'Too Late to Collaborate': Challenges to the Discovery of in-progress Research

Corinna Breitinger[1], Bela Gipp[1,2], Patrick Wortner[2], Harald Reiterer[1]

[1]University of Konstanz, Germany (first.last@uni-konstanz.de)

[2]University of Wuppertal, Germany (last@uni-wuppertal.de)

ABSTRACT

Identifying the ongoing research in one's field is an essential yet time-intensive information seeking task. Today's digital libraries support researchers in the search and discovery of published academic research. However, they are unable to support researchers in the discovery of ongoing research, i.e., research that has not yet been published. The discovery of ongoing research thus remains a manual information seeking task lacking standardized processes or automated support systems. We present findings from an initial qualitative study on how computer science researchers from four disciplines currently go about identifying ongoing research projects within their fields and the challenges they face. A major challenge we identify is what we term the discovery-confidentiality trade-off. On the one hand, researchers express a need to discover ongoing research projects in their domain to identify collaboration partners and to avoid performing duplicate research. However, at the same time, researchers are hesitant to reveal details about their own in-progress research for fear of idea plagiarism. We discuss several key factors influencing this trade-off, such as trust and timeliness. We argue that these factors must be accounted for in the design of future academic search and recommendation solutions to support researchers in the timely identification of ongoing research.

1 INTRODUCTION

Being informed of ongoing research in one's field is an essential information seeking (IS) task for researchers. Researchers informed about current research projects in their fields can avoid duplicate research, sooner receive inspiration for future research directions, and identify potential *collaborators* with whom to exchange expertise or resources for mutual benefit. In many disciplines, however, it is not uncommon for several years to pass between framing a research idea and concluding that research in the form of a publication. Because of this delay, it remains a challenging IS-task to stay informed of current research projects in one's domain.

Identifying suitable collaboration partners has become increasingly important as research projects become more complex and often require expensive scientific instruments or expertise from multiple disciplines [5]. The benefits of seeking out collaboration in research have been widely examined [1,3,20]. Research collaboration has been shown to increase productivity [13], as commonly measured by the number of published peer-reviewed articles [5]. Collaborative publications have also been shown to receive more citations than single-author publications [5], and researchers who frequently collaborate tend to attract more research funding [15].

In addition to the benefits of collaboration, prior research has extensively examined the information-seeking behavior of researchers and scholars from a range of scientific disciplines [7,11,18], typically with the underlying aim of improving library services for researchers [10]. However, prior research on the information-seeking behavior of researchers has exclusively focused on the discovery of *published* research, i.e., research projects that have been completed and thus are almost always no longer current. We view this lack of research on *ongoing* research discovery as a shortcoming that we address in this paper.

Automated systems have been proposed to support the discovery of published literature in one's field [4], as well as the discovery of collaborators in the form of scientific collaboration recommendation [2,17,21], or expert finding systems (EFS), such as researcher expertise search systems [16,19,22]. However, existing systems make use of one, or a combination, of the following: researcher's publications [12,16,22], which typically represent completed projects, past collaboration behavior, i.e., co-authorship networks [12,14,17], or information publicly available online [16,22], e.g. on university's department websites. While recently published research might be similar to ongoing projects, there is no way for another researcher to be certain. Publications may also be representative of only a fraction of a researcher's current work and will provide no information on research that has recently changed direction.

More recently, academic social networking sites (ASNS), such as Academia.edu and ResearchGate are gaining popularity. ResearchGate allows users to define projects and can thus act as a tool to discover ongoing research. While research projects described on ASNS or university websites have the potential of being up-to-date, finding this information remains tedious. Additionally, verifying the accuracy of manually created research project descriptions and correcting false information on ResearchGate are a cumbersome manual task [23]. Furthermore, not all researchers are able or willing to provide details about their unpublished work openly on the Web.

Because existing automated systems primarily examine research data of the past, they are of little help in (1) identifying researchers who have changed their research direction or (2) identifying in-progress research projects for the purpose of collaboration recommendation. The reliable discovery of ongoing research thus remains a mostly manual information seeking task.

In summary, prior research has extensively examined researchers' *collaboration* and *information-seeking* behaviors. However, no research has been performed on how researchers, specifically computer scientists, go about the *information seeking task of in-progress research discovery to identify potential collaborators*.

In this paper, we examine how computer science (CS) researchers currently address this task. Specifically, our study answers the following research questions:

1) What strategies do CS researchers currently employ to discover *ongoing*, i.e., yet unpublished, research projects in their respective field?

2) What challenges do CS researchers face when attempting to identify ongoing research projects in their field?

In this under-researched domain, our findings shed light on the question of how CS researchers may be better supported in the task of identifying ongoing academic research projects.

The remainder of this paper is structured as follows. Section 2 describes our study design. Section 3 presents the findings and describes the factors underlying the observed *discovery-confidentiality* trade-off. Section 4 provides a discussion, including research and design implications.

2   METHOD

We conducted 40-minute semi-structured interviews with 16 computer science researchers from six European Universities. Interview-based approaches have been commonly used to investigate information seeking behavior in academia [7,8]. Since the behaviors and challenges faced when identifying ongoing research projects differ depending on the academic discipline, this initial study focused only on computer scientists. The participants were doctoral or postdoctoral researchers, as well as three professors, each representative of one of four computer science domains: Human-Computer Interaction (HCI), Information Science (Info. Sci.), Data Analysis and Visualization (Dat. Ana.) and broadly defined Software Engineering (Softw. Eng.). By choosing four domains, we allowed for a comparison of potential differences in information seeking behavior across CS domains for the task of identifying ongoing research projects.

Table 1: Overview of Study Participant's Backgrounds

ID	Gender	Level	Research domain affiliation[1]
1	m	Prof.	Inf. Sci.
2	m	Ph.D.	Inf. Sci.
3	m	Ph.D.	Inf. Sci.
4	m	postdoc	Inf. Sci.
5	m	Ph.D.	HCI
6	m	postdoc	HCI
7	f	Ph.D.	HCI
8	f	Ph.D.	HCI
9	f	postdoc	Dat. Ana.
10	m	Ph.D.	Dat. Ana.
11	m	Ph.D.	Dat. Ana.
12	f	postdoc	Dat. Ana.
13	f	Prof.	Softw. Eng.
14	m	Prof.	Softw. Eng.
15	m	Ph.D.	Softw. Eng.
16	m	Ph.D.	Softw. Eng.

[1] We chose these computer science domains as the focus for our initial study, as these domains are likely also among the most relevant for our readership.

We asked participants about their strategies for seeking ongoing research projects, the importance of this task, and the challenges they encountered. In return, they were compensated for their time. The interviews were audio recorded and the transcribed data was coded using an inductive grounded theory approach [6]. The process of data gathering and analysis was cyclical with interviews performed over the course of several months to adhere to the theoretical sampling methodology of grounded theory with the aim of maximizing similarities and differences among the groups in the gathered data. The data was open coded, and the axial and selective coding examined by a second researcher, who re-analyzed the data using the theoretical lens that had been identified.

3   FINDINGS

In this section, we first describe the strategies employed by computer science researchers when seeking information on ongoing research projects in their respective field. Subsequently, we describe the identified *transparency-confidentiality* trade-off experienced by researchers. We discuss factors influencing this trade-off and suggest implications for the future research and design of academic search and research recommendation systems.

3.1 Ongoing Research Discovery

We found that all researchers interviewed described an individual set of strategies to help them address the task of identifying in-progress research in their field. The most used strategies were: a) *attending conferences* (mentioned by 15 participants), b) *talking to colleagues* (14), c) *browsing researcher's online profiles*, e.g. on university websites, funding websites, Google Scholar, or ResearchGate (14), d) *relying on senior researchers* to inform them (5), and e) *attending talks* given by visiting researchers (4).

Many strategies thus continue to rely heavily on **face-to-face encounters** (a; b; e), or **personal connections** (b; d). P10 and others noted that ongoing research discovery often happens by chance as you meet other researchers: *"Apart from [attending conferences], I'm not really looking or searching for what's going on. It's more like 'you get to know other people'."*

Regarding the importance of connections, P7 said: *"[the task is] a lot like 'it's not what you know it's who you know'."* While P11 observed: *"I would only say [it's possible to be informed about ongoing work] by word of mouth, or if you know somebody."*

While the combination of used strategies differed more strongly when compared *across* the four examined CS domains than *within* the same domain, each participant described his or her individual combination of strategies to solve the task. This lack of a standardized appropriation of search strategies for in-progress research discovery may pose its own challenge to identifying in-progress research, especially at an inter-disciplinary level.

Lastly, none of the interviewed researchers viewed ongoing research discovery as an easy or straightforward task. P7 noted the inability of existing search systems to support with this task:

"When you're [performing] a literature search online, you're often looking at past research. It might not help you make out a relevant collaborator right now." P7 also observed the impracticality of her preferred strategy, making use of personal connections suggested by an advisor: *"it gets harder the larger the research community."* Since for larger domains, it becomes unlikely even for a senior researcher to have an overview of all projects and researchers involved. P12 described another limitation of leveraging one's research network to identify ongoing research: *"In my opinion, [discovering ongoing research] is only possible, if you know the people."*

Despite the range of strategies described by participants, several researchers felt they did not know exactly *how* they could find ongoing research, or what was the 'right way' to address this task. As a result, they had a feeling that they were always missing some information. Regarding the use of university websites, P7

observed: *"It was impossible to verify that I had even looked at all of the right things or that I had looked in the right places."*

Apart from strategy c) 'browsing researcher's profiles', all named strategies relied on previous work, e.g., already having attended a conference, or already having established connections. Researchers new to a domain, or with no connections to rely on, must browse publicly available information. All strategies also fell short when it came to the researchers' own reflections about their effectiveness, e.g. 'chance' or 'luck' was often mentioned when thinking back upon how they had identified ongoing research in the past.

3.2 The Discovery-Confidentiality Trade-off

Having examined the strategies that CS researchers employ for ongoing research discovery, including their weaknesses, we discuss in more depth an underlying struggle of which nearly all researchers interviewed were acutely aware. We term this dichotomy the *discovery-confidentiality* trade-off. On the one hand, researchers expressed a need to discover ongoing research in their field, yet on the other, they wished to keep details about their own research confidential. We subsequently discuss the implications of these findings on the design of recommendation system.

We found researchers expressed a **need to discover** in-progress research in their field: *"I would say that it's the most important thing you have to do when you're a researcher."* (P9). The ability to discover ongoing research was perceived important for several reasons. P9 followed up her statement with *"You don't want to replicate something that's already existing"*, hinting at the fear of performing duplicate research. This fear was repeatedly stated as a motivator: *"As for every scientist, if you do stuff that has already been done, that's a problem."* (P2). If you are not aware of ongoing research *"it might happen that I work on something and then someone else is working on the same thing, and then they publish it before me. Then my work is worthless."* (P1).

Gaining inspiration for one's own work was another reason researcher wanted to discover ongoing research efforts: P7: *"You want to discover ideas and develop them further. So, it's important to know what's out there."* P6 also noted that it lets him *"see in which direction my research field is moving."* P12 explained: *"I think that's very important, [because] you see what the hot topics are, and [..] once you know what other people are doing you have some ideas of how you can support them."*

The task of finding collaborators was perceived as very important by the majority of researchers. P9 stated, *"The more information you have about other people, what they are actually working on, the easier it is for you to initiate collaboration."*. P4 agreed: *I suppose if I could be matched with researchers doing the same thing as me that would save a lot of time."* Interestingly, mainly the professors (P1, P13) also confirmed the prior-mentioned importance of identifying collaborators *"to secure funding."*

While the interviewed researchers described many benefits of collaboration, they at the same time expressed a wish to keep details about their own research **confidential**. This internal struggle was either directly or indirectly mentioned by 13 of the 16 interviewees:

> P10 stated, *"I think, a lot of researchers do not actually tell other people what they are currently working on, because it's kind of a trade-off: You want to search for collaborators, but on the other hand, you don't want to show your ideas."*

> P11: *"...there's like a bad feeling there, if you have something new, you don't want to share too much information, because somebody might try to copy the idea or steal the idea."*

> P3: *"It's [..] a trade-off: you need to tell them [other researchers] something about what you're currently working on or trying to do. But especially, if it's a person that you don't know, then, of course, you don't want to tell them like the whole thing. So that's definitely something that I'm asking myself always."*

We identified two key factors influencing this trade-off: trust and timeliness. Naturally, low levels of trust influenced the desire for confidentiality, while the need for timely ongoing research discovery influenced the desire to share openly. Unless a level of **trust** has already been established, with a potential collaborator, researchers wished to keep in-progress activities confidential: *"If you don't know the other party, then you don't know whether they are going to steal your idea. I think that is one, or the major concern."* (P10). According to P9: *"trust is the most important issue in starting collaboration. Especially at the beginning."*

Participants worried primarily about their ideas being plagiarized, but also about their ideas being inadvertently appropriated by others – a phenomenon commonly known as 'cryptomnesia': *"I think one other issue is that you come to a similar idea, although you don't realize it. So, it's not like "ah, I know it's his work, and I'll try to sell it now in my paper," but it's more like you come to a similar idea, although you don't realize it."* (P9).

While trust, and the associated fear of one's ideas being plagiarized, was the most prominent factor influencing the desire for 'confidentiality' in this trade-off, **timeliness** was identified as a key factor influencing the desire to discover ongoing research in the first palace. Identifying relevant ongoing projects 'too late' was an experience that researchers found frustrating when attempting to discover potential collaboration partners: *"...not everyone wants to share [ongoing research]. At least, they don't want to share it until it's published. And then it might be too late."* (P6).

P2 also noted he was disappointed when members of a research group with excellent expertise in his area discovered one of his research projects described on a university website too late in order to collaborate on that project: *"...for this publication deadline we were aiming for, it was too late to involve those people. It would have taken too much time to bring them into the project. [..] if we would have met them earlier, it would have been quite helpful."*

In addition to the need for ongoing research to be discovered at the 'right time' to be relevant – preferably as early as possible – researchers also expressed a feeling of 'wasting time' when seeking out collaboration opportunities: *"That it takes time and it takes some effort"* (P9). The time and effort required could lead some researchers to want to give up attempting to discover ongoing work altogether, as noted by P6: *"...it can sometimes waste a lot of time, not just organizing the collaboration, but just trying to find the 'optimal' collaboration partner, and you'd be better off just doing it alone."* Three researchers also stated that they do not use ResearchGate as much as they might like because of lacking time to curate their own profiles and projects.

Experiencing the *discovery-confidentiality* trade-off to some extent appears inevitable when seeking to identify ongoing research in one's field. However, we suggest that automated support systems of the future that are tailored explicitly to this task could significantly alleviate the factors underlying this trade-off.

4 IMPLICATIONS FOR RESEARCH & DESIGN

We showed that the ability to identify ongoing research is deemed an important task by researchers to prevent performing duplicate research, to gain inspiration, and to find collaborators at the 'right time.' Currently, the task of seeking potential collaborators for in-progress research relies heavily on existing connections and face-to-face communication between researchers, as shown in Section 3.1.

We identified trust as a key factor influencing the degree of confidentiality researchers are willing to 'trade' for higher chances of in-progress research discovery, since sharing details of ongoing work entails a risk that ideas are stolen. Furthermore, we identified timeliness of ongoing research discovery as an important factor. No automated academic collaboration recommendation systems currently consider a researcher's full portfolio of in-progress work to generate recommendations [4] for related projects in real-time and before publication. Our study findings thus help specify initial boundary conditions for such a technical solution to support the discovery of ongoing research, specifically, the importance of designing for trust and timeliness.

To address the trust requirement, recommendation approaches for ongoing research must offer users full control over sharing research data and, if required, data confidentiality. Research ideas and preliminary results should only be revealed in a user-controllable process if recommended researchers agree to initiate contact and share data to explore a collaboration. Any data exchanged should additionally be securely timestamped and verifiable, e.g. using trusted timestamping [9].

Addressing timeliness entails that the discovery of ongoing research must be supported before it is 'too late' to collaborate, i.e., such a system must analyze work-in-progress research, not only published research in order to solve this challenge. Researchers should be able to indicate how far along in a research project they are and for which specific topics within a larger project they are seeking collaborators. A real-time research collaboration recommendation system must additionally provide convenience to address the identified concern of 'wasting time' and 'effort.' The effort required of users must be lower when compared to the current strategies. Thus, user profiles must be automatically generated and matched in a fully confidential and privacy-preserving manner with only minimal effort required from the user.

In future research, we aim to explore and adapt privacy-preserving feature matching methods to analyze researchers' work-in-progress (WIP) resources, i.e., the data that researchers implicitly or explicitly generate or use during their ongoing research efforts, in order to generate timely recommendations of similar in-progress research work and potential collaborators.

5 CONCLUSION

Identifying ongoing research, even before it has been published, is an information seeking task not being actively addressed by today's academic search or recommendation systems. This leads researchers to rely heavily on personal connections and face-to-face communication to discover in-progress research efforts. We discuss our findings from a qualitative study of the behavior and concerns of computer science researchers to better understand the challenges they face in this task. We identify a shared internal struggle that we term the discovery-confidentiality trade-off. On the one hand, researchers express a need to discover ongoing research in their field, yet on the other, they wish to keep details about their own research confidential. We describe the researcher's perception of this trade-off and discuss its underlying factors, namely the importance of trust and timeliness in the discovery of ongoing research. This paper presents first steps towards the design of new approaches and systems that could better support ongoing research discovery by providing timely recommendations, while ensuring the confidentiality of a user's in-progress research data to satisfy the trust requirement.
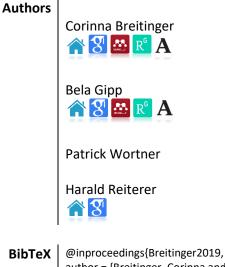
ACKNOWLEDGMENTS

REFERENCES

[1]   Abramo, G., D'Angelo, C.A., and Di Costa, F. Research collaboration and productivity: is there correlation? *Higher Education 57*, 2 (2009), 155–171.
[2]   Afzal, M.T. and Maurer, H.A. Expertise Recommender System for Scientific Community. *Journal of Universal Computer Science 17*, 11 (2011), 1529–1549.
[3]   Beaver, D. and Rosen, R. Studies in scientific collaboration: Part I. The professional origins of scientific co-authorship. *Scientometrics 1*, 1 (1978), 65–84.
[4]   Beel, J., Gipp, B., Langer, S., and Breitinger, C. Research-paper Recommender Systems: A Literature Survey. *International Journal on Digital Libraries*, 1432–5012 (2015), 1–34.
[5]   Bozeman, B., Fay, D., and Slade, C.P. Research collaboration in universities and academic entrepreneurship: the-state-of-the-art. *The Journal of Technology Transfer 38*, 1 (2013), 1–67.
[6]   Chesebro, J.W. and Borisoff, D.J. What makes qualitative research qualitative? *Qualitative Research Reports in Communication 8*, 1 (2007), 3–14.
[7]   Ellis, D., Cox, D., and Hall, K. A comparison of the information seeking patterns of researchers in the physical and social sciences. *Journal of documentation 49*, 4 (1993), 356–369.
[8]   Ge, X. Information-seeking behavior in the digital age: A multidisciplinary study of academic researchers. *College & Research Libraries 71*, 5 (2010), 435–455.
[9]   Gipp, B., Meuschke, N., and Gernandt, A. Decentralized Trusted Timestamping using the Crypto Currency Bitcoin. *Proceedings of the iConference 2015*, (2015).
[10]   Haines, L.L., Light, J., O'Malley, D., and Delwiche, F.A. Information-seeking behavior of basic science researchers: implications for library services. *Journal of the Medical Library Association: JMLA 98*, 1 (2010), 73.
[11]   Hemminger, B.M., Lu, D., Vaughan, K.T.L., and Adams, S.J. Information seeking behavior of academic scientists. *Journal of the Association for Information Science and Technology 58*, 14 (2007), 2205–2225.
[12]   Kong, X., Jiang, H., Yang, Z., Xu, Z., Xia, F., and Tolba, A. Exploiting publication contents and collaboration networks for collaborator recommendation. *PloS one 11*, 2 (2016).
[13]   Lee, S. and Bozeman, B. The impact of research collaboration on scientific productivity. *Social Studies of Science 35*, 5 (2005), 673–702.
[14]   Li, J., Xia, F., Wang, W., Chen, Z., Asabere, N.Y., and Jiang, H. Acrec: a co-authorship based random walk model for academic collaboration recommendation. *Proceedings of the 23rd International Conference on World Wide Web*, (2014), 1209–1214.
[15]   Liao, C.H. How to improve research quality? Examining the impacts of collaboration intensity and member diversity in collaboration networks. *Scientometrics 86*, 3 (2010), 747–761.
[16]   Liu, P., Curson, J., and Dew, P. Use of RDF for expertise matching within academia. *Knowledge and Information Systems 8*, 1 (2005), 103–130.
[17]   Makarov, I., Bulanov, O., and Zhukov, L.E. Co-author Recommender System. *International Conference on Network Analysis*, (2016), 251–257.
[18]   Niu, X., Hemminger, B.M., Lown, C., et al. National study of information seeking behavior of academic researchers in the United States. *Journal of the Association for Information Science and Technology 61*, 5 (2010), 869–890.
[19]   Punnarut, R. and Sriharee, G. A researcher expertise search system using ontology-based data mining. *Proceedings of the Seventh Asia-Pacific Conference on Conceptual Modelling-Volume 110*, (2010), 71–78.
[20]   Sonnenwald, D.H. Scientific collaboration. *Annual Review of Information Science and Technology 41*, 1 (2007), 643–681.
[21]   Tang, J., Wu, S., Sun, J., and Su, H. Cross-domain collaboration recommendation. *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, (2012), 1285–1293.
[22]   Tang, J., Zhang, J., Zhang, D., Yao, L., Zhu, C., and Li, J. Arnetminer: An expertise oriented search system for web community. *Proceedings of the 2007 International Conference on Semantic Web Challenge-Volume 295*, (2007), 1–8.
[23]   Teixeira da Silva, J. ResearchGate projects: unregulated academic social media. *Social Communication 3*, 1 (2017), 6–13.

Additional Information

Related Research | https://www.gipp.com/pub/

Authors

Corinna Breitinger

Bela Gipp

Patrick Wortner

Harald Reiterer