Bursting Scientific Filter Bubbles: Boosting Innovation via Novel Author Discovery

Anonymous Author(s)
Affiliation
Address
email

Abstract

Isolated silos of scientific research and the growing challenge of information overload limit awareness across the literature and hinder innovation. Algorithmic curation and recommendation, which often prioritize relevance, can further reinforce these informational "filter bubbles." In response, we describe Bridger, a system for facilitating discovery of scholars and their work, to explore design tradeoffs between relevant and novel recommendations. We construct a faceted representation of authors with information gleaned from their papers and inferred author personas, and use it to develop an approach that locates commonalities ("bridges") and contrasts between scientists — retrieving partially similar authors rather than aiming for strict similarity. In studies with computer science researchers, this approach helps users discover authors considered useful for generating novel research directions, outperforming a state-of-art neural model. In addition to recommending new content, we also demonstrate an approach for *displaying* it in a manner that boosts researchers' ability to understand the work of authors with whom they are unfamiliar. Finally, our analysis reveals that Bridger connects authors who have different citation profiles, publish in different venues, and are more distant in social co-authorship networks, raising the prospect of bridging diverse communities and facilitating discovery.

1 Introduction

> "Opinion and behavior are more homogeneous within than between groups… Brokerage across structural holes provides a vision of options otherwise unseen." (Burt, 2004)

The volume of papers in computer science continues to sky-rocket, with the DBLP computer science bibliography listing hundreds of thousands of publications in the year 2020 alone.[1] In particular, the field of AI has seen a meteoric growth in recent years, with new authors entering the field every hour [27]. Researchers rely largely on search and recommendation services like Google Scholar and Semantic Scholar to keep pace with the growing literature and the authors who contribute to it. The literature retrieval services algorithmically decide what information to serve to scientists [1, 5], using information such as citations and textual content as well as behavioral traces such as clickthrough data, to inform machine learning models that output lists of ranked papers or authors.

By relying on user behavior and queries, these services adapt and reflect human input and, in turn, influence subsequent search behavior. This cycle of input, updating, engagement, and response can lead to an amplification of biases around searchers' prior awareness and knowledge [12]. Such biases include selective exposure [7], homophily [16], and the aversion to information from novel domains

[1]https://dblp.org/statistics/publicationsperyear.html

Submitted to NeurIPS 2021 AI for Science Workshop.

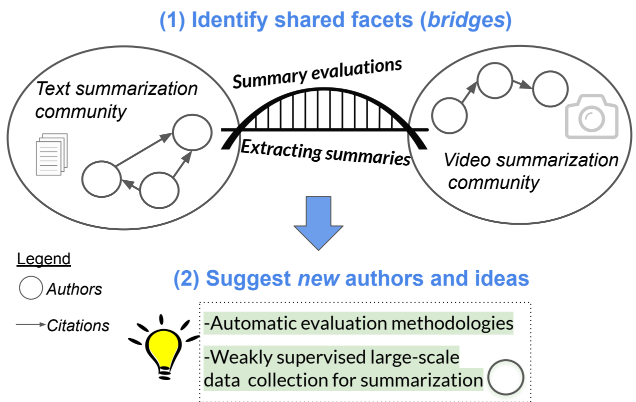

Figure 1: *Bursting scientific bubbles with Bridger.* The overarching goal is to (1) find commonalities among authors working in different areas and unaware of one another, and (2) suggest novel and valuable authors and their work, unlikely discovered otherwise due to their disparities.

that require more cognitive effort to consider [10, 13]. By reinforcing these tendencies, systems that filter and rank information run the risk of engendering so-called *filter bubbles* [21] that fail to show users novel content outside their narrower field of interest.

These bubbles and silos of information can be costly to individual researchers and for the evolution of science as a whole. They may lead scientists to concentrate on narrower niches [14], reinforcing citation inequality and bias [20] and limiting cross-fertilization among different areas that could catalyze innovation [10, 13, 11]. Addressing filter bubbles in general, in domains such as social media and e-commerce recommendations, is a hard and unsolved problem [8, 4, 32]. The problem is especially difficult in the scientific domain. The scientific literature consists of complex models and theories, specialized language, and an endless diversity of continuously emerging concepts. Connecting blindly across these cultural boundaries requires significant cognitive effort [28], translating to time and resources most researchers are unlikely to have to enter unfamiliar research territory.[2]

Our vision in this paper is to develop an approach that **boosts scientific innovation and builds bridges across scientific communities**, by helping scientists **discover authors that spark new ideas** for research. Working toward this goal, we developed Bridger, illustrated in Figure 1. Our main contributions include:

- **A multidimensional author representation for matching authors along specific facets**. Our novel representation includes information extracted automatically from papers, including tasks, methods and resources, and automatically inferred *personas* that reflect the different focus areas on which each scientist works. Each of these aspects is embedded in a vector space based on its content, allowing the system to *identify authors with commonalities along specific dimensions* and not others, such as authors working on similar tasks but not using similar methods.
- **Boosting discovery of useful authors and ideas from novel areas.** We explore the utility of our author representation in experiments with computer science researchers interacting with Bridger. We find that this representation helps users discover authors considered novel *and* relevant, *assisting users in finding potentially useful research directions*. Bridger outperforms a strong neural model currently employed by a public scholarly search engine for search and recommendation[3]— despite Bridger's focus on surfacing *novel* content and the built-in biases associated with this novelty. We conduct interviews with researchers, studying the tradeoffs between novelty and relevance in scientific content recommendations and discussing challenges for author discovery systems.
- **Exploring how to effectively depict recommended authors.** In addition to assessing *what* authors to recommend to spark new research ideas, we also consider *how* to display authors in a way that enables users to rapidly understand what new authors work on. We employ Bridger as an experimental platform to explore which facets are displayed to users, investigating various design

[2]The challenge of limited time to explore novel directions is also discussed in our interviews with researchers; see §D.

[3]`https://twitter.com/SemanticScholar/status/1267867735318355968`.

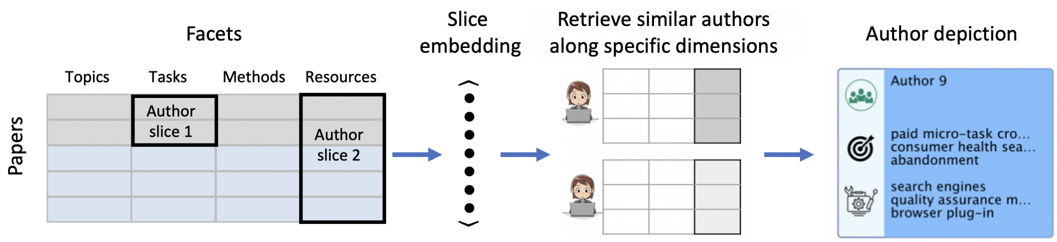

Figure 2: Bridger's author representation, retrieval, and depiction. Users are represented in terms of a matrix with rows corresponding to papers, and columns corresponding to facets. Bridger finds suggested authors who match along "slices" of the user's data – facets, subsets of papers, or both.

choices and tradeoffs. We obtain substantially better results in terms of user *understanding of profiles of unknown authors*, when displaying information taken from our author representation.

- **Evidence of bridging across research communities.** Finally, we conduct in-depth analyses revealing that Bridger surfaces novel and valuable authors and their work that are unlikely to be discovered in the absence of Bridger due to publishing in different venues, citing and being cited by non-overlapping communities, and having greater distances in the social co-authorship network.

Taken together, the ability to uncover novel and useful authors and ideas, and to serve this information to users in an effective and intuitive manner, suggests a future where automated systems are put to work to build bridges across communities, rather than blindly reinforcing existing filter bubbles.

2 Bridger: Approach Overview

In this section we present a brief overview of our novel faceted representation of authors, and methods for using this representation for author discovery by matching researchers along specific dimensions (Figure 2). See Appendix for full details.

2.1 Author representations

We represent an author, $\mathcal{A}$, as a set of *personas* capturing the multiple themes an author can work on in different papers, in which each persona is encoded with facet-wide aggregations of term embeddings across a set of papers. Terms are spans of text referring to *methods, tasks and resources* automatically extracted from each paper i. Each term t is located in a "cell" in the matrix illustrated in Figure 2, that depicts outlines of "slices" in bold — subsets of rows and columns in the illustrated matrix, corresponding to personas (subsets of rows) and facets (columns). We aggregate the terms into author-level *facets* that capture different aspects of $\mathcal{A}$: $\mathcal{V}_{\mathcal{A}} = \{\mathbf{m}, \mathbf{t}, \mathbf{r}\}$, where $\mathbf{m}$ is an aggregate embedding of $\mathcal{A}$'s *method* facets, $\mathbf{t}$ is an embedding capturing $\mathcal{A}$'s *tasks*, and $\mathbf{r}$ represents $\mathcal{A}$'s *resources*.

2.2 Approaches for recommending authors

For a given author $\mathcal{A}$, we are interested in automatically suggesting new authors working on areas that are relevant to $\mathcal{A}$ but also likely to be interesting and spark new ideas.

Commonalities and contrasts model We explore a formulation of the author discovery problem in terms of matching authors along specific dimensions that allow more fine-grained control – such as by using only a subset of views in $\mathcal{V}_{\mathcal{A}}$, or only a subset of $\mathcal{A}$'s papers, or both — as in the row and column *slices* seen in Figure 2. This decomposition of authors also enables us to explore *contrasts* along specific author dimensions, e.g., finding authors who use similar tasks to $\mathcal{A}$ but use very different methods or resources.

2.3 Depicting Recommended Authors

Researchers, flooded with constant streams of papers, typically have a very limited attention span to consider whether some new author or piece of information is relevant to them. It is thus important that the information we display for each author (such as their main methods, tasks, resources, and also papers) is *ranked*, and that we provide users with rankings explaining how the retrieved authors relate to them. In systems that help people find authors, such as Microsoft Academic Graph, Google Scholar, and AMiner [30], authors are often described in terms of a few high-level topics. We show that methods, tasks and resources, when ranked, can greatly help understand what authors work on. See experimental findings in Appendix §B. We now present our main findings on author discovery.

3 Novel Author Discovery: Experiment Results

Twenty computer-science researchers participated in our experiment. We showed them recommended authors using two author-ranking strategies (§A.3), one based on similar tasks alone (`sT`) and the other on similar tasks with contrasting (distant) methods (`sTdM`). We compare these strategies to the SPECTER (`ss`) baseline, a strong neural representation for scientific paper retrieval [5].

We examine the proportion of users who preferred each of the `sT` and `sTdM` conditions in comparison to `ss`. The facet-based approaches lead to a boost despite comparing against an advanced baseline geared at relevance to which users are naturally primed. For the `sT` condition, 60 percent of participants preferred bridger author suggestions compared to 40 percent who preferred the specter author suggestions. For the `sTdM` condition, 78 percent preferred bridger. We also compare the results from `sT` and `sTdM` conditions based on personas P for user $\mathcal{A}$, versus the user's non-persona-based results presented above, finding them to further boost results.

3.1 Evidence of Bursting Bubbles

We empirically find that authors suggested by `sT` and `sTdM` tend to be very different than those suggested by `ss`, according to metrics measuring proximity to the user based on citation distances to the user, as well as publication venue distances and the shortest path length between the user and the matched author in the coauthorship graph. Findings of this analysis, shown in Figure 3, suggest that Bridger surfaces novel authors from more diverse, distant fields and research communities than SPECTER.

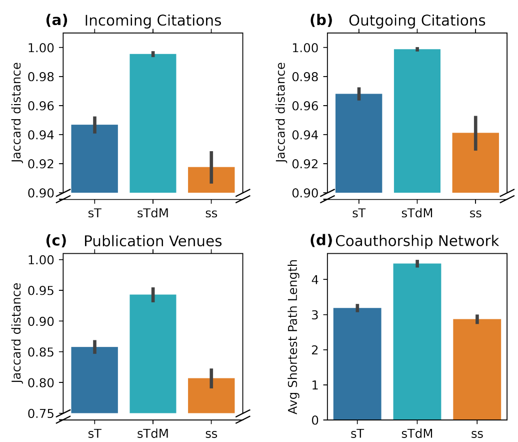

Figure 3: *Bridger suggests authors that are more likely to bridge gaps between communities in comparison to the baseline.* Clockwise: (a, b) Jaccard distance between suggested authors' papers and the user's papers for incoming citations (a) and outgoing citations (b); greater distance means that suggested authors are less likely to be cited by or cite the same work. (c) Jaccard distance for publication venues. (d) Shortest path length in the coauthorship graph between author and user.

4 Conclusion

We presented Bridger, a framework for facilitating discovery of novel and valuable scholars and their work. Bridger consists of a faceted author representation, allowing users to see authors who match them along certain dimensions (e.g., tasks) but not others. Bridger also provides "slices" of a user's papers, enabling them to find authors who match the user only on a subset of their papers, and only on certain facets within those papers. Our experiments with computer science researchers show that the facet-based approach was able to help users discover authors with work that is considered more interesting and novel, substantially more than a relevance-focused baseline representing state-of-art retrieval of scientific papers. Importantly, we show that authors surfaced by Bridger are indeed from more distant communities in terms of publication venues, citation links and co-authorship social ties. These results suggest a new and potentially promising avenue for mitigating the problem of isolated silos in science.

References

[1] Jöran Beel and Bela Gipp. Google scholar's ranking algorithm: an introductory overview. In *Proceedings of the 12th international conference on scientometrics and informetrics (ISSI'09)*, volume 1, pages 230–241. Rio de Janeiro (Brazil), 2009.

[2] Isabel Cachola, Kyle Lo, Arman Cohan, and Daniel S Weld. Tldr: Extreme summarization of scientific documents. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*, pages 4766–4777, 2020.

[3] Arie Cattan, Sophie Johnson, Daniel Weld, Ido Dagan, Iz Beltagy, Doug Downey, and Tom Hope. Scico: Hierarchical cross-document coreference for scientific concepts. *arXiv preprint arXiv:2104.08809*, 2021.

[4] Wanyu Chen, Pengjie Ren, Fei Cai, Fei Sun, and Maarten de Rijke. Improving end-to-end sequential recommendations with intent-aware diversification. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pages 175–184, 2020.

[5] Arman Cohan, Sergey Feldman, Iz Beltagy, Doug Downey, and Daniel Weld. SPECTER: Document-level Representation Learning using Citation-informed Transformers. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 2270–2282, Online, July 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-main.207. URL `https://www.aclweb.org/anthology/2020.acl-main.207`.

[6] Alessandro Epasto, Silvio Lattanzi, and Renato Paes Leme. Ego-Splitting Framework: from Non-Overlapping to Overlapping Clusters. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '17*, pages 145–154, Halifax, NS, Canada, 2017. ACM Press. ISBN 978-1-4503-4887-4. doi: 10.1145/3097983.3098054. URL `http://dl.acm.org/citation.cfm?doid=3097983.3098054`.

[7] Dieter Frey. Recent research on selective exposure to information. *Advances in experimental social psychology*, 19:41–80, 1986.

[8] Yingqiang Ge, Shuya Zhao, Honglu Zhou, Changhua Pei, Fei Sun, Wenwu Ou, and Yongfeng Zhang. Understanding echo chambers in e-commerce recommender systems. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2261–2270, 2020.

[9] Suchin Gururangan, Ana Marasović, Swabha Swayamdipta, Kyle Lo, Iz Beltagy, Doug Downey, and Noah A. Smith. Don't Stop Pretraining: Adapt Language Models to Domains and Tasks. In *ACL*, 2020. doi: 10.18653/v1/2020.acl-main.740.

[10] Tom Hope, Joel Chan, Aniket Kittur, and Dafna Shahaf. Accelerating innovation through analogy mining. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '17, pages 235–243. Association for Computing Machinery. ISBN 978-1-4503-4887-4. doi: 10.1145/3097983.3098038. URL `https://doi.org/10.1145/3097983.3098038`.

[11] Tom Hope, Ronen Tamari, Hyeonsu Kang, Daniel Hershcovich, Joel Chan, Aniket Kittur, and Dafna Shahaf. Scaling creative inspiration with fine-grained functional facets of product ideas. *arXiv e-prints*, pages arXiv–2102, 2021.

[12] Lanu Kim, Jevin D West, and Katherine Stovel. Echo chambers in science? In *American Sociological Association*, 2017.

[13] Aniket Kittur, Lixiu Yu, Tom Hope, Joel Chan, Hila Lifshitz-Assaf, Karni Gilon, Felicia Ng, Robert E. Kraut, and Dafna Shahaf. Scaling up analogical innovation with crowds and AI. 116(6):1870–1877. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1807185116. URL `https://www.pnas.org/content/116/6/1870`. Publisher: National Academy of Sciences Section: Social Sciences.

[14] Joel Klinger, Juan Mateos-Garcia, and Konstantinos Stathoulopoulos. A narrowing of ai research? *arXiv preprint arXiv:2009.10385*, 2020.

[15] Yi Luan, Luheng He, Mari Ostendorf, and Hannaneh Hajishirzi. Multi-Task Identification of Entities, Relations, and Coreference for Scientific Knowledge Graph Construction. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Brussels, Belgium, 2018. Association for Computational Linguistics. doi: 10.18653/v1/D18-1360. URL `http://aclweb.org/anthology/D18-1360`.

[16] Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1):415–444, 2001.

[17] Rada Mihalcea and Paul Tarau. TextRank: Bringing Order into Text. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, pages 404–411, Barcelona, Spain, July 2004. Association for Computational Linguistics. URL `https://www.aclweb.org/anthology/W04-3252`.

[18] Fionn Murtagh and Pierre Legendre. Ward's hierarchical agglomerative clustering method: which algorithms implement ward's criterion? *Journal of classification*, 31 (3):274–295, 2014.

[19] Mark Neumann, Daniel King, Iz Beltagy, and Waleed Ammar. ScispaCy: Fast and Robust Models for Biomedical Natural Language Processing. *undefined*, 2019. URL `/paper/ScispaCy%3A-Fast-and-Robust-Models-for-Biomedical-Neumann-King/de28ec1d7bd38c8fc4e8ac59b6133800818b4e29`.

[20] Mathias Wullum Nielsen and Jens Peter Andersen. Global citation inequality is on the rise. *Proceedings of the National Academy of Sciences*, 118(7), 2021.

[21] Eli Pariser. *The filter bubble: What the Internet is hiding from you*. Penguin UK, 2011.

[22] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830, 2011.

[23] Nils Reimers and Iryna Gurevych. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *EMNLP/IJCNLP*, 2019. doi: 10.18653/v1/D19-1410.

[24] Ariel S. Schwartz and Marti A. Hearst. A simple algorithm for identifying abbreviation definitions in biomedical text. In *Biocomputing 2003*, pages 451–462. WORLD SCIENTIFIC, December 2002. ISBN 978-981-238-217-7. doi: 10.1142/9789812776303_0042. URL `https://www.worldscientific.com/doi/abs/10.1142/9789812776303_0042`.

[25] Arnab Sinha, Zhihong Shen, Yang Song, Hao Ma, Darrin Eide, Bo-June (Paul) Hsu, and Kuansan Wang. An overview of microsoft academic service (MAS) and applications. pages 243–246. ACM Press. ISBN 978-1-4503-3473-0. doi: 10.1145/2740908.2742839. URL `http://dl.acm.org/citation.cfm?doid=2740908.2742839`.

[26] Shivashankar Subramanian, Daniel King, Doug Downey, and Sergey Feldman. S2and: A benchmark and evaluation system for author name disambiguation. *arXiv preprint arXiv:2103.07534*, 2021.

[27] Xuli Tang, Xin Li, Ying Ding, Min Song, and Yi Bu. The pace of artificial intelligence innovations: Speed, talent, and trial-and-error. *Journal of Informetrics*, 14(4):101094, 2020.

[28] Daril Vilhena, Jacob Foster, Martin Rosvall, Jevin West, James Evans, and Carl Bergstrom. Finding cultural holes: How structure and culture diverge in networks of scholarly communication. 1:221–238. ISSN 23306696. doi: 10.15195/v1.a15. URL `http://www.sociologicalscience.com/articles-vol1-15-221/`.

[29] David Wadden, Ulme Wennberg, Yi Luan, and Hannaneh Hajishirzi. Entity, Relation, and Event Extraction with Contextualized Span Representations. In *EMNLP/IJCNLP*, 2019. doi: 10.18653/v1/D19-1585.

[30] Huaiyu Wan, Yutao Zhang, Jing Zhang, and Jie Tang. Aminer: Search and mining of academic social networks. *Data Intelligence*, 1(1):58–76, 2019.

[31] Kuansan Wang, Zhihong Shen, Chiyuan Huang, Chieh-Han Wu, Darrin Eide, Yuxiao Dong, Junjie Qian, Anshul Kanakia, Alvin Chen, and Richard Rogahn. A Review of Microsoft Academic Services for Science of Science Studies. *Frontiers in Big Data*, 2, 2019. ISSN 2624-909X. doi: 10.3389/fdata.2019.00045. URL `https://www.frontiersin.org/articles/10.3389/fdata.2019.00045/full#h7`. Publisher: Frontiers.

[32] Ziwei Zhu, Jianling Wang, and James Caverlee. Measuring and mitigating item under-recommendation bias in personalized ranking systems. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 449–458, 2020.

A Appendix

In this section we present our novel faceted representation of authors, and methods for using this representation for author discovery by matching researchers along specific dimensions (Figure 2). We also present methods for depicting the recommended authors when showing them to users.

A.1 Paper representations

Paper Information Each paper P contains rich, potentially useful information. This includes raw text such as in a paper's abstract, incoming and out-going citations, publication date, venues, and more. One key representation we derive from each paper P is a vector representation $\tilde{P}$, using a state-of-art scientific paper embedding model. This neural model captures overall coarse-grained topical information on papers, shown to be powerful in clustering and retrieving papers [5].

Another key representation is based on fine-grained facets obtained from papers. Let $\mathcal{T}_{P_i} = \{t_1, t_2, \ldots\}$ be a set of *terms* appearing in paper i. Each term is associated with a specific *facet* (category). We consider several categories of terms in this paper: coarse-grained paper topics inferred from the text [31], and fine-grained spans of text referring to *methods, tasks and resources* — core aspects of computer science papers [3] — automatically extracted from paper i with a scientific named entity recognition model [29]. Each term t is located in a "cell" in the matrix illustrated in Figure 2, with facets corresponding to the columns and papers to rows. Each term $t \in \mathcal{T}_{P_i}$ is also embedded in a vector space using a neural language model (see §A.5), yielding a $\tilde{t}$ vector for each term.

A.2 Author representations

We represent an author, $\mathcal{A}$, as a set of *personas* in which each persona is encoded with facet-wide aggregations of term embeddings across a set of papers. Figure 2 illustrates this with outlines of "slices" in bold — subsets of rows and columns in the illustrated matrix, corresponding to personas (subsets of rows) and facets (columns).

Author personas Each author $\mathcal{A}$ can work in multiple areas. In our setting, this can be important for understanding the different interests of authors, enabling more control on author suggestions. We experiment with a clustering-based approach for constructing *personas*, $\mathrm{P}_{\mathcal{A}}$, based on inferring for each set of author papers $\mathcal{P}_{\mathcal{A}}$ a segmentation into K subsets reflecting a common theme — illustrated as subsets of rows in the matrix in Figure 2. We also experiment with a clustering based on the network of co-authorship collaborations in which $\mathcal{A}$ takes part. See §A.5 for details on clustering. As discussed later (§B), we find that the former approach in which authors are represented with clusters of papers elicits considerably better feedback from scholars participating in our experiments.

Co-authorship information Each paper P is in practice authored by multiple people, i.e., it can belong to multiple authors $\mathcal{A}$. Each author assumes a *position* k for a given paper, potentially reflecting the strength of affinity to the paper. As discussed below (§A.5), we make use of this affinity in determining what weight to assign terms $\mathcal{T}_{P_i}$ for a given paper and given author.

Author-level facets Finally, using the above information on authors and their papers, we construct multiple author-level *facets* that capture different aggregate aspects of $\mathcal{A}$. More formally, in this paper we focus our experiments on author facets $\mathcal{V}_\mathcal{A} = \{\mathbf{m}, \mathbf{t}, \mathbf{r}\}$, where $\mathbf{m}$ is an aggregate embedding of $\mathcal{A}$'s *method* facets, $\mathbf{t}$ is an embedding capturing $\mathcal{A}$'s *tasks*, and $\mathbf{r}$ represents $\mathcal{A}$'s *resources*. In addition, we also construct these facets separately for each one of the author's personas $\mathrm{P}_\mathcal{A}$ — corresponding to "slice embeddings" over subsets of rows and columns in the matrix illustrated in Figure 2. In analyses of our experimental results (§C), we also study other types of information such as citations and venues; we omit them from the formal notations to simplify presentation.

A.3 Approaches for recommending authors

For a given author $\mathcal{A}$ using Bridger, we are interested in automatically suggesting new authors working on areas that are relevant to $\mathcal{A}$ but also likely to be interesting and spark new ideas. We are given a user $\mathcal{A}$, their set of personas $\mathrm{P}_\mathcal{A}$, and for each persona its faceted representation $\mathcal{V}_\mathcal{A} = \{\mathbf{m}, \mathbf{t}, \mathbf{r}\}$. We are also given a large pool of authors across computer science, $\{\mathcal{A}_1, \mathcal{A}_2, \ldots\}$, from which we aim to retrieve author suggestions to show $\mathcal{A}$.

Baseline model We employ SPECTER [5], a strong neural model to which we compare, trained to capture overall topical similarity between papers based on text and citation signals (see Cohan et al. [5] for details) and used for serving recommendations as part of a large public academic search system. For each of author $\mathcal{A}$'s papers P, we use this neural model to obtain an embedding $\tilde{P}$. We then derive an aggregate author-level representation $\tilde{\mathbf{p}}$ (e.g., by weighted averaging that takes author-term affinity into account, see §A.5). Similar authors are computed using a simple distance measure over the dense embedding space. As discussed in the introduction and §**??**, this approach focuses on retrieving authors with the most overall similar papers to $\mathcal{A}$. Intuitively, the baseline can be thought of as "summing over" both the rows and columns of the author matrix in Figure 2. By aggregating across all of $\mathcal{A}$'s papers, information on finer-grained sub-interests may be lost. In addition, by being trained on citation signals, it may be further biased and prone to favor highly-cited papers or authors.

To address these issues, we explore a formulation of the author discovery problem in terms of matching authors along specific dimensions that allow more fine-grained control – such as by using only a subset of views in $\mathcal{V}_\mathcal{A}$, or only a subset of $\mathcal{A}$'s papers, or both — as in the row and column *slices* seen in Figure 2. This decomposition of authors also enables us to explore *contrasts* along specific author dimensions, e.g., finding authors who use similar tasks to $\mathcal{A}$ but use very different methods or resources.

- **Single-facet matches** For each author $\mathcal{A}_i$ in the pool of authors $\{\mathcal{A}_1, \mathcal{A}_2, \ldots\}$, we obtain their respective aggregate representations $\mathcal{V}_{\mathcal{A}_i} = \{\mathbf{m}, \mathbf{t}, \mathbf{r}\}$. We then retrieve authors with similar embeddings to $\mathcal{A}$ along one dimension (or matrix columns in Figure 2; e.g., $\mathbf{r}$ for resources), ignoring the others. Unlike the baseline model, which aggregates *all* information appearing in $\mathcal{A}$'s papers – tasks, methods, resources, general topics, and any other textual information – this approach is able to disentangle *specific* aspects of an author, potentially enabling discovery of more novel, remote connections that can expose users to more diverse ideas and cross-fertilization opportunities.
- **Contrasts** Finding matches along *one* dimension does not guarantee retrieving authors who are *distant* along the others. As an example, finding authors working on *tasks* related to *scientific knowledge discovery* and *information extraction from texts*, could be authors who use a diverse range of *resources*, such as *scientific papers*, *clinical notes*, etc. While the immense diversity in scientific literature makes it likely that focusing on similarity along one dimension only will still surface diverse results in terms of the other (see results in §C), we seek to further ensure this.
  To do so, we apply a simple approach inspired by recent work on retrieving inspirations [11]: We first retrieve the top K authors $\{\mathcal{A}_1, \mathcal{A}_2, \ldots \mathcal{A}_K\}$ that are most similar to $\mathcal{A}$ along one dimension (e.g., $\mathbf{t}$), for some relatively *large* K (e.g., $K = 1000$). We then rank this narrower list inversely by another dimension (e.g., $\mathbf{r}$), and show user $\mathcal{A}$ authors from the top of this list. Intuitively, this approach helps balance relevance and novelty by finding authors who are *similar* enough along one dimension, and within that subset find authors who are relatively *distant* along another.
- **Persona-based matching** Finally, to account for the different focus areas authors may have, instead of aggregating over *all* of an author's papers, we perform the same single-view and contrast-based retrieval using the author's personas $\mathrm{P}_\mathcal{A}$ — or, in other words, row-and-column slices of the matrix in Figure 2.

A.4 Depicting Recommended Authors

Our representation allows us to explore multiple design choices not only for *which* authors we show users, but also *how* we show them. In our experiments (§B, §C), we evaluate authors' facets and personas in terms of their utility for helping researchers learn about new authors, and for controlling how authors are filtered.

Term ranking algorithms to explain what authors work on Researchers, flooded with constant streams of papers, typically have a very limited attention span to consider whether some new author or piece of information is relevant to them. It is thus important that the information we display for each author (such as their main methods, tasks, resources, and also papers) is *ranked*, such that the most important or relevant terms appear first. We explore different approaches to rank the displayed terms, balancing between *relevance* (or centrality) of each term for a given author, and *coverage* over the various topics the author works on. We compare between several approaches, including a customized relevance metric we design, in a user study with researchers (§B). We discuss in more detail the ranking approaches we try in §A.5.

Retrieval explanations In addition to term ranking approaches aimed at explaining to users of Bridger what a new suggested author works on, we also provide users with two rankings that are geared for explaining how the retrieved authors relate to them. First, we allow users to rank author terms $\mathcal{T}$ by how similar they are to their own list of terms (for each facet, separately). Second, users can also rank each author's *papers* by how similar they are to their own — showing the most similar papers first. These explanations can be regarded as a kind of "anchor" for increasing trust, which could be especially important when suggesting novel, unfamiliar content.

A.5 Implementation details

A.5.1 Data

We use data from the Microsoft Academic Graph (MAG) [25]. We use a snapshot of this dataset from March 1, 2021. We also link the papers in the dataset to those in an a large public academic search engine.[4] We limit the papers and associated entities to those designated as Computer Science papers. We focus on authors' recent work, limiting the papers to those published between 2015 and 2021, resulting in 4,650,474 papers from 6,433,064 authors. Despite using disambiguated MAG author data, we observe the challenge of author ambiguity still persists [26]. In our experiments, we thus exclude participants with very few papers (see §C), since disambiguation errors in their papers stand out prominently.

A.5.2 Term Extraction

We extract terms (spans of text) referring to tasks, methods, and resources mentioned in paper abstracts and titles, using the state-of-art DyGIE++ IE model [29] trained on SciERC [15]. We extracted 10,445,233 tasks, 20,705,854 methods, and 4,978,748 resources from 3,594,975 papers. We also use MAG topics, higher-level coarse-grained topics available for each paper in MAG. We expand abbreviations in the extracted terms using the algorithm in [24] implemented in ScispaCy [19].

A.5.3 Scoring papers by relevance to an author

The papers published by an author have varying levels of importance with regard to that author's overall body of publications. To capture this, we use a simple heuristic that takes into account two factors: the author's position in a paper as a measure of affinity (see §A.2), and the paper's overall impact in terms of citations. More formally, for each author $\mathcal{A}$, we assign a weight $w_{\mathcal{A},P}$ to each paper P in $P_{\mathcal{A}}$, $w_{\mathcal{A},P} = \texttt{pos}_{\mathcal{A},P} \times \texttt{Rank}_P$, where $\texttt{pos}_{\mathcal{A},P}$ is 1.0 if $\mathcal{A}$ is first or last author on P and 0.75 otherwise, and $\texttt{Rank}_P$ is MAG's assigned paper Rank (a citation-based measure of importance, see [31] for details), normalized by min-max scaling to a value between .5 and 1.

[4] Redacted for anonymity.

A.5.4 Author similarity

We explore several approaches for author similarity and retrieval, all based on paper-level aggregation as discussed in §A.3. For the document-level SPECTER baseline model discussed in §A.3, we obtain 768-dimensional embeddings for all of the papers. To determine similarity between authors, we take the average embedding of each author's papers, weighted by the paper relevance score described above. We then compute the cosine similarity between this author and the average embedding of every other author. For our faceted approach, we compute similarities along each authors' facets, using embeddings we create for each term in each facet. The model used to create embeddings was CS-RoBERTa [9], which we fine-tuned for the task of semantic similarity using the Sentence-BERT framework [23]. For each author or persona, we calculate an aggregate representation along each facet by taking the average embedding of the terms in all of the papers, weighted by the relevance score of each associated paper.

A.5.5 Identification of personas

We infer author personas using two different approaches. For the first approach we cluster the co-authorship network using the ego-splitting framework in [6]. In a second approach, we cluster each authors' papers by their SPECTER embeddings using agglomerative clustering with Ward linkage [18] on the Euclidean distances between embedding vectors.[5] In our user studies, we show participants their personas and the details of each one (papers, facets, etc.).[6] To make this manageable, we sort the clusters (personas) based on each cluster's most highly ranked paper according to MAG's assigned rank, and show participants only their top two personas.

A.5.6 Term ranking for Author Depiction

We evaluate several different strategies to rank terms (methods, tasks, resources) shown to users in Experiment I (§B):

- **TextRank:** For each term t in an author's set of papers, we create a graph $G_F = (V, E)$ with vertices V the terms and weighted edges E, where weight w_{ij} is the euclidean distance between the embedding vectors $\tilde{t_i}$ and $\tilde{t_j}$. We score each term t_i according to its PageRank value in G_F [17].
- **TF-IDF** For each t, we compute TF-IDF across all authors, considering each author as a "document" (bag of terms) in the IDF (inverse document frequency) term, counting each term once per paper. We calculate the TF-IDF score for each term for each author, and use this as the term's score.
- **Author relevance score** For each t, we calculate the sum of the term's relevance scores (§ A.5.3) derived from their associated papers. If a term is used in multiple papers, the associated paper's score is used for each summand.
- **Random** Each term t is assigned a random rank.

B Experiment I: Author Depiction

In systems that help people find authors, such as Microsoft Academic Graph, Google Scholar, and AMiner [30], authors are often described in terms of a few high-level topics. In advance of exploring how we might leverage facets to engage researchers with a diverse set of authors, we performed a user study to gain a better understanding of what information might prove useful when depicting authors. We started from a base of Microsoft Academic Graph (MAG) topics, and then added their extracted facets (tasks, methods, resources). We investigated the following research questions:

- **RQ1**: Do tasks, methods, and/or resources complement MAG topics in depicting an author's research?
- **RQ2**: Which term ranking best reflects an author's interests?
- **RQ3**: Do tasks, methods, and/or resources complement MAG topics in helping users gain a better picture of the research interests of *unknown* authors?
- **RQ4**: Do personas well-reflect authors' different focus areas?

[5] Implemented in the scikit-learn Python library [22]. Distance threshold of 85.

[6] Some authors do not have detected personas; we observe this to often be the case with early-career researchers.

B.1 Experiment Design

Thirteen computer-science researchers were recruited for the experiment through Slack channels and mailing lists. Participants were compensated $20 over PayPal for their time. Study sessions were one-hour, semi-structured interviews recorded over Zoom. The participants engaged in think-aloud throughout the study. They evaluated a depiction of a known author (e.g., research mentor) for accuracy in depicting their research, as well as depictions of five *unknown* authors for usefulness in learning about new authors.

Throughout all parts of the experiment, the interviewer asked follow-up questions regarding the participant's think-aloud and reactions.[7] To address **RQ1** and **RQ2**, the participants first evaluated the accuracy of a known author's depiction.

Step I. To begin, we presented the participant with only the top 10 MAG topics for the known author. We asked them to mark any topic that was unclear, too generic, or did not reflect the author's research well. Next, we provided five more potential lists of terms. One of these lists consisted of the next 10 top topics. The other four presented 10 tasks, each selected as the top-10 ranked terms using the strategies described in §A.5. We asked participants to rank the five lists (as a whole) in terms of how well they complemented the first list (with an option to select none).

Step II. The process then repeated for five more potential lists to complement the original topics and the highest-ranked second list selected in Step I — this time, with methods instead of tasks. If the participant ranked a methods list highest, we then presented the participant with a resources list that used the same ranking strategy preferred by the participant for methods, and asked whether or not this list complemented those shown so far.

Step III. To address **RQ3**, participants next evaluated the utility of author depictions for five unknown authors. To describe each unknown author, we provided topics, tasks, methods, and resources lists with 10 terms each. The non-topics lists were ranked using TF-IDF as a default. The participant noted whether or not each additional non-topics list complemented the preceding lists in helping them understand what kind of research the unknown author does.

Step IV. Finally, for **RQ4**, we asked participants to evaluate the known author's distinct personas presented in terms of tasks, which were ranked using TF-IDF. On a Likert-type scale of 1-5, participants rated their agreement with the statement, "The personas reflect the author's different research interests (since the year 2015) well."

B.2 Results

B.2.1 Results for RQ1

The majority of participants found that tasks, methods, and resources complemented topics to describe a known author's research. For both tasks and methods, 11 of 13 participants felt that seeing information about that facet, more so than additional top MAG topics or no additional information, complemented the original top ten MAG topics. The prevailing grievance with the additional MAG topics was that they were too general. Furthermore, 7 of 9 participants who evaluated a resources list thought that it complemented the preceding lists.

B.2.2 Results for RQ2

Participants overall preferred the relevance score ranking strategy for tasks and methods. We compared the four ranking strategies and MAG topics baseline strategy for both tasks and methods. For each participant, we awarded points to each strategy based on its position in the participant's ranking of the five strategies. We awarded the least favorite strategy one point and the most favorite strategy five points. Since there were 13 participants, a strategy could accumulate up to 65 points. Separately, we counted how many times each strategy was a participant's favorite strategy (Figure 4c, d). With regards to tasks, TextRank and TF-IDF accrued the most points from participants, with the relevance score trailing close behind (Figure 4a). Meanwhile, the MAG topics baseline accrued the least points, even fewer than the random task ranking strategy. In addition, relevance score and TextRank were chosen most often as the favorite task ranking strategy (Figure 4c). With regards to

[7] The script for Experiment I can be found in our supplementary materials.

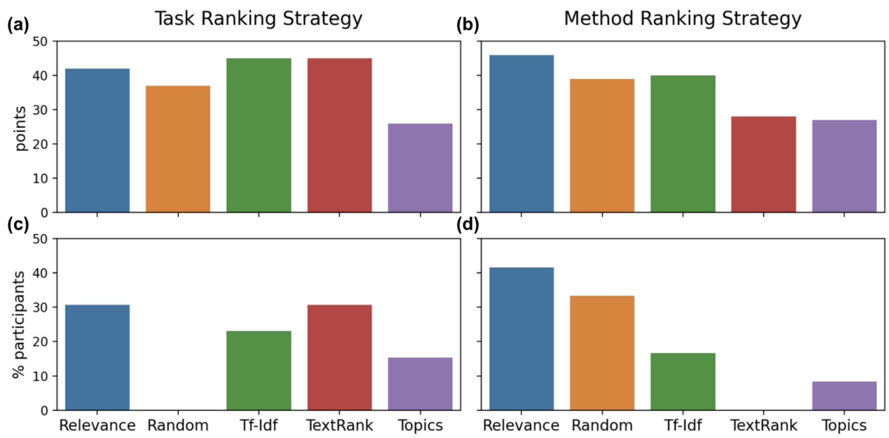

Figure 4: Points awarded to each ranking strategy for tasks (a) and methods (b), and percentage of participants who favored each strategy most for tasks (c) and methods (d).

methods, the relevance score ranking strategy performed best in terms of both total points (Figure 4b) and favorite strategy (Figure 4d).

B.2.3 Results for RQ3

Participants generally found tasks, methods, and resources helpful to better understand what kind of research an unknown author does. To calculate how many participants were in favor of including tasks, methods, and resources to help them better understand an author, we determined the average of each participant's binary response per facet. Adding up the 13 responses for each facet, we saw that the majority of participants thought each additional facet helped them understand the unknown author better. All 13 participants found the tasks helpful, eight found the methods helpful, and 12 found the resources helpful. As an example, P12 connected an unknown author's topics, tasks, and methods to better understand them: "*I wouldn't have known they were an information retrieval person from the [topics] at all.... The previous things [in topics and tasks] that mentioned translation and information retrieval and kind of separately... This [methods section] connects the dots for me, which is nice.*" Interestingly, methods were not viewed to be as useful as tasks or resources. The majority of participants cited unfamiliar terms as a key issue.

B.2.4 Results for RQ4

Participants indicate preference for personas selected based on papers rather than co-authorship. After the experiment, six participants were informally asked to compare the experiment's personas selected based on co-authorship with the personas based on paper-based clustering (see §A.5). Four of them preferred the updated version. Furthermore, one of the users who preferred the old version still thought the updated version had better personas themselves and merely did not like the updated personas' ordering. In addition, all six participants liked seeing the personas in terms of papers. In our experiment in §C, we observed much higher satisfaction with the updated personas in comparison to the original personas of this experiment.

C Experiment II: Author Discovery

We now turn to our main experiment, exploring whether facets can be employed in Bridger to spur users to discover valuable and novel authors and their work. We use our two author-ranking strategies (§A.3), one based on similar tasks alone (`sT`) and the other on similar tasks with contrasting (distant) methods (`sTdM`). We compare these strategies to the SPECTER (`ss`) baseline. More specifically, we investigated the following research questions:

- **RQ5**: Do `sT` and `sTdM`, in comparison to SPECTER, surface suggestions of authors that are considered novel and valuable, coming from research communities more distant to the user?

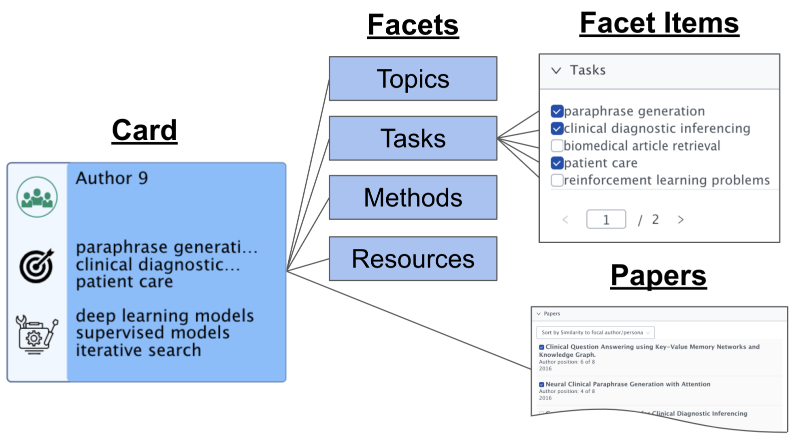

Figure 5: Illustration of information shown to users in Experiment II, §C. When the user clicks on an author card, an expanded view is displayed with 5 sections: papers, topics, and our extracted facets — tasks, methods, and resources.

- **RQ6**: Does sorting based on personas help users find more novel and valuable author suggestions?

C.1 Experiment Design

Twenty computer-science researchers participated in the experiment after recruitment through Slack channels and mailing lists. Participants were compensated $50 over PayPal for their time.

All participants were shown results based on their overall papers (without personas) consisting of 12 author cards they evaluated one by one. Four cards were included for each of `sT`, `sTdM`, and `ss`. We only show cards for authors who are at least 2 hops away in the co-authorship graph from the user, filtering authors with whom they had previously worked.

For participants who had at least two associated personas, we also presented them with authors suggested based on each separate persona: four author cards for each of their top two personas (two under `sT` and two under `sTdM`). Whether the participants saw the personas before or after the non-persona part was randomized.

Each author card provides a detailed depiction of that author (see Figure 2). The author's name and affiliation is hidden in this experiment to mitigate bias. As shown in Figure 5, cards showcase five sections of the author's research: their papers, MAG topics, and our extracted facet terms. We also let users view the tasks and methods ranked by *similarity* to them, which could be helpful to explain why an author was selected and better understand commonalities.

The cards showed up to five items for each section, with some sections having a second page, depending upon data availability. Papers could be sorted based on recency or similarity to a participant / persona. To avoid biasing participants, the only information provided for each paper was its title, date, and the suggested author's position on each paper (e.g., first, last).

Each of these items (papers and terms) had a checkbox, which the user was instructed to check if it fulfilled two criteria: 1) potentially interesting and valuable for them to learn about or consider in terms of utility, and 2) not too similar to things they had worked on or used previously. Following a short tutorial,[8] participants evaluated each author shown by checking the aforementioned checkboxes (see Figure 5, right). While evaluating the first and last author (randomized), the participant engaged in a protocol analysis methodology (sharing their thinking as they worked). Participants with personas were also asked, based on each persona's top five associated papers, whether they each reflected a coherent focus area, and whether they seemed useful for filtering author suggestions.[9]

[8] The tutorial slides are available in our supplementary materials.

[9] See supplementary materials for the source code used for generating the data for Experiment II, as well as the code for the interactive application used in the evaluation, and the script used to direct the participants.

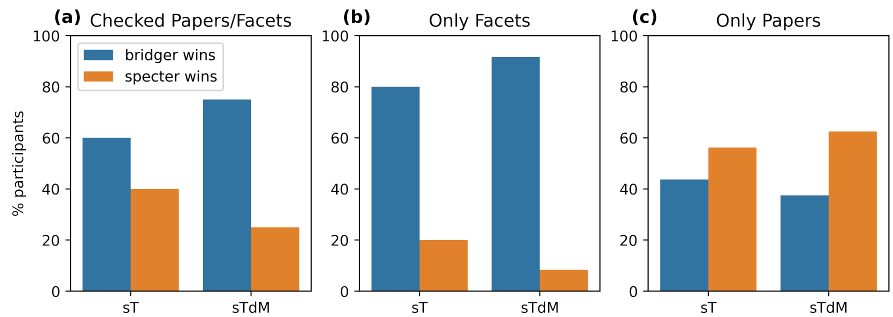

Figure 6: *More users prefer Bridger for suggesting novel, interesting authors.* Percent of the participants who preferred author suggestions surfaced by faceted conditions (`sT` and `sTdM`, blue bars) compared to a baseline non-faceted paper embedding (`ss`, orange bars). On average, users prefer the former suggestions, leading to more discovery of novel and valuable authors and their work (a). When broken down further, we find users substantially preferred the facet items shown for authors in our condition (b), and preferred the paper embedding baseline when evaluating papers (c). See §C for discussion.

C.2 Quantitative Results

For each author card evaluated by a user, we calculate the ratio of checked boxes to total boxes in that card. Then, for each user, we calculate the average of these ratios per condition (`sT`, `sTdM`, `ss`), and calculate a user-level preference S specifying which of the three conditions received the highest average ratio. Using this score, we find the proportion of users who preferred each of the `sT` and `sTdM` conditions in comparison to `ss`. This metric indicates the user's preference between Bridger- and SPECTER-recommended authors in terms of novelty and value (RQ5).

Figure 6(a), shows results by this metric. The facet-based approaches lead to a boost over the non-faceted `ss` approach, with users overall preferring suggestions coming from the facet-based conditions. This is despite comparing against an advanced baseline geared at relevance, to which users are naturally primed.

We break down the results further by slightly modifying the metric to account for the different types of item types users could check off. In particular, we distinguish between the task/method/resource/topic checkboxes, and the paper checkboxes. For each of these two groups, we compute S in the same way, ignoring all checkboxes that are not of that type (e.g., counting only papers). This breakdown reveals a more nuanced picture. For the task, method, resource and topic facets, the gap in favor of `sT` grows considerably (Figure 6b). In terms of papers only, `ss`, which was trained on aggregate paper-level information, achieves a marginally better outcome compared to `sT`, with a slightly larger gap in comparison to `sTdM` (Figure 6c). Aside from being trained on paper-level information, SPECTER also benefits from the fact that biases towards filter bubbles can be particularly strong with regard to papers. Unlike with facets, users must tease apart aspects of papers that are new and interesting to them versus aspects that are relevant but familiar. See §D.1.3 for more discussion and concrete examples.

Importantly, despite obtaining better results overall with the faceted approach, we stress that our goal in this paper is not to "outperform" SPECTER, but mostly to use it as a reference point — a non-faceted approach used in a real-world academic search and recommendation setting.

Personas We also compare the results from `sT` and `sTdM` conditions based on personas P for user $\mathcal{A}$, versus the user's non-persona-based results presented above (RQ6). We compare the set of authors found using personas with authors retrieved without splitting into personas (equivalent to the union of all personas). Table 1 shows the number of users for which the average proportion of checked items was higher for the persona-matched authors than for the overall-matched authors (for at least one of the personas). For most participants, users signalled preference for persona-matched authors when looking at one or both of their personas. Interestingly, for papers we see a substantial boost in preference for both conditions, indicating that by focusing on more refined *slices* of the user's papers, we are able to gain better quality along this dimension too.

Item type	sT	sTdM
All	58%	75%
Paper	83%	67%
Topic	58%	75%
Task	42%	50%
Method	67%	58%
Resource	50%	67%

Table 1: Percentage of users with personas (N=12), for which the average proportion of checked items was higher for the persona-matched authors than for the overall-matched authors. Users saw suggested authors based on two of their personas. The suggestions came from either the `sT` or `sTdM` conditions. Reported here are counts of users who showed preference for one or both personas.

C.3 Evidence of Bursting Bubbles

The matched authors displayed to users were identified based either on `sT` and `sTdM` or the baseline SPECTER-based approach (`ss`). These two groups differed from each other substantially according to several empirical measures of similarity. We explore the following measures, based on author dimensions in our data that we do not use as part of the experiment: (1) Citation distance: A measure of distance in terms of citations that the user has in common with the matched author (Jaccard distance: 1 minus intersection-over-union). This is calculated both for incoming and outgoing citations. (2) Venue distance: The Jaccard distance between user and matched author for publication venues. (3) Coauthor shortest path: The shortest path length between the user and the matched author in the coauthorship graph. Findings of this analysis, shown in Figure 3, suggest that Bridger surfaces novel authors from more diverse, distant fields and research communities than SPECTER (RQ5).

In the following section, we conclude by diving deeper into user interviews we conducted, revealing more evidence and insights into user preferences and surfacing potential issues and challenges for building author discovery systems.

D User Interviews: Analysis & Discussion of Author Discovery

D.1.1 Bridges Across Scientific Filter Bubbles

Bridger authors encourage more diverse connections. Under the Bridger conditions, participants noted diverse potentially useful research directions that connected their work to other authors not only within their own subareas, but also other areas. This was especially true under the `sTdM` condition. For instance, P9, who works on gradient descent for convex problems, saw a `sTdM` author's paper discussing gradient descent but for deep linear neural networks, which imply non-convex problems. They remarked, "*This is a new setup. It's very different, and it's super important . . . definitely something I would like to read . . .* " Considering a paper under a `sTdM` author, P6 observed an interesting contrast with their work: "*I think my work has been bottom-up, so top-down would be an interesting approach to look at.*" As another example, P2 drew a connection between the mathematical area of graph theory and their area of human-AI decision-making under the `sTdM` condition: "*This could be interesting mostly because . . . they're using graph theory for decision making . . . something I have not considered in the past.*" P19 remarked of an `sTdM` author's paper, "This one actually seems quite interesting because it seems like explicitly about trying to bridge the gap between computational neuroscience models, understanding the neocortex, and computing. So that seems like it's... going to actually chart the path for me between my work and the stuff I think about like artificial neural networks and machines."

In reacting to `sTdM` authors, many participants were able to go further than simply state their interest in a connection and also describe *how* they would utilize the connection. Looking at a `sTdM` author, P6 explained how the author's neuroscience work could motivate work in their area of natural language processing: "*I might learn from that [paper] how people compose words, and that might be inspiring for work on learning compositional representation . . .* " P18 checked off a paper titled "Multidisciplinary Collaboration to Facilitate Hypothesis Generation in Huntington's Disease" under a `sTdM` author "*because new ways to think about generating hypotheses could be interesting.*" Seeing the topic 'spike-timing-dependent plasticity' under a `sTdM` author, P19 mused, "*I would like to* *understand how spike-timing-dependent plasticity works and whether that could lead to a better learning rule for other types of neural nets, like the ones I work with on language, so that seems fun.*" P12 described a `sTdM` author's paper about knowledge-driven search applications as useful to them because "*One of my primary research areas is knowledge base completion. However, that's not an end application. An end application would be a search application which kind of uses my method to complete the knowledge base, and gives the user the end result. . . .*" Though the `sTdM` condition presented more of a risk in terms of surfacing authors with which the user could draw connections, it also surfaced the more far-reaching connections.

The `sT` condition also helped participants ponder new connections, though perhaps not as distant. P8 said of a `sT` author's work, "*I've worked a bit on summarization, so I want to know whether the approaches that I've worked on are applicable to real-time event summarization, which is a task I don't know about.*" Also reflecting on a `sT` author, P1 explained, "*I've done a lot of work with micro tasks and these seem more maybe larger scale, like physical tasks or like planning travel. . . . There are so many problems . . . that I could apply my techniques to.*" Other times, participants would connect one facet of their work to a different facet of the suggested author's work. In discussing a question-answering paper from a `sT` author, P8 explained, "*I don't have experience with [the method] adversarial neural networks [used in this paper], but question answering is a task that I've worked on, so I would want to check this.*" Conversely, if participants found new connections with SPECTER, they tended to be more immediate connections to authors in their area. As an example, when checking off the paper "Efficient Symmetric Norm Regression via Linear Sketching" from a SPECTER-suggested author, P9 observed, "*I have used sketching techniques and I have [also] used norm regression, but [on] this specific problem I have not.*" P9 also identified some of the papers from the suggested author as co-authored by their advisor.

D.1.2 Facets Help Elicit New Research Directions But Require More Context

Describing an author's work with short, digestible items in the form of tasks, methods, and resources helped participants find interesting new research directions. For instance, P14 expressed that a `sTdM` author's paper associated with medical image diagnosis would not be useful for them to consider because "*breaking into that space for me would require a lot of work.*" However, when they later saw 'medical image diagnosis' as a task, they commented, "*As a task, I could see some usefulness there. There could be other approaches that might more quickly catch my interest.*" Committing to interest in the task required much less effort. Moreover, participants were able to peruse more of an author's interesting tasks and methods that they did not necessarily find in their top papers. Reacting to one `sT` author, P3 did not see any papers related to 'biomedical question answering,' but they did see 'biomedical question answering system' as a method. They then noted, "*I'm going to click 'biomedical question answering' because that's not what I have worked on before, but I'm interested in learning about it.*"

Tasks, methods and resource facets support discovery better than topics. While participants occasionally thought certain tasks, methods, or resources were too generic, participants were much more likely to complain that topics were too high-level to spark ideas for new, profitable research directions. P3 summarized, "*I think many of them are quite generic, so I can say I already worked on it,*" and P7 noted, "*'Artificial intelligence' is too broad. I think everything comes under that.*"

Terms with unknown meaning often garner interest, but all facets and papers require more context. Participants commonly identified tasks, methods, and resources as interesting, even when they did not fully understand their meaning. When P4 saw the method 'least-general generalization of editing examples' from a `sT` author, they stated, "*Don't know what this means exactly, but it sounds interesting.*" P13 marked their interest in the task "folksonomy-based recommender systems" under a `sTdM` author after having commented, "*I'm curious [about folksonomy] simply because I'm ignorant.*"

`sTdM` also surfaced distant resources that sparked interest. In seeing the resource 'synaptic resources' under a `sTdM` author, for example, P19 simply said, "*I'd like to know what that is.*" Nonetheless, many participants also struggled with indiscernible terms. For example, P20 said of the resource 'NAIST text corpus' under a `sT` author, "*I'm not sure what this is, and I can't guess from the name. And it wasn't mentioned in the title of the papers.*" P2 explained that a paper did not "*seem that interesting, but mostly because I don't understand all of these words.*" Thus, providing term definitions may be helpful. For additional context, multiple participants expressed interest in having abstracts available, and P15 suggested including automated summaries [2].

D.1.3 Biases Toward Scientific Filter Bubbles

Time constraints in the fast-moving world of research inhibit exploration beyond the filter bubble. Despite clear interest in an author's distant research, a couple of participants were hesitant to make connections. In reacting to a `sT` author, P11 recognized, "*There's just a bunch of really interesting kind of theory application papers in this list that I'm not familiar with. . . . I would maybe scan a little bit of these, but it's so far off that it's harder to make room to read someone that far away, but still cool.*"

Unknown background knowledge can make it intimidating to consider new areas. Engaging with distant authors' work requires a large cognitive load that can make uncovering connections difficult. P18 provided the following example: "*Maybe there's some theoretical computer science algorithm that if I knew to apply it to my problem would speed things up or something like that, but I wouldn't know enough to recognize it as interesting.*" Echoing findings in §D.1.2, this comment suggests that unfamiliar terms can especially hinder making interesting connections, and that highlighting the most useful aspects of a distant author's research may facilitate building far-reaching connections.

Preconceived notions of an area hinder consideration of connections to that area. Because Bridger's authors are selected to be more different from the user than SPECTER's authors, they often met with hard-line resistance, without full consideration of potential links. Looking at a `sTdM`-suggested author, the natural language processing (NLP) researcher P20 said, "*This is not really an NLP paper, so I would pass.*" Similarly, P17 rejected `sTdM` suggestions, saying "*I don't know anything about neuroscience, and I'm not going to start now probably.*"

Difficulty teasing apart novel aspects from paper titles helps SPECTER. Although participants were asked to only check off interesting papers that suggested something new for them to explore, biases towards filter bubbles can be particularly strong with regard to papers because users must tease apart papers' new and interesting aspects from their relevant but familiar aspects. Even if a paper is directly connected to a user's research, they may be tempted to check off a paper because they have not seen that *exact* paper or because it has minute differences from their work. In contrast, when judging a particular facet item, participants need only contemplate the novelty of the term itself, without distraction or fixation on other terms [10, 13, 11]. As an example, P17 swiftly separated a task's general relevance from its lack of novelty to know not to check it. They explained, "*'Scientific article summarization'- It is relevant, [but] I'm already familiar with it.*" This bias helps explain the overall preference for SPECTER when considering only papers (Figure 6(c)).

D.1.4 Personas

All participants with personas stated at least one would be helpful. Upon first view of their personas, of the 12 participants who had them, seven described their two personas as distinct, coherent identities that would be useful for filtering author suggestions. As an example, P2 characterized their personas as related to "*human-AI collaboration or decision-making*" and "*error analysis and machine learning debugging*" respectively. The other 5 participants described one persona as coherent and seemingly useful for filtering authors. Concerns about their other personas were related to coherence, granularity, overlap with the other persona, and preference for the non-persona results after already looking through them and their first persona. Though the persona author suggestions performed relatively well in generating novel connections (Table 1), a few participants commented that they did not see the connection between suggested authors and their persona. For example, under a persona associated with lexical semantics, P6 commented on a `sTdM` paper, "*'Causality' is not a topic I would work on in lexical semantics.*" Diverse author suggestions may be more confusing under personas because users look for connections specific to that persona; indicating to users when these author suggestions are for exploratory purposes may be helpful.