usability and investigate how participants interact with the features offered to explore literature networks and find relevant work. We recruited U.S.-based participants from Prolific, an online crowdsourcing platform designed specifically for academic research. The participants had diverse knowledge backgrounds and research experiences (Table 1). They had at least one year of research experience. 76 are current students, and 46 are not currently enrolled. For the highest education level completed (or currently completing), 21 listed Doctorate, 42 Master's, 44 Bachelor's, 12 high school graduates or the equivalent, and 3 others (JD, MD, and DDS). Notably, 3 participants were professors, and 2 were postdoctoral researchers. All participation was voluntary and anonymous. The study was approved by Georgia Tech's IRB, with survey responses and logged tool interactions stored in accordance with official institute policy.

### 3.1 Procedure

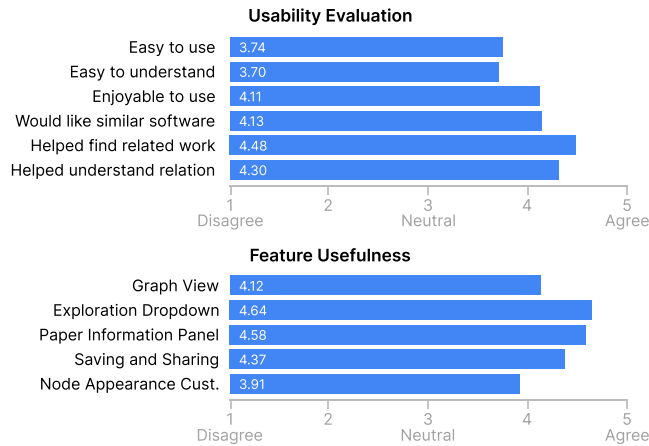
After participants signed their consent forms electronically and filled out their background questionnaire, they were provided with a short tutorial video (1 minute 23 seconds) highlighting the features that ARGO SCHOLAR provides. Participants were then asked to answer two simple multiple-choice questions to ensure an understanding of the instructions. Those who answered both correctly within three attempts were allowed to continue the study. Participants were then asked to use the tool for at least 10 minutes to build their own literature network, based on a topic of their choosing. Their resulting network must have included at least 10 relevant papers, and they were encouraged to include more. After exploration, the participants submitted their result snapshot as a URL (as described in subsection 2.4). Finally, participants were asked to fill out a survey regarding the usability of the tool and its features. The study took no more than 20 minutes to complete; many participants chose to explore the tool for longer. Each qualified participant was compensated \$4 with a bonus of \$2 for networks that demonstrated a clear effort to use the tool to discover relevant papers.

### 3.2 Results and Lessons Learned

**Usability and usefulness are sustained at a large scale.** Figure 2 summarizes the results of the exit survey of 5-point Likert-scale questions regarding the utility and usefulness of ARGO SCHOLAR's features. Overall, the results were positive and very encouraging

<sup>1</sup>Citations that indicate that the cited work is used or extended in the new effort [27]

<sup>2</sup>Semantic Scholar supports 100 requests per 5 minutes per IP address



**Figure 2: Average ratings from 122 participants regarding the usability and usefulness of ARGO SCHOLAR. Most participants found ARGO SCHOLAR provided an easy and enjoyable way to find related work and understand relations between papers and rated the Exploration Dropdown, Paper Information Panel, and Saving and Sharing features the most favorably.**

given that our study, with 122 participants, is at a scale that is significantly larger than previous studies [7, 25] and that ARGO SCHOLAR was not designed as a walk-up-and-use system. Participants favorably rated ARGO SCHOLAR as enjoyable to use and that they would like to use software similar to ARGO SCHOLAR in the future. They were particularly positive about ARGO SCHOLAR’s ability to help them find related work (at 4.48/5) and understand how papers are related to each other (4.30/5). All key features of ARGO SCHOLAR were rated above a 4 except for the *Node Appearance Customization* panel. *Exploration Dropdown* was rated the highest (at 4.64/5). One general trend observed was that current students rated the tool higher for almost all categories compared to those who were not currently enrolled. We hypothesize this may be due to a greater need or desire for such a research tool. Participants in the humanities sometimes commented on a lack of papers available, which is indicative of the topics that Semantic Scholar may skew towards.

**Incremental graph-based exploration is effective for diverse disciplines.** Due to ARGO SCHOLAR’s web-based deployment and live connection to the 200M papers on Semantic Scholar, we are able to evaluate the literature search approach based on incremental exploration with participants from much more diverse backgrounds and research experiences (Table 1) and at a scale that is a magnitude larger than previous studies [7, 13, 25]. Our results reaffirm that the incremental exploration technique generally works for participants across a wide range of disciplines and helps them find related work and understand relationships between such articles. For example, a *biological sciences* participant commented “...it helped me to find some relevant papers I hadn’t come across yet. This site could be very helpful for my dissertation.” This notion was supported by a *speech-language pathology* participant: “I have never been able to map out citations and references for research papers so quickly in an organized fashion.”

**No one-size-fits-all tool for literature exploration.** Many participants enjoyed the node-link diagram display of their citation

**Table 1: Participant background organized by National Science Foundation (NSF) research areas [4]**

NSF Research Category	No.
Social, Behavioral and Economic Sciences (SBE)	51
Biological Sciences (BIO)	24
Computer and Information Science and Engineering (CISE)	21
Engineering (ENG)	9
Mathematical and Physical Sciences (MPS)	6
Education and Human Resources (EHR)	5
Integrative Activities (OLA)	3
Environmental Research and Education (ERE)	2
Geosciences (GEO)	1

networks. For example, a *Doctor of Medicine* participant proclaimed “Love the node aspect and makes showing connections to other papers super easy!” And a *professor* participant commended ARGO SCHOLAR for being “...a nice tool to visualize the literature and I could see myself using this and showing it to graduate students to help them with their literature reviews.” Some were not as receptive to such visualization: “This type of visualization is really not how my mind works. I can see that it could be very powerful for some people, though”. A few participants also noted that ARGO SCHOLAR “seemed a little difficult to use at first, but after a few minutes ... everything become more enjoyable”; however, it is encouraging that participants who did face a learning curve still found the system useful and enjoyable.

**Novel feature requests and inspirations for literature exploration tools.** Many participants expressed explicit interest in a novel feature that supports “make my own ... lines to connect some of my own connections” to express their views of how papers are related. Participants also asked for a “sticky note function” prevalent in only a handful of tools [10, 25]. We also note the high rating and feedback (e.g., *Very easy to publish ... Can save lots of time working with my colleagues*) and suggest that sharing functions could be valuable additions to literature exploration tools. These suggestions reinforce the need for certain functions and provide inspirations for future work. Many of these suggestions were not reported in previous research studies [7, 20, 25], and we attribute such findings to the increased scale and diversity provided by our study [8].

## 4 CONCLUSION AND DISCUSSION

We introduce ARGO SCHOLAR, an open-sourced, web-based visualization tool that allows users to interactively and incrementally explore literature to gain insight and uncover relevant work. Users can access our tool using their web browser without the need for any software installation, easily share their exploration results, and connect directly to Semantic Scholar’s live database of more than 200 million papers. Through large-scale user evaluations, we show the effectiveness of ARGO SCHOLAR in helping users understand paper connections within a field while enabling freedom in exploring further and distilled lessons learned from our study. We note that our usage of Semantic Scholar’s database may result in the inherent increased availability of papers in the sciences for our tool, so future work will be focused on finding sources that supplement humanities research areas and implementing our participants’ suggestions. We look forward to more students, researchers, and practitioners adopting ARGO SCHOLAR to help their literature reviews and searches and make sense of literature across disciplines.



## REFERENCES

- [1] 2021. ConnectedPapers | Find and explore academic papers. <https://www.connectedpapers.com/>. Accessed: 2022-03-17.
- [2] 2021. Google Scholar. <https://scholar.google.com/>. Accessed: 2022-04-11.
- [3] 2022. LitMaps - Discover Science Faster. <https://www.litmaps.co/>. Accessed: 2022-03-17.
- [4] 2022. NSF Research Areas. [https://www.nsf.gov/about/research\\_areas.jsp](https://www.nsf.gov/about/research_areas.jsp). Accessed: 2022-04-27.
- [5] Waleed Ammar, Dirk Groeneveld, Chandra Bhagavatula, Iz Beltagy, Miles Crawford, Doug Downey, Jason Dunkelberger, Ahmed Elgohary, Sergey Feldman, Vu A. Ha, Rodney Michael Kinney, Sebastian Kohlmeier, Kyle Lo, Tyler C. Murray, Hsu-Han Ooi, Matthew E. Peters, Joanna L. Power, Sam Skjonsberg, Lucy Lu Wang, Christopher Wilhelm, Zheng Yuan, Madeleine van Zuylen, and Oren Etzioni. 2018. Construction of the Literature Graph in Semantic Scholar. In *NAACL*.
- [6] Kumaripaba Athukorala, Eve E. Hoggan, Anu Lehtiö, Tuukka Ruotsalo, and Giulio Jacucci. 2013. Information-seeking behaviors of computer scientists: Challenges for electronic literature search tools. In *ASIST*.
- [7] Duen Horng Chau, Aniket Kittur, Jason I. Hong, and Christos Faloutsos. 2011. Apolo: making sense of large network data by combining rich user interaction and machine learning. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2011).
- [8] Paolo Federico, Florian Heimerl, Steffen Koch, and Silvia Miksch. 2017. A Survey on Visual Approaches for Analyzing Scientific Literature and Patents. *IEEE Transactions on Visualization and Computer Graphics* 23 (2017), 2179–2198.
- [9] George W. Furnas and Daniel M. Russell. 2005. Making sense of sensemaking. *CHI '05 Extended Abstracts on Human Factors in Computing Systems* (2005).
- [10] Robert J. Gove, Cody Dunne, Ben Shneiderman, Judith L. Klavans, and B. Dorr. 2011. Evaluating visual and statistical exploration of scientific literature networks. *2011 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)* (2011), 217–224.
- [11] Nitesh Goyal, Gilly Leshed, and Susan R. Fussell. 2013. Effects of visualization and note-taking on sensemaking and analysis. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2013).
- [12] Michael Gusenbauer. 2018. Google Scholar to overshadow them all? Comparing the sizes of 12 academic search engines and bibliographic databases. *Scientometrics* 118 (2018), 177–214.
- [13] Jiangen He, Q. Ping, Wen Lou, and Chaomei Chen. 2019. PaperPoles: Facilitating adaptive visual exploration of scientific publications by citation links. *Journal of the Association for Information Science and Technology* 70 (2019).
- [14] Emi Ishita, Yasuko Hagiwara, Yukiko Watanabe, and Yoichi Tomiura. 2018. Which Parts of Search Results do Researchers Check when Selecting Academic Documents? *Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries* (2018).
- [15] Sanjay Ram Kairam, Nathalie Henry Riche, Steven Mark Drucker, R. Fernandez, and Jeffrey Heer. 2015. Refinery: Visual Exploration of Large, Heterogeneous Networks through Associative Browsing. *Computer Graphics Forum* 34 (2015).
- [16] Thomas Kluyver, Benjamin Ragan-Kelley, Fernando Pérez, Brian E. Granger, Matthias Bussonnier, Jonathan Frederic, Kyle Kelley, Jessica B. Hamrick, Jason Grout, Sylvain Corlay, Paul Ivanov, Damián Avila, Safia Abdalla, Carol Willing, and Jupyter Development Team. 2016. Jupyter Notebooks - a publishing format for reproducible computational workflows. In *ELPUB*.
- [17] Kevin Li, H. Yang, Anish Upadhayay, Zhiyan Zhou, Jon Saad-Falcon, and Duen Horng Chau. 2021. Argo Scholar: Interactive Visual Exploration of Literature in Browsers. *ArXiv abs/2110.14060* (2021).
- [18] Siwei Li, Zhiyan Zhou, Anish Upadhayay, Omar Shaikh, Scott Freitas, Haekyu Park, Zijie Jay Wang, Susanta Routray, Matthew Hull, and Duen Horng Chau. 2020. Argo Lite: Open-Source Interactive Graph Exploration and Visualization in Browsers. *Proceedings of the 29th ACM International Conference on Information & Knowledge Management* (2020).
- [19] Kyle Lo, Lucy Lu Wang, Mark Neumann, Rodney Michael Kinney, and Daniel S. Weld. 2020. S2ORC: The Semantic Scholar Open Research Corpus. In *ACL*.
- [20] Rina Nakazawa, Takayuki Itoh, and Takafumi Saito. 2015. A Visualization of Research Papers Based on the Topics and Citation Network. *2015 19th International Conference on Information Visualisation* (2015), 283–289.
- [21] Robert S. Pienta, James Abello, Minsuk Kahng, and Duen Horng Chau. 2015. Scalable graph exploration and visualization: Sensemaking challenges and opportunities. *2015 International Conference on Big Data and Smart Computing (BIGCOMP)* (2015), 271–278.
- [22] Peter Piroli and Stuart Card. 2005. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of international conference on intelligence analysis*, Vol. 5. McLean, VA, USA, 2–4.
- [23] Napol Rachatasumrit, Jonathan Bragg, Amy X. Zhang, and Daniel S. Weld. 2022. CiteRead: Integrating Localized Citation Contexts into Scientific Paper Reading. *27th International Conference on Intelligent User Interfaces* (2022).
- [24] Hannah Snyder. 2019. Literature review as a research methodology: An overview and guidelines. *Journal of Business Research* (2019).
- [25] Nicole Sultanum, Christine Murad, and Daniel J. Wigdor. 2020. Understanding and Supporting Academic Literature Review Workflows with LitSense. *Proceedings of the International Conference on Advanced Visual Interfaces* (2020).
- [26] Sanna Talja, Pertti Vakkari, Jenny Fry, and Paul Wouters. 2007. Impact of research cultures on the use of digital library resources. *Journal of the Association for Information Science and Technology* 58 (2007), 1674–1685.
- [27] Marco Valenzuela, Vu A. Ha, and Oren Etzioni. 2015. Identifying Meaningful Citations. In *AAAI Workshop: Scholarly Big Data*.
- [28] Nees Jan van Eck and Ludo Waltman. 2014. CitNetExplorer: A new software tool for analyzing and visualizing citation networks. *J. Informetrics* 8 (2014), 802–823.
- [29] Claes Wohlin. 2014. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In *EASE '14*.
- [30] Xiaolong Luke Zhang, Yan Qu, C. Lee Giles, and Piyou Song. 2008. CiteSense: supporting sensemaking of research literature. In *CHI*.