"It answers questions that I didn't know I had": PhD students' evaluation of an information-sharing knowledge graph

Stanislava Gardasevic

*Department of Communication and Information Sciences,
University of Hawai'i at Manoa, Honolulu, Hawai'i, USA, and*

Manika Lamba

*School of Information Sciences, University of Illinois at Urbana-Champaign,
Champaign, Illinois, USA*

Abstract

Purpose – Interdisciplinary PhD programs can be challenging as the vital information needed by students may not be readily available; it is scattered across the university's websites, while tacit knowledge can be obtained only by interacting with people. Hence, there is a need to develop a knowledge management model to create, query and maintain a knowledge repository for interdisciplinary students. This study aims to propose a knowledge graph containing information on critical categories (faculty, classes, dissertations, etc.) and their relationships, extracted from multiple sources, essential for interdisciplinary PhD students. This study evaluates the usability of a participatory-designed knowledge graph intended to facilitate information exchange and decision-making.

Methodology – The authors used data from multiple sources (such as university websites, faculty profiles, publication and dissertation metadata and crowdsourced data) to generate a knowledge graph in the Neo4J Bloom platform. The authors recruited 15 interdisciplinary PhD students using convenience sampling from the University of Hawai'i at Mānoa at various PhD stages to design and populate the knowledge graph. Next, the authors conducted a mixed methods study to perform its usability evaluation. First, the authors engaged the students in a participatory design workshop to identify relevant graph queries. Second, the authors conducted semi-structured interviews to determine the usability of the knowledge graph and rate the queries. Each interview was coded with structural and thematic codes and was further analyzed using sentiment analysis in R programming language.

Findings – The usability findings demonstrate that interaction with this knowledge graph benefits PhD students by notably reducing uncertainty and academic stress, particularly among newcomers. Knowledge graph supported them in decision-making, especially when choosing collaborators (e.g. supervisor or dissertation committee members) in an interdisciplinary setting. Key helpful features are related to exploring student–faculty networks, milestones tracking, rapid access to aggregated data and insights into crowdsourced fellow students' activities. However, they showed concerns about crowdsourced data privacy and accessibility. Although participants expressed the need for more qualitative data in the graph, they noted it helped identify people to talk to about the topics of their interest.

Originality – The knowledge graph provides a solution to meet the personalized needs of doctoral researchers and has the potential to improve the information discovery and decision-making process substantially. It also includes the tacit knowledge exchange support missing from most current approaches, which is critical for this population and establishing interdisciplinary collaborations. This approach can be applied to other interdisciplinary programs and domains globally.

Keywords PhD students, Usability evaluation, Knowledge graph, Interdisciplinary programs, Neo4J Bloom

Paper type Research paper

Introduction

Obtaining a PhD can be perceived as a self-discovery journey, promising the potential of a rewarding academic career and freedom (Wood *et al.*, 2016). But, for many, graduate studies can be very stressful and strenuous (Bair and Haworth, 2005; Grady *et al.*, 2014). A significant percentage of PhD students have reported experiencing anxiety and depression (Levecque *et al.*, 2017; Woolston, 2019), loneliness or imposter syndrome (Chakraverty, 2020). Also, the degree may take a long time to obtain. For example, in the USA, the median range to complete PhD is 6–12 years, depending on the discipline (National Center for Science and Engineering Statistics, 2021), while the dropout rate is estimated at a disturbing 50% (Cassuto, 2013). Success factors are often related to navigating the dissertation process (Young *et al.*, 2019), and the so-called "departmental culture," including student/faculty relationship, student satisfaction with the program, student-to-student interactions (Bair and Haworth, 2005), including peer mentorship (Preston *et al.*, 2014). The problem with navigating "departmental culture" exacerbates in interdisciplinary PhD programs (such as those that involve faculty from multiple departments, as in the case of iSchools), because students have to "span boundaries between areas, departments, and knowledge bases" (Gardner *et al.*, 2011).

This paper details the development and evaluation of a knowledge graph that tackles those issues by answering the information needs of the PhD student population, allowing them to access and share crucial, often-tacit knowledge and identify potential community connections.

PhD students need support to make well-informed decisions when selecting a supervisor and collaborators and to navigate the program and dissertation process. Still, not all necessary information can be found online. The most pertinent knowledge exchange happens through in-person conversations, serendipitously (Twidale *et al.*, 1997) and is often referred to as "social" information sharing (Talja, 2002). Previous research indicates that women, first-generation college students and underrepresented minority students may have a stronger need to obtain knowledge about "the unwritten, unofficial, and often unintended lessons, values, and perspectives that make up the 'hidden curriculum' in graduate education" (Wood *et al.*, 2016) but have less access to social networks where this knowledge circulates.

This study engaged participatory approaches to involve students in an interdisciplinary PhD program in the design, development and evaluation of a *knowledge graph* created to address the above-mentioned issues. Knowledge graph is the core of a potential information system intended to democratize access to insider (tacit) knowledge needed for PhD students to complete their degrees successfully, especially for those who do not have equal access to the networks where the knowledge exchange occurs. A knowledge graph model and data set (Gardasevic, 2022) were developed to help students with decision-making on each of the important steps in the program.

Our research followed the recommendations of the ISO (2019) standard for human-centered design principles and activities for enhancing computer-based interactive systems.

The previous stages encompassed the requirements gathering study and the design and creation of the knowledge graph. This study presents the evaluation segment, where PhD students reflect on the usefulness of the data set by using it to answer both specific queries and engage in exploratory searches.

Case background – the interdisciplinary PhD program

This research is situated in the Interdisciplinary PhD Program in Communication and Information Sciences (CIS) at the University of Hawai'i at Mānoa (UHM). The program has 20–30 students at any given point in time (with over 100 alumni) and includes faculty members from four units across three colleges: *Social Sciences*, *Natural Sciences* and the *Shidler College of Business*. Even though there are over 40 faculty members appointed to the CIS graduate faculty, only the "core" faculty are interested in interdisciplinary topics such as social or health informatics, information policy, human–computer interaction, etc. The degree requirements for this program include:

- attending weekly interdisciplinary seminars;

[first stage]:

- taking three core courses;
- passing three comprehensive exams;
- publishing a paper under the supervision of one of the CIS-affiliated faculty;

[pre-proposal stage]:

- selecting a dissertation research supervisor and four committee members;
- defending a dissertation proposal; and

[all-but-dissertation or "ABD" stage]:

- defending the dissertation.

CIS PhD students have the freedom to choose among a myriad of research avenues, which creates a challenge for each student to carve their own path through four different "departmental cultures" (Bair and Haworth, 2005), as well as program requirements taken as the blueprint for the graph design. The official program website suggests the five-year timeline for its completion.

Communication and information sciences knowledge graph and related work

In our previous work, we engaged community members via a series of interviews and workshops to develop the knowledge graph model representing essential categories, relationships and data relevant to an interdisciplinary PhD student population (Figure 1). This knowledge graph is a multilayered/multiplex network, representing *People* through 12 layers or dimensions.

Most recommender systems research focuses on scholarly networks that cover relatively few dimensions, mainly based on publication metadata (e.g. Ortiz Vivar *et al.*, 2022). Knowledge graphs in educational settings have been similarly limited in their dimensions of interest. For example, research by Rahdari *et al.* (2021) proposes a recommender system matching prospective graduate students with faculty who have expertise in an area and are affiliated with a particular school/department. Research by Jordão *et al.* (2014) focuses on visualizing patterns elicited from course-taking activity. The KCUBE project developed a knowledge graph to support advising students in computer science regarding curriculum

Figure 1.
CIS knowledge graph model
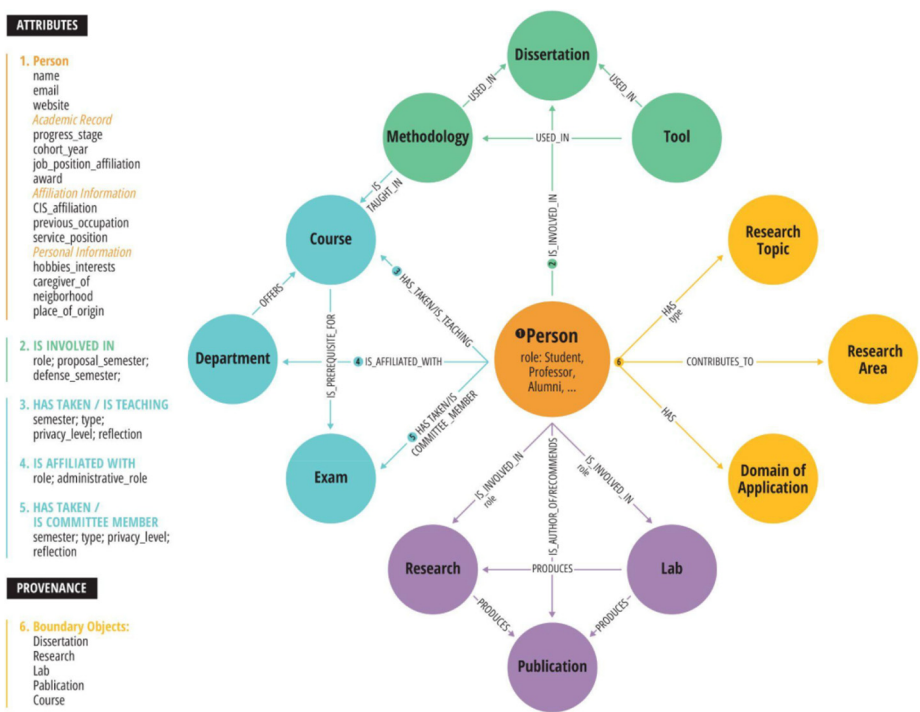
Source: Gardasevic and Gazan (2023)

and career planning (Li *et al.*, 2022). However, to understand the full complexity of social structures, researchers suggest that three dimensions are considered minimum (Dickison *et al.*, 2016). Considering scholars' many hats (e.g. teaching, mentoring, researching, publishing, administrating science, etc.) (Börner, 2010), the graph produced for this study attempts to comprehensively represent these activities through the prism of PhD students' information needs.

Due to the current state of the art in the area of automatic tagging (Banerjee *et al.*, 2022), much of the data in the CIS graph was manually curated and captured with the Neo4J graph database. The refining attributes in this graph are intended to support not only information discovery and decision-making but also social connection-building and information sharing in the community, as knowledge seekers are hesitant to ask for information unless they feel they have a close enough relationship with those who could share it (Keppler and Leonardi, 2023).

Relevant literature
Knowledge graphs as medium for information discovery
The research on the dynamics of multidimensional social and knowledge networks is a promising avenue, especially for facilitating transdisciplinary collaborations (Contractor, 2009). Knowledge graphs are often used for knowledge organization and developing systems to support scientific information discovery and expert recommendation (e.g. Callahan *et al.*, 2023; Osborne *et al.*, 2013).

Another affordance of the graph representation of local social connections is the possibility to identify the people in one's domain who can offer direct advice – especially other students who are invaluable sources of information, yet traditionally excluded from expert recommender systems (Wood *et al.*, 2016). Preston *et al.* (2014) outlined that "peer mentorship is an under-utilized resource with great capacity to foster human and social capital within and between cohorts of graduate students."

Graph-based knowledge organization systems have tended to be monodisciplinary (Osborne *et al.*, 2013), aiming to represent scholars through their universal set of activities (Corson-Rikert *et al.*, 2012), which can be applied to local settings. In this study, we were inspired by the Hawaiian epistemological approach to knowledge (Meyer, 2001), where "*knowing*" cannot be separated from the physical place where it emerged and depends on one's relationships and lineage, providing a bottom-up approach to knowledge organization system creation.

Knowledge graph to support interdisciplinary science and education

Interdisciplinary science is becoming increasingly common (Porter and Rafols, 2009), more successful at making breakthroughs and generating more relevant outcomes than monodisciplinary research (Fortunato *et al.*, 2018). Even though the CIS knowledge graph model was developed to satisfy the specific needs of the PhD students in a particular interdisciplinary program, the categories and relationships, as declared in the knowledge graph schema, can be applied to other academic domains and settings. For example, it can be applied to facilitate interdisciplinary or multidisciplinary research in fields such as information science/iSchools (Wiggins and Sawyer, 2012), life sciences (Franc-Dąbrowska *et al.*, 2020) or climate change research (Bruine de Bruin and Morgan, 2019).

Knowledge graph evaluation-related work

Knowledge graph evaluation methods, most commonly using a probabilistic approach, deal with knowledge graph completion and error detection and the research is mainly focused on DBpedia, which is the most prominent knowledge graph (Paulheim, 2016). Previous research by Sarrafzadeh *et al.* (2014) used mixed methods to examine general users' exploratory search behavior when interacting with knowledge graphs and pertinent corresponding documents based on Wikipedia articles; the authors consequently developed so-called hierarchical knowledge graphs (Sarrafzadeh *et al.*, 2016, 2020) that were evaluated via experimental approach. Our contribution to this body of research is methodological, i.e. engaging the intended end users in a qualitative evaluation study of a knowledge graph that was created based on their information needs. Even though qualitative approaches are applied to gather requirements and evaluate visual analytical tools to support decision-making in other areas (Dimara *et al.*, 2022; Hong *et al.*, 2020), there is a lack of literature that uses such an approach to investigate the end users' experiences with knowledge graph tools, which is critical when it comes to their adoption (Li *et al.*, 2023).

Research questions

The research questions guiding this study were formulated to evaluate the usefulness and potential impact of the knowledge graph created based on extensive requirements gathering study and inputs of end users:

RQ1. What are the reactions and impressions of CIS PhD students to the knowledge graph?

RQ2. Which knowledge graph features are perceived as helpful? How do students talk about their next steps in the PhD program upon interacting with knowledge graph?

RQ3. What information was missing from the knowledge graph?

Methodology

The entire methodological approach as conducted in the overall study is shown in Figure 2. The study was divided into three stages, pertinent to human-centered design approach, and different data collection methods were applied in each stage:

(1) In the requirement gathering stage, we conducted interviews and the program website usability study with PhD students and Gardasevic (2020).

(2) In the graph design and data ingestion stage, we conducted three workshops, scraped the web, indexed publications and dissertations and collected and ingested the crowdsourced data into Neo4J graph database (Gardasevic and Gazan, 2023).

(3) The methods conducted in this study's final evaluation stage are described in detail in the following text, and they encompass the participatory design workshop, a survey and semi-structured interviews with CIS PhD students.

Data collection

Our data collection approach in this phase was three-fold:

(1) a participatory design workshop;

(2) follow-up survey; and

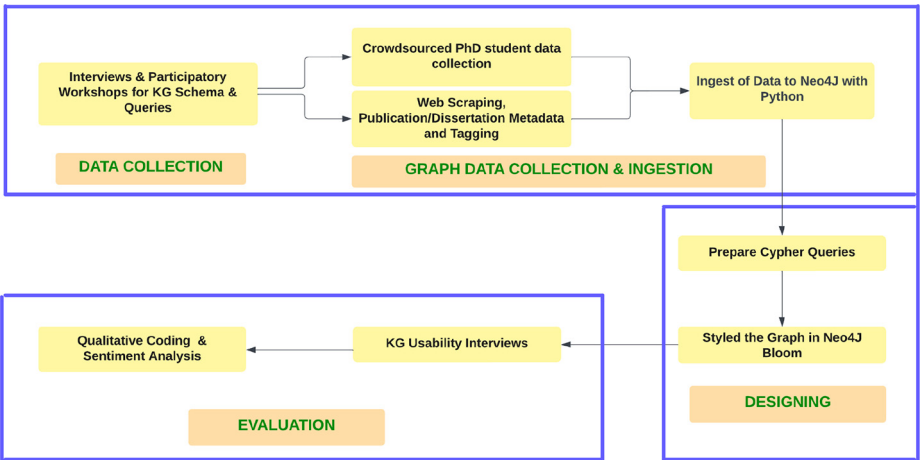(3) in-depth interviews with CIS students following the knowledge graph evaluation protocol.

Figure 2.
Summary of the methodological approach

Source: Figure created by authors

These approaches were chosen to gather and address both group and individual insights about the proposed technology. The follow-up survey allowed the participants to voice their opinions asynchronously and anonymously.

The interdisciplinary CIS PhD students were invited to participate in this study via CIS student mailing list and were asked to sign the consent form. All of the data was collected by the first author.

Participatory design workshop
In June 2022, a participatory design workshop convened, drawing the active engagement of *15 PhD students,* split between in-person ($n = 8$) and online ($n = 7$) participation.

The workshop was concurrently scheduled with the group's weekly interdisciplinary seminar, providing a convenient opportunity for students to participate without making it mandatory. The participants were not offered any monetary or other incentives.

The workshop agenda included a preview of the knowledge graph model, specific queries and a community discussion on potential system functionalities. Furthermore, students were actively involved in two distinct voting exercises: one focusing on refining the multiplex graph visualization design and the other prioritizing useful queries for this population. The students were invited to suggest other specific questions they found significant.

Follow-up survey. The workshop was followed up by the post-workshop survey (in Google Forms) that was distributed to the CIS student mailing list, and contained several open and closed-ended questions evaluating the usefulness of the graph demonstrated. Some survey results ($n = 8$) showed that participants perceived the tool as useful. To investigate further which features students found useful, we developed a graph evaluation protocol that addresses the research questions.

Semi-structured interviews and knowledge graph evaluation protocol
In spring 2023, the authors interviewed *eight* CIS PhD students (38% of the population) using the *graph evaluation protocol* [1]. We applied convenience sampling, and the first author directly emailed potential participants to invite them to participate in the study. Each interview was 82 min long on average, and participants were offered a US$50 gift card [2] for their partaking. We used a QuickTime player to screen and voice record each interview and the interactions with the graph.

Even though the graph was primarily designed to support new students and their information needs, we invited students from different stages in the interdisciplinary CIS PhD program to evaluate it. Among the study participants, four were from the latest cohort (or "new" students); two were in the second year of the program (or "mid-stage" students) and two were in their finishing stage ("all-but-dissertation" or ABD students). Of this population, six were domestic/US students, and two were international. We stopped interviewing new participants as we noticed a thematic saturation (Saunders *et al.*, 2018) that emerged from the data collected from participants in each of the stages.

We used the Neo4J Bloom graph exploration tool, its styling capabilities and natural language querying options to showcase and evaluate the CIS knowledge graph. Led by the results from the participatory design workshop, we created a list of 20 queries voted as useful at least three times (Appendix 1) that were the crux of the *graph evaluation protocol*. The *protocol* encompassed:

- identifying categories of students' interest in the controlled vocabulary entries [3] used to tag *publications* and *dissertations*;
- asking for current information needs and exploring the graph to seek answers;

- rating each of the 20 queries on a five-point Likert scale, with occasional prompts to reflect on their ratings; ending with; and
- five summative questions.

Three interviews were conducted in person, during which students engaged in hands-on exploration of the graph in two instances. Throughout all the interviews, the researcher played the intermediary role in guiding the participants through the graph exploration exercises. This was perceived appropriate considering the complexity of the interface tool (Neo4J Bloom) and the fact that the researcher provided an "*informal advising*" session that accompanied the evaluation protocol, sharing her views, advice and experiences related to the program, as she was demonstrating the graph queries and facilitating the exploration.

Figure 3 shows an example of the interface, demonstrating the exploration of a particular faculty member and the *research areas* he publishes in (yellow nodes). Such queries were often expanded to show other listed categories based on participants' interests.

We were not evaluating Neo4J Bloom's usability, but rather the information that can be obtained from the knowledge graph.

Data analysis
The interviews were transcribed using the Otter.ai [4] transcription tool, after which both authors revised the automatically generated text for accuracy. The transcriptions were anonymized, and analyzed in Atlas.ti cloud software. We coded each interview (with structural and thematic codes) and grouped them into overall themes reported in the following section. Structural coding assigns concise phrases to data segments relevant to specific questions of inquiry (Saldaña, 2016), and in our study, these codes correspond to graph categories. Thematic coding involves identifying, analyzing and reporting patterns or
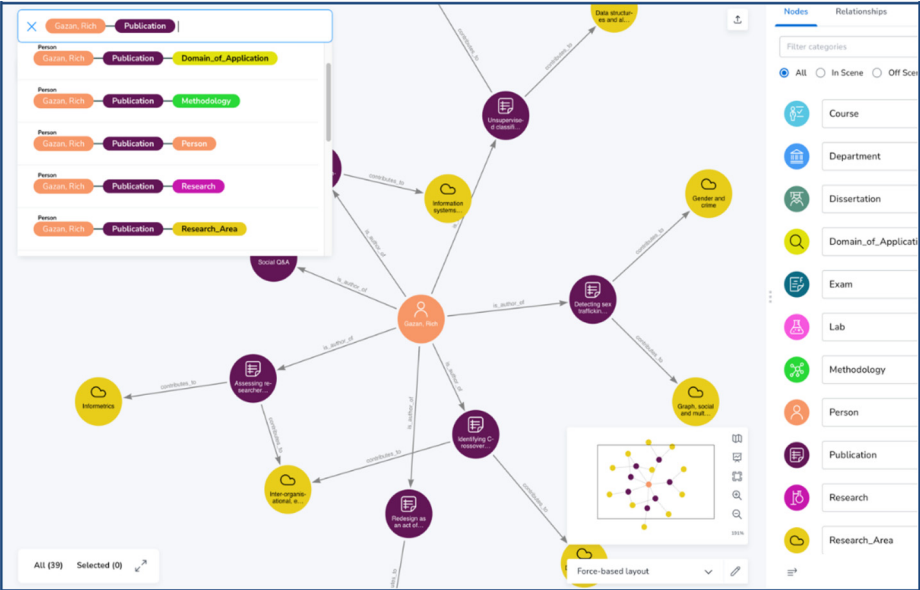

Figure 3.
Example of the knowledge graph interface in Neo4J Bloom

Source: Figure created by authors

themes within collected data to capture important aspects in relation to the research questions (Braun and Clarke, 2006).

Furthermore, Appendix 2 demonstrates how we conducted sentiment analysis using the *syzuhet* library (Jockers, 2015) in R. This analysis served two purposes:

(1) assessing how the narrative was structured to convey sentiments across participants' responses; and

(2) identifying emotions (*disgust, anger, anticipation, surprise, trust, fear, sadness and joy*) and sentiments (*positive and negative*) within the participants' responses.

Positionality and limitations

The first author is an insider to the program, a PhD student herself and a study participants' peer. Aside from the obvious advantages of this approach that encompass her familiarity with jargon, social rules and norms, the drawbacks of the approach have to do with the "courtesy bias" on the participants' side that might have resulted in more favorable ratings, than it would have been if the researcher was an outsider.

The second author is an outsider to the program, yet a recent PhD graduate who went through the same process, just in a different country and culture; her insights serve as an objective check and set of eyes on the data and results.

Findings

This section gives an overview of qualitative coding and sentiment analysis results. The codes (reported as *bolded text*) were organized in themes pertinent to the research questions and reported as subsections and sections, respectively. We report the number of code occurrences over all interviews as "*c*."

Reactions and impressions of communication and information sciences PhD students to the knowledge graph

Seven out of eight participants found the graph exploration results *interesting* ($c = 35$), and four were *surprised* ($c = 18$). On the other hand, five participants got results that showed *known information* ($c = 10$), while four got to *confirm their beliefs* ($c = 9$).

The discussions accompanying the graph exploration tasks called for reflective commentary on the participants' side. To demonstrate the overall flow of conversation and the participants' sentiments during the interview, we refer to Figure 4, for which we used a binning average method (Jockers, 2015) to track changes in polarity in participants' responses. Figure 4 shows that the new students have ended the interview and graph exploration exercise (conditionally taken as an intervention) with a much more optimistic outlook when compared to ABD students.

Figure 5(a)–(b) illustrates the distribution of emotions and sentiments expressed by the participants during the interview, providing valuable insights into their overall experience of witnessing the knowledge graph in a live demo. The findings reveal a highly positive reception among the participants [Figure 5(a)]. The analysis reveals that "*trust*" emerged as the most prevalent emotion among the participants when the knowledge graph was used for information retrieval. This suggests that the participants had confidence in the knowledge graph's capabilities and reliability to provide accurate and relevant information. Such trust is crucial for successfully adopting and accepting knowledge graph tools (Li *et al.*, 2023). Following *trust*, "*anticipation*" and "*joy*" were among the prominent emotions expressed by the participants [Figure 5(b)]. This finding indicates that the participants were excited about
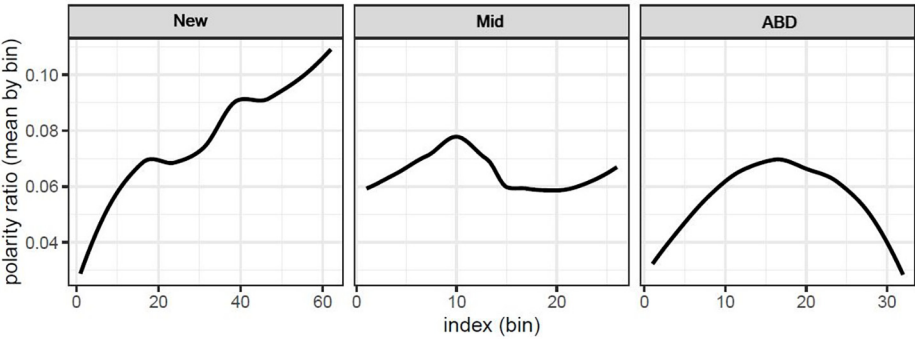
Source: Figure created by authors

the possibilities offered by the knowledge graph and felt a sense of pleasure or happiness while interacting with it. Further, *"anticipation"* often suggests an eagerness to explore and discover new information through the knowledge graph.

Different approaches to the knowledge graph based on the program stage
Students in different stages of the program had different experiences with the graph. That was evident both from the sentiment analysis results, structural coding and the results from the query ratings (Appendix 1). For example, when participants were asked to start a search on a category of their interest, the new students wanted to learn more about *courses/exams* and *faculty*. By contrast, the second-year students started searches to identify pathways of their *peers* (e.g. where and when they published papers), *faculty* engagements in *dissertations* or research-related questions (e.g. exploring *methods* of their interest). The *faculty research and publishing* information was considered more valuable to the students closer to finishing the degree and conducting research themselves.

PhD life hardships and peer support via *knowledge graph*. The two most prominent difficulties pertinent to the CIS program are its interdisciplinary nature and the lack of support. Following is one reflection on the program organization, where a new student lacked guidance when they first started the program:

> It was a bit disorganized, as a new student was not sure what I was doing. So much new information.

Even though the guidelines clearly indicate the expected milestones for PhD students, many newcomers experience uncertainty, even "imposter syndrome," in the first year of their program. They benefit from hearing experiences and tips from those who have already gone through them; still, such exchange does not happen often in this program. One student described their PhD journey as being a "super lonely path,' especially in an interdisciplinary program such as CIS, where there are no labs or student collaborations:

> In the CIS, you are doing your own thing. Sometimes people would have no idea what you're doing except for you and your supervisor.

Thus, students valued the access to information crowdsourced from their peers, which served as motivation for some or as a comforting factor for others – a reinforcement that they are on the right path.
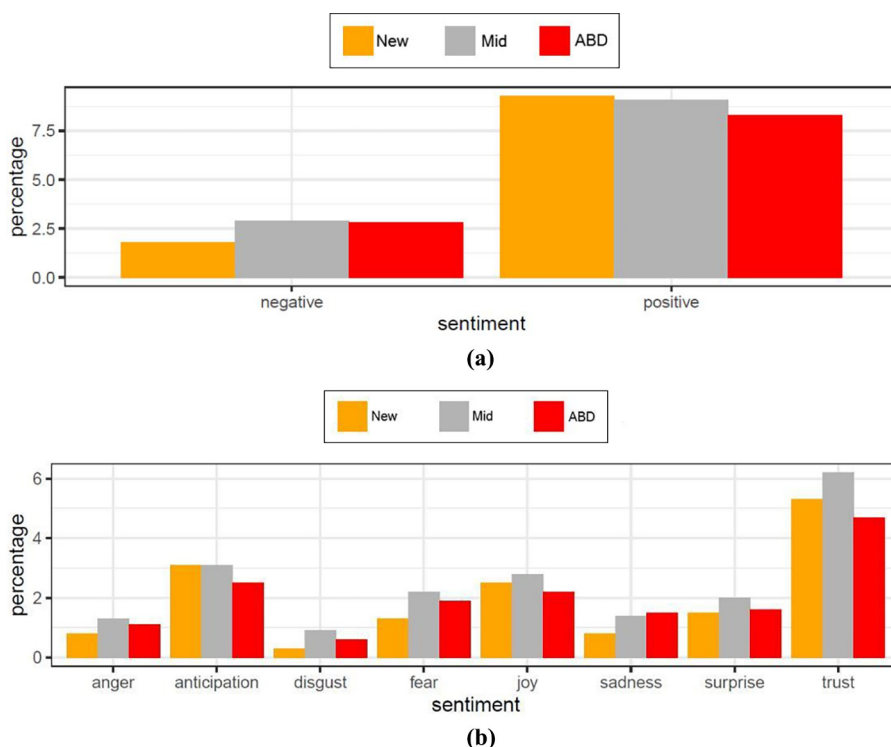
Figure 5.
Percentage of
emotions and
sentiments in the
participants'
responses

Notes: (a) Sentiments; (b) emotions
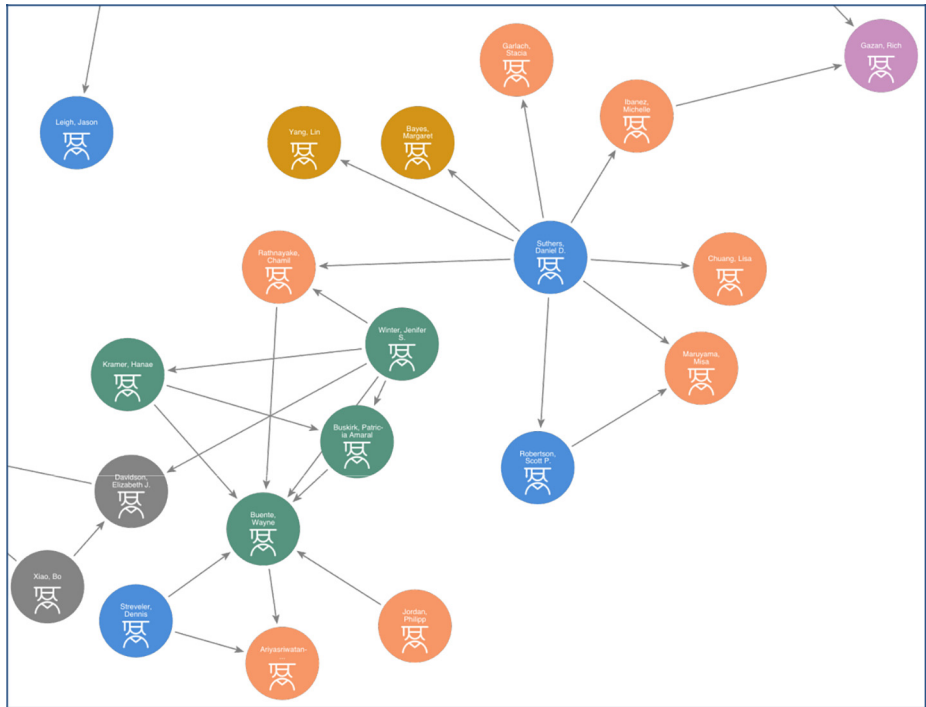**Source:** Figure created by authors

Helpful features of communication and information sciences knowledge graph
Throughout the interviews, the words *"helpful"* and *"useful"* were explicitly used over 50
times to refer to different aspects of the knowledge graph. The *whole graph exploration
experience,* including the visualization of connections among people and concepts in the
graph, was perceived as the most helpful, compared to any individual feature. A "mid-stage"
student commented on that with:

> A lot of people who are [. . .] part of these types of networks probably aren't even aware of who's
> in this whole network. So, for you to have kind of a bird's eye view of who's all involved, helps
> you to be able to take the right steps [. . .] towards that problem-solving.

The following text outlines specific themes participants found helpful when interacting with
the knowledge graph.

Previous students' activities. Information that was found very useful by most PhD
students was the data on the activities of their peers and predecessors. Newer students who
needed to plan where to publish papers and consider the faculty members for collaboration
found the queries around *student publications* and their *co-authorships with faculty*
informative, considering there is no easy way to access this data currently. Figure 6 shows
the screenshot of a part of the co-authorship query in which a student could recognize which
faculty members publish with their students (orange and yellow nodes) vs those who work
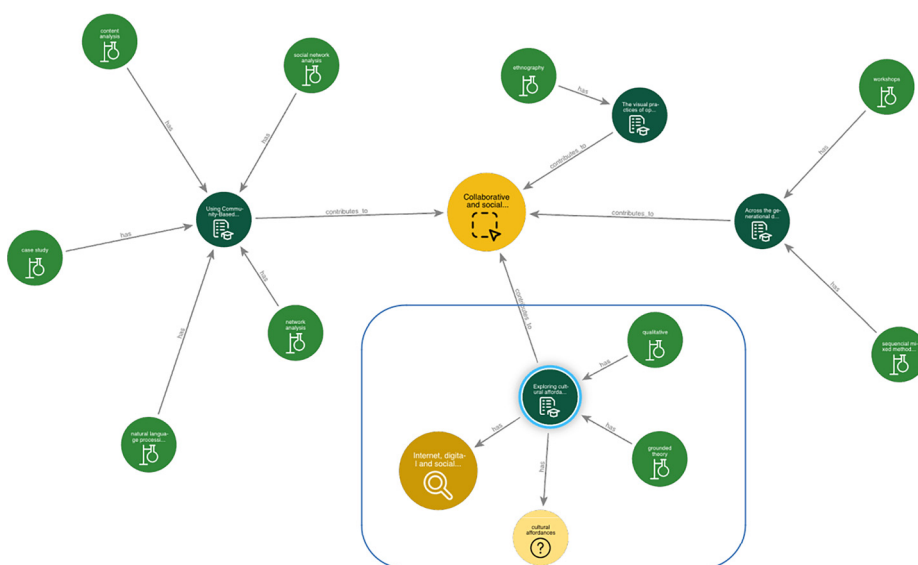
Source: Figure created by authors

with other faculty in the same department (same color nodes) or have cross-departmental collaborations.

Similarly, participants appreciated the ability to discover and explore *previous students' dissertations*, their committees and other aspects, such as *methods* and *tools* used to perform research. For example, Figure 7 shows the screenshot of a query where a participant looked at dissertations in the *area* of "Collaborative and Social Media" and the research methods used in them. This query was expanded to investigate the dissertation of interest further by showing its *research topic* and *domain of application*.

Student reflections on courses/exams. Seven out of eight participants found the *students' reflections on classes* useful. A new student commented on this feature:

> When you go through the class availability, and all of the hundreds of frickin' departments; [...] it's really overwhelming to figure out what type of courses are useful [...] I think something like this would show me- other people in our department took these types of courses-here are some that might be helpful.

Similarly, *students' reflections on "directed readings"* with faculty members gave new students insights into individual professors' supervision styles. This information was found particularly useful, even critical, next to other aspects that can support students' *decision-making* when it comes to picking the supervisor or committee members:

> The other piece that was helpful was just reading feedback from other peers on the selection of their chairs or working with the chairs. It's one lens, one perspective, but it's a data point for me to consider. So, that brings me just kind of awareness before I make the decision.

Source: Figure created by authors

Faculty exploration. Exploration of *people* was one of the features that students found very useful, especially when *exploring faculty* based on various aspects of their activities, such as *research topics/areas/domains* they work on, *methodologies* they work with and their current engagements in the *dissertation* committees. Upon graph exploration, a "mid-stage" student who is working on forming a dissertation committee identified potential collaborators they had not considered before:

> Yeah, [the query] narrowed it down a little bit [. . .] But definitely, I would look at [two professors shown in the recall] as being people who are kind of outside of what I was looking at before.

Supporting interdisciplinary research. When invited to comment on the utility of the graph, one participant emphasized the importance of its functionality to identify people from different disciplines to collaborate with on *interdisciplinary research,* as monodisciplinary work has its boundaries. To that, a "mid-stage" student said:

> I want to make an interdisciplinary group on a topic that we might all understand- that's important to me.

Figure 8 shows the screenshot of a query that represents faculty from different departments (indicated by colors) publishing in the same *domain of application* (i.e. HCI).

Aggregated data saves time. Three participants reflected on the tool's utility to save time when looking up/searching for the information they need, as the graph aggregates the data from multiple sources to allow for this functionality (Figure 9). Two participants found the controlled vocabularies elicited from this corpus useful to explore categories of their interest. Answering one of the protocol follow-up questions (*Did you find the data you wouldn't be able to find otherwise?),* participants reflected on the hardships of finding not only specific information but rather all the information that was discussed during the interview. An ABD student commented:

Figure 8.
Faculty from different departments publishing on human–computer interaction
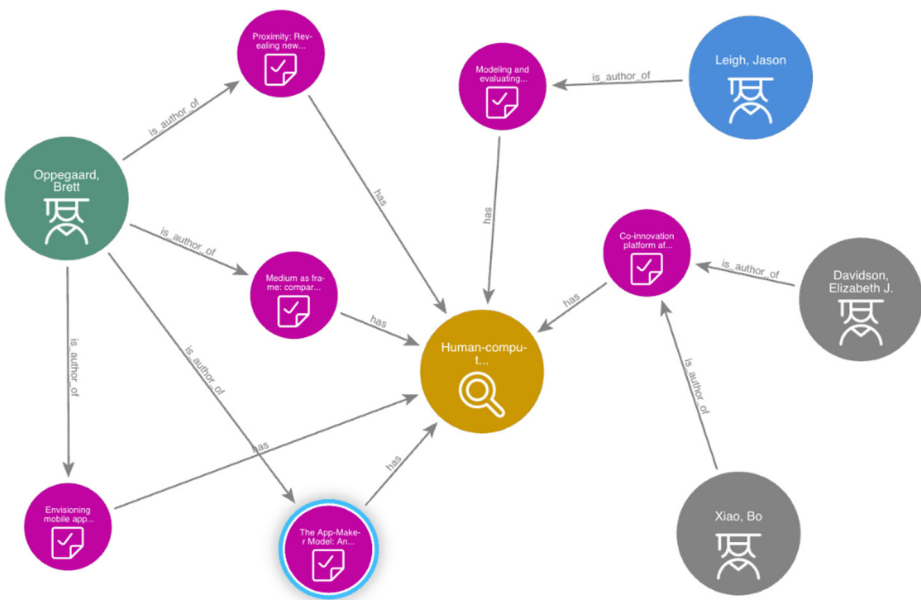
Source: Figure created by authors

Figure 9.
Sources of data for CIS Knowledge graph
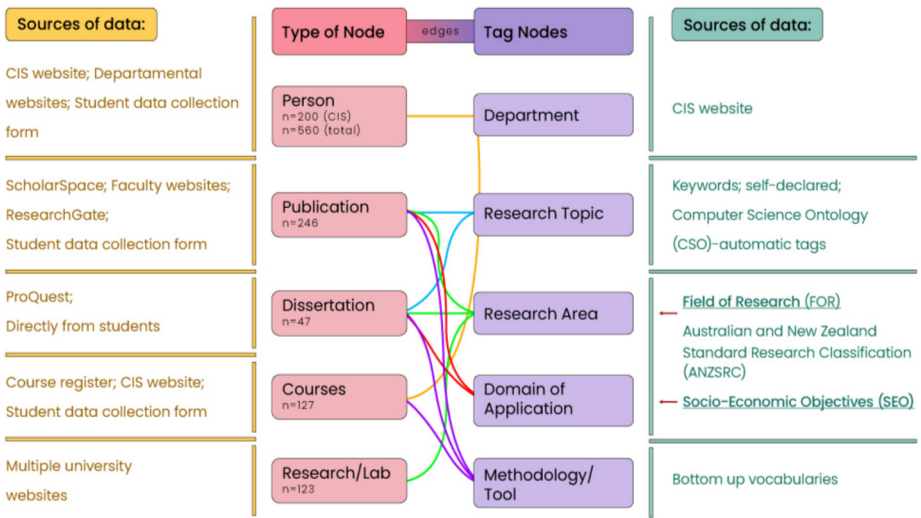
Source: Figure created by authors

So, I think a lot of it would be difficult to find, at least all together. Like, you could do this yourself, and it would take forever.

Planning for future steps based on knowledge graph insights
Decision-making support. Six students expressed the value of the graph, especially when it supports their decision-making process. Codes *helpful* and *decision-making* overlapped five times. When commenting on the criteria for the ratings of queries demonstrated, a "mid-stage" student said:

The closer it comes to helping me make a decision, the higher the value it has for me.

We demonstrated the queries that were voted as useful by PhD students and personalized them to showcase specific categories that were of participants' interest. This helped them to anticipate the next steps in the program. A "new" student interviewee commented on the procedure:

This process gave me a lot more information that I didn't even think I needed to know. It answers questions that I didn't know I had.

Participants found it particularly useful to inform decisions on choosing faculty to collaborate with-as a chair or a committee member ($c = 17$). Three students commented on the usefulness of the queries that showed the faculty's current mentoring engagements (example of the query screenshot in Figure 10), as the recall may indicate if they are already "stretched thin" in that segment. One "mid-stage" student participant said:

[This query] is very useful for me [. . .] I'd want to create a committee with a lot of time for me, right? I think [the professor] working on four dissertations- that's borderline.

Personal timeline, projections and reflections. Throughout the interviews, all participants reflected on their *personal timeline* in the program – not only the time to the degree but also the time to fulfill the set requirements ($c = 42$). A "new" student was reminded of their next step:

That's just like eye-opening. I didn't even [. . .] I honestly kind of forgot about the publishable paper thing. But I guess that's along that line right now.

Four participants talked about being *timeline-driven.* In such cases, this tool was considered to be very useful regarding their steps planning. A "new" student commented on that:

I'm very timeline-driven [. . .] so it's really helpful; it gives me a framework because I tend to over-plan, and this puts it in perspective for me.

The two international students emphasized the usefulness of looking at other students' timelines, as for the international student population, it is helpful to see the "true picture" of the potential length of the program, which helps them with planning around visas, scholarships and financing. Three participants found the use of the previous students' timelines to compare with them. One of the "mid-stage" student interviewees said:

I like to see where I stand amongst my peers. I think it's more important to me because I'm older. And, you know, I'm competing with a bunch of young kids, I guess.

Information missing in the knowledge graph
At the end of the evaluation protocol, participants were asked to comment on the information they would like to know that is not represented or cannot be represented by the graph. Two participants stated that the graph was comprehensive and
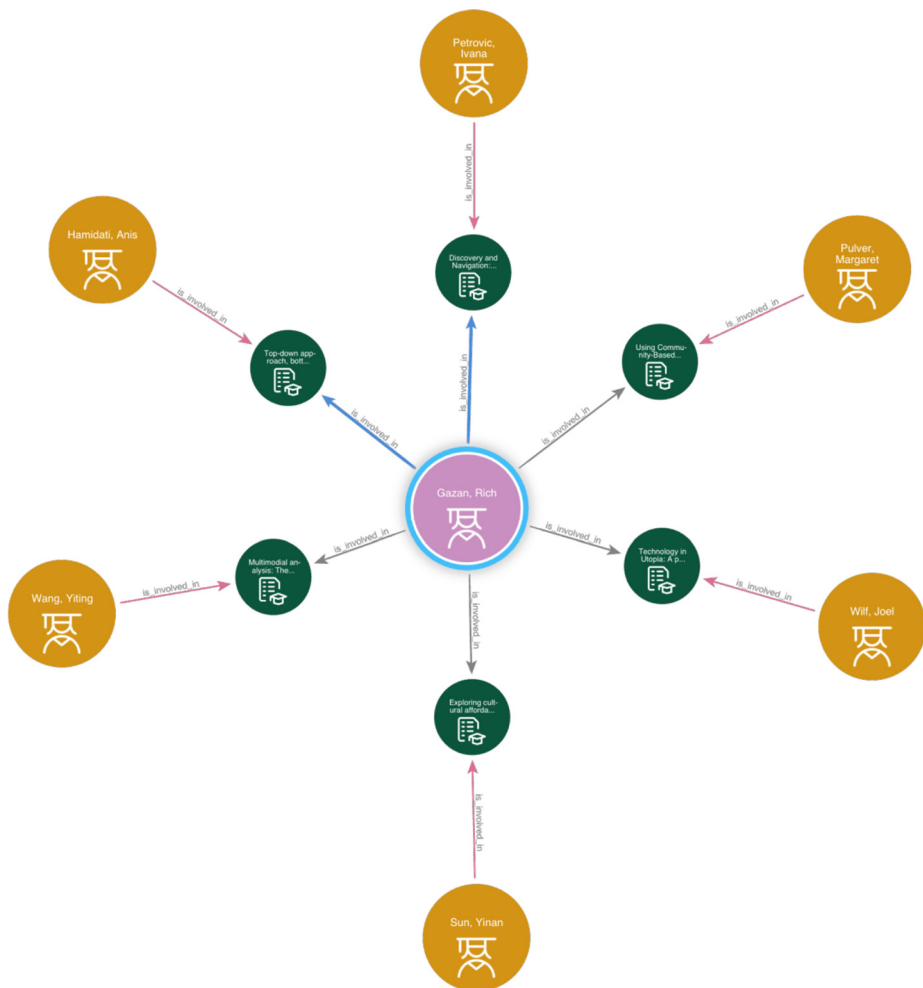
Figure 10.
Current students of a
faculty member (blue
edges indicate their
chairing role)

Source: Figure created by authors

captured all the information a CIS PhD student may need. Two participants stated it would be impossible to read negative feedback about professors, as people would refrain from writing about this. An ABD participant said that the graph is a good starting point, yet one needs to do further independent investigations on issues that are not easy to conclude from query recalls (e.g. why is a professor mentoring so few students).

The knowledge graph was noted to be good at showing facts and directing students to people who might provide more insight into events behind the numbers and statistics; to that, a "mid-stage" student commented:

I need to talk to each of these people probably, to get that [personalized information].

Similarly, students were interested in knowing more about one's reasoning behind picking a dissertation supervisor or a committee member. However, when asked if they would leave such a comment themselves, another student laughingly said:

> Oh, yeah, I mean, I am being sort of hypocritical here. I would love to get that quick and easy information, but you cannot count on me to be a contributor to that.

We noticed significant differences in the level of comfort when it comes to providing written reflections on faculty and courses between domestic students – who are generally open to it, and international students – who are very much wary of this approach and would avoid the feature altogether. One student talked about the possible misuse of this feature, as it would allow for misinformation-spreading mechanisms in the case of a very competitive program.

Three interviewees were interested to learn more about the personal experiences of students in the program, especially those who quit or were expelled; as well as the professional pathways of alumni (as they follow their degree choices).

Discussion

Previous research noted the gaps in the knowledge graph design and evaluation studies, outlining the fact that they are disconnected from the communities they are supposed to serve (Li *et al.*, 2023; McGee *et al.*, 2019). Thus, our study addresses these issues by conducting a qualitative study grounded in participatory approaches for designing the graph, its exploration scenarios and evaluation of its utility. These approaches were successful in creating a useful database that supports PhD students in information discovery and decision-making, as well as providing potential for information sharing and social connection-building in an interdisciplinary research community. The findings of this study helped us understand the thought process behind interactions with such a tool to improve future knowledge graph iterations and get an idea for the novel multiplex graph interface design.

The implication of this research is especially significant as the methodological contribution to the field, as the knowledge graph research is predominantly addressed via quantitative approach; the human-centered design steps described in this and our previous papers (omited1, omitted2) introduces the iterative development of the graph model and database population that can be replicated to other domains.

Overall reactions and impressions to knowledge graph

The overall impressions of CIS PhD students on the knowledge graph were positive. This finding was evident through the query ratings, comments coding results and sentiment analysis. Even though previous research has shown that users lose trust in knowledge graphs upon interacting with the complex graph visualizations (Li *et al.*, 2023), the sentiment analysis shows that the participants trust the CIS knowledge graph, which might be due to the choice of queries. The overall acceptance rate and positive feedback for such a tool demonstrate the advantage of the effort made to understand the end-user's needs, which are often overlooked when creating knowledge graphs (Li *et al.*, 2023).

The ability to read about the experiences and get ideas on the pathways of other students was appreciated by all participants, although most informative for new students. This is due to the fact that such information is hard to find elsewhere. PhD students found some insights comforting and seemed more confident they would achieve milestones they were initially unclear about. A new student commented on the tool:

This is a good navigator; a navigation tool that helps you to know [. . .] how to go about it; They need this. This is very handy. Very relevant [. . .] I've been struggling with all these things here, and you picked it out.

It is a promising result, indicating that the designing of a tool to support the activities of PhD students can potentially alleviate some of the stress and strain they are experiencing. Possibly, such an approach might improve student retention by increasing satisfaction with their academic programs, including the fulfillment of their PhD program expectations (Bair and Haworth, 2005).

Knowledge graph as a helpful tool
Participants found the tool helpful, especially seeing the connections/links between actors in the domain that one may not have been aware of.

The graph was perceived as helpful as much as it supported the decision-making, in particular when considering which faculty to invite to be their supervisors or to serve as dissertation committee members. This feature is particularly important as we know that many of the challenges and success factors of PhD students are directly related to the student–mentor relationship (Bair and Haworth, 2005; Grady *et al.*, 2014; Young *et al.*, 2019). The findings emphasize the value of alternative metrics of faculty, such as their mentoring engagements, which are currently not sufficiently recognized when compared to mainstream impact metrics based on publishing (e.g. citation count, h-index, etc.).

Also, the graph was helpful in personal timeline planning and motivation; students could compare their paths to those of their peers.

We will use these findings to inform the features of the knowledge graph interface, emphasizing those that may support intrinsic motivations for contributing and updating crowdsourced data (Pinto and dos Santos, 2018). For this population, those are:

- progress checks and comparisons with other students;
- projection of future steps and planning; and
- personal graph development through time.

Missing from the graph
All participants agreed that more context is needed to understand better what was happening with people behind the numbers they have seen in the metrics shown. The qualitative component of actual human experience is beyond the ability of any graph to represent.

Sometimes, information may or may not be explicitly encoded in a graph, but students would find value in simply knowing who to ask. The graph makes it easier to identify people with multiple potential shared interests and ask them about their degree choices – for example, why someone chose one supervisor over another. This way, students would be supported in initiating conversations, potentially fostering relational confidence and trusted information exchange (Keppler and Leonardi, 2023), even peer mentorship that is perceived as beneficial for PhD students' satisfaction and retention (Preston *et al.*, 2014).

Although the "previous students timeliness" feature was perceived as beneficial for the study participants to plan for success and get motivated to proceed, the downfalls of having this data available are potential enhancement of the negative feelings caused by imposter syndrome (Chakraverty, 2020) or social comparison (Kim *et al.*, 2021).

Limitations and future work

The limitations of our approach are mainly about generalizability. Even though the CIS program is small and not representative of the broader PhD education realm, it may be considered a critical example of an interdisciplinary program in which faculty from different departments are affiliated. In such programs, PhD students have to navigate various departmental cultures (Gardner *et al.*, 2011). However, to provide further use of this approach, it can be applied to other interdisciplinary programs, e.g. those offered by iSchools [5]. The iSchool related research has highlighted the potential benefits of a community portal as a forum for peer information exchange among students from various disciplines (Naughton *et al.*, 2010) and the need to support the identity formation among students in this interdisciplinary setting, and both can be addressed via this knowledge graph approach (Choi, 2015).

As our approach bears limitations of manual tagging and data input, more research is needed to identify the appropriate mechanism to ingest and update relevant publicly available data, by modifying existing approaches (e.g. Banerjee *et al.*, 2022; Müller, 2023).

Questions of data privacy and transparency must be addressed. For example, students' comments on courses (and faculty mentorship styles) were found to be extremely useful yet potentially harmful features. Also, other stakeholders could be potential users of this knowledge graph, e.g. faculty, administrators, graduate division officers and alumni. Considering the flexibility of knowledge graph as a tool (Li *et al.*, 2023), such changes could easily be implemented through another round of requirements gathering with the mentioned stakeholders.

Conclusion

This paper focused on the evaluation of a knowledge graph designed to support PhD students in their progress toward the degree and their integration into the research and educational community. We have discussed some of the observed advantages and pitfalls of having knowledge and social networks represented in a knowledge graph. We demonstrated that involving end-users in the design process can help students anticipate and plan for the next steps in their degree and potentially connect with one another.

The findings have shown that the end users found the tool helpful, especially as it supports them in making data-driven decisions, be more confident of their next steps and it saves them time when looking for information. Still, the students noted the lack of the qualitative aspect behind the data aggregated in the knowledge graph, and would like to know more of the experiences of the previous students.

This tool can be used for making more personalized recommendations than is possible through generic policies and degree requirements, as well as improving the visibility of the work done by researchers in the local setting. Such an approach overcomes the socio-technical issues that may emerge due to a lack of understanding of end users' needs but also a methodological deficiency often overlooking the qualitative approach to research by technology developers when designing knowledge graph exploration software.

Notes

1. Graph evaluation protocol: https://drive.google.com/file/d/1FuzWFTBVFYUVSx-uQ0ZVirydo66cw80Z/view.

2. The funds were received through the National Science Foundation, under Grant No. 1933803, distributed by the Science of Science Summer School (S4) program, where the co-authors met and started collaborating.

3. List of controlled vocabularies: https://docs.google.com/spreadsheets/d/1qgIuqEpebvc
   DswAC3wzdDPAfCvwx1dIvwLRCqjHKmkE/edit?usp=sharing

4. https://otter.ai/

5. https://www.ischools.org/

References

Bair, C.R. and Haworth, J.G. (2005), "Doctoral student attrition and persistence: a meta-synthesis of research", in Smart, J.C. (Ed.), *Higher Education: Handbook of Theory and Research*, Kluwer Academic Publishers, Dordrecht, Vol. 19, pp. 481-534, doi: 10.1007/1-4020-2456-8_11.

Banerjee, B., Ingram, W.A., Wu, J. and Fox, E.A. (2022), "Applications of data analysis on scholarly long documents", *Presented at the 2022 IEEE International Conference on Big Data (Big Data), IEEE, Osaka, Japan*, pp. 2473-2481, doi: 10.1109/BigData55660.2022.10020935.

Börner, K. (2010), *Atlas of Science: visualizing What we Know*, MIT Press, Cambridge, MA, available at: https://mitpress.mit.edu/9780262014458/atlas-of-science/

Braun, V. and Clarke, V. (2006), "Using thematic analysis in psychology", *Qualitative Research in Psychology*, Vol. 3 No. 2, pp. 77-101, doi: 10.1191/1478088706qp063oa.

Bruine De Bruin, W. and Morgan, M.G. (2019), "Reflections on an interdisciplinary collaboration to inform public understanding of climate change, mitigation, and impacts", *Proceedings of the National Academy of Sciences*, Vol. 116 No. 16, pp. 7676-7683, doi: 10.1073/pnas.1803726115.

Callahan, T.J., Tripodi, I.J., Stefanski, A.L., Cappelletti, L., Taneja, S.B., Wyrwa, J.M., Casiraghi, E., Matentzoglu, N.A., Reese, J., Silverstein, J.C. and Hoyt, C.T. (2023), "An Open-Source knowledge graph ecosystem for the life sciences", available at: https://arxiv.org/ftp/arxiv/papers/2307/2307.05727.pdf

Cassuto, L. (2013), "Ph.D. attrition: how much is too much? The chronicle of higher education", available at: www.chronicle.com/article/ph-d-attrition-how-much-is-too-much/ (accessed 13 November 2023).

Chakraverty, D. (2020), "PhD student experiences with the impostor phenomenon in STEM", *International Journal of Doctoral Studies*, Vol. 15, pp. 159-179, doi: 10.28945/4513.

Choi, H. (2015), "An academic nomad?: identity formation of iSchool students", iConference, *Newport Beach CA USA*, available at: www.ideals.illinois.edu/items/73928

Contractor, N. (2009), "The emergence of multidimensional networks", *Journal of Computer-Mediated Communication*, Vol. 14 No. 3, pp. 743-747, doi: 10.1111/j.1083-6101.2009.01465.x.

Corson-Rikert, J., Mitchell, S., Lowe, B., Rejack, N., Ding, Y. and Guo, C. (2012), *The VIVO Ontology*, VIVO, Springer International Publishing, Cham, pp. 15-33, doi: 10.1007/978-3-031-79435-3_2.

Dickison, M.E., Magnani, M. and Rossi, L. (2016), *Multilayer Social Networks*, Cambridge University Press, Cambridge, doi: 10.1017/CBO9781139941907.

Dimara, E., Zhang, H., Tory, M. and Franconeri, S. (2022), "The unmet data visualization needs of decision makers within organizations", *IEEE Transactions on Visualization and Computer Graphics*, Vol. 28 No. 12, pp. 4101-4112, doi: 10.1109/TVCG.2021.3074023.

Fortunato, S., Bergstrom, C.T., Börner, K., Evans, J.A., Helbing, D., Milojević, S., Petersen, A.M., Radicchi, F., Sinatra, R., Uzzi, B. and Vespignani, A. (2018), "Science of science", *Science*, Vol. 359 No. 6379, p. eaao0185, doi: 10.1126/science.aao0185.

Franc-Dąbrowska, J.A., Gołębiewska, B., Gołębiewski, J. and Drejerska, N. (2020), "PhD schools as an innovative model of education at universities of life sciences", in Babić, V. and Nedelko, Z. (Eds), *Advances in Higher Education and Professional Development*, IGI Global, Hershey, Pennsylvania, pp. 342-362, doi: 10.4018/978-1-7998-2708-5.ch015.

Gardasevic, S. (2020), "User-driven efforts in creating knowledge graph information system", *Digital Library Perspectives*, Vol. 36 No. 2, pp. 97-111, doi: 10.1108/DLP-12-2019-0043.

Gardasevic, S. (2022), "CIS graph database and model [dataset]", Figshare, doi: 10.6084/m9. figshare.21663401.v4

Gardasevic, S. and Gazan, R. (2023), "Community design of a knowledge graph to support interdisciplinary PhD students", in Sserwanga, I., Goulding, A., Moulaison-Sandy, H., Du, J.T., Soares, A.L., Hessami, V. and Frank, R.D. (Eds), *Information for a Better World: Normality, Virtuality, Physicality, Inclusivity*, Springer Nature Switzerland, Cham, Vol. 13972, pp. 473-490, doi: 10.1007/978-3-031-28032-0_36.

Gardner, S.K., Szostak, R. and Repko, A. (2011), "'A jack of all trades and a master of some of them': successful students in interdisciplinary PhD programs", available at: https://our.oakland.edu/ handle/10323/4465

Grady, R.K., La Touche, R., Oslawski-Lopez, J., Powers, A. and Simacek, K. (2014), "Betwixt and between: the social position and stress experiences of graduate students", *Teaching Sociology*, Vol. 42 No. 1, pp. 5-16, doi: 10.1177/0092055X13502182.

Hong, S.R., Hullman, J. and Bertini, E. (2020), "Human factors in model interpretability: industry practices, challenges, and needs", *Proceedings of the ACM on Human-Computer Interaction, 4 (CSCW1)*, pp. 1-26, doi: 10.1145/3392878.

International Standard Organization (ISO) (2019), available at: www.iso.org/obp/ui/#iso:std:iso:9241: −11:ed-2:v1:en (accessed 13 November 2023).

Jockers, M.L. (2015), "Syuzhet: extract sentiment and plot arcs from text, github", available at: https:// github.com/mjockers/syuzhet (accessed 13 November 2023).

Jordão, V., Gonçalves, D. and Gama, S. (2014), "EduVis: visualizing educational information", *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational, presented at the NordiCHI '14: The 8th Nordic Conference on Human-Computer Interaction*, ACM, Helsinki Finland, 1011-1014, doi:10.1145/2639189.2670263.

Keppler, S.M. and Leonardi, P.M. (2023), "Building relational confidence in remote and hybrid work arrangements: novel ways to use digital technologies to foster knowledge sharing", in Baym, N. and Ellison, N. (Eds), *Journal of Computer-Mediated Communication*, Vol. 28 No. 4, p. zmad020, doi: 10.1093/jcmc/zmad020.

Kim, H., Schlicht, R., Schardt, M. and Florack, A. (2021), "The contributions of social comparison to social network site addiction", in Gruebner, O. (Ed.), *Plos One*, Vol. 16 No. 10, p. e0257795, doi: 10.1371/journal.pone.0257795.

Levecque, K., Anseel, F., De Beuckelaer, A., Van Der Heyden, J. and Gisle, L. (2017), "Work organization and mental health problems in PhD students", *Research Policy*, Vol. 46 No. 4, pp. 868-879, doi: 10.1016/j.respol.2017.02.008.

Li, H., Appleby, G., Brumar, C.D., Chang, R. and Suh, A. (2023), "Knowledge graphs in practice: characterizing their users, challenges, and visualization opportunities", available at: http://arxiv. org/abs/2304.01311

Li, Q., Baciu, G., Cao, J., Huang, X., Li, R.C., Ng, P.H.F., Dong, J., Zhang, Q., Sin, Z.P. and Wang, Y. (2022), "KCUBE: a knowledge graph university curriculum framework for student advising and career planning", in Li, R.C., Cheung, S.K.S., Ng, P.H.F., Wong, L.P. and Wang, F.L. (Eds), *Blended Learning: Engaging Students in the New Normal Era*, Springer International Publishing, Cham, Vol. 13357, pp. 358-369, doi: 10.1007/978-3-031-08939-8_31.

McGee, F., Ghoniem, M., Melançon, G., Otjacques, B. and Pinaud, B. (2019), "The state of the art in multilayer network visualization", *Computer Graphics Forum*, Vol. 38 No. 6, pp. 125-149, doi: 10.1111/cgf.13610.

Meyer, M.A. (2001), "Our own liberation: reflections on Hawaiian epistemology", *The Contemporary Pacific*, Vol. 13 No. 1, pp. 124-148.

Müller, K. (2023), "How to use Chat-GPT and python to build a knowledge graph in Neo4j based on your own articles", *Medium*, 26 August, available at: https://towardsdatascience.com/how-to-use-chat-gpt-and-python-to-build-a-knowledge-graph-in-neo4j-based-on-your-own-articles-c622bc4e2eaa (accessed 13 November 2023).

National Center for Science and Engineering Statistics (2021), "Doctorate recipients from U.S. Universities: 2020", available at: https://ncses.nsf.gov/pubs/nsf22300/ (accessed 13 November 2023).

Naughton, R., Hall, C., Zhao, H. and Lin, X. (2010), "PhD portal: developing an online iSchool doctoral student community", iConference, *Urbana-Champaign IL USA*.

Ortiz Vivar, J., Segarra, J., Villazón-Terrazas, B. and Saquicela, V. (2022), "REDI: towards knowledge graph-powered scholarly information management and research networking", *Journal of Information Science*, Vol. 48 No. 2, pp. 167-181, doi: 10.1177/0165551520944351.

Osborne, F., Motta, E. and Mulholland, P. (2013), "Exploring scholarly data with Rexplore", in Salinesi, C., Norrie, M.C. and Pastor, Ó. (Eds), *Advanced Information Systems Engineering*, Vol. 7908 Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 460-477, doi: 10.1007/978-3-642-41335-3_29.

Paulheim, H. (2016), "Knowledge graph refinement: a survey of approaches and evaluation methods", *Semantic Web*, Vol. 8 No. 3, pp. 489-508, doi: 10.3233/SW-160218.

Pinto, L.F.S. and Dos Santos, C.D. (2018), "Motivations of crowdsourcing contributors", *Innovation and Management Review*, Vol. 15 No. 1, pp. 58-72, doi: 10.1108/INMR-02-2018-004.

Porter, A.L. and Rafols, I. (2009), "Is science becoming more interdisciplinary? Measuring and mapping six research fields over time", *Scientometrics*, Vol. 81 No. 3, pp. 719-745, doi: 10.1007/s11192-008-2197-2.

Preston, J.P., Ogenchuk, M.J. and Nsiah, J.K. (2014), "Peer mentorship and transformational learning: PhD student experiences", *Canadian Journal of Higher Education*, Vol. 44 No. 1, pp. 52-68, doi: 10.47678/cjhe.v44i1.182924.

Rahdari, B., Brusilovsky, P. and Javadian Sabet, A. (2021), "Connecting students with research advisors through User-Controlled recommendation", *Presented at the RecSys '21: Fifteenth ACM Conference on Recommender Systems*, ACM, Amsterdam Netherlands, pp. 745-748, doi: 10.1145/3460231.3478879.

Saldaña, J. (2016), *The Coding Manual for Qualitative Researchers*, SAGE Publications, available at: https://study.sagepub.com/saldanacoding3e (accessed 13 November 2023).

Sarrafzadeh, B., Vechtomova, O. and Jokic, V. (2014), "Exploring knowledge graphs for exploratory search", *Proceedings of the 5th Information Interaction in Context Symposium, presented at the IIiX '14: Fifth Information Interaction in Context Symposium*, ACM, Regensburg Germany, pp. 135-144, doi: 10.1145/2637002.2637019.

Sarrafzadeh, B., Roegiest, A. and Lank, E. (2020), "Hierarchical knowledge graphs: a novel information representation for exploratory search tasks", available at: http://arxiv.org/abs/2005.01716

Sarrafzadeh, B., Vtyurina, A., Lank, E. and Vechtomova, O. (2016), "Knowledge graphs versus *hierarchies: an analysis of user behaviours and perspectives in information seeking*", Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval, presented at the CHIIR '16: Conference on Human Information Interaction and Retrieval, *ACM, Carrboro NC USA*, pp. 91-100, doi: 10.1145/2854946.2854958.

Saunders, B., Sim, J., Kingstone, T., Baker, S., Waterfield, J., Bartlam, B., Burroughs, H. and Jinks, C. (2018), "Saturation in qualitative research: exploring its conceptualization and operationalization", *Quality and Quantity*, Vol. 52 No. 4, pp. 1893-1907, doi: 10.1007/s11135-017-0574-8.

Talja, S. (2002), "Information sharing in academic communities: types and levels of collaboration in information seeking and use", *New Review of Information Behaviour Research*, Vol. 3, pp. 143-160.

Twidale, M.B., Nichols, D.M. and Paice, C.D. (1997), "Browsing is a collaborative process", *Information Processing and Management*, Vol. 33 No. 6, pp. 761-783, doi: 10.1016/S0306-4573(97)00040-X.

Wiggins, A. and Sawyer, S. (2012), "Intellectual diversity and the faculty composition of iSchools", *Journal of the American Society for Information Science and Technology*, Vol. 63 No. 1, pp. 8-21, doi: 10.1002/asi.21619.

Wood, C.V., Campbell, P.B. and McGee, R. (2016), "An incredibly steep hill': how gender, race, and class shape perspectives on academic careers among beginning biomedical PhD students", *Journal of Women and Minorities in Science and Engineering*, Vol. 22 No. 2, pp. 159-181, doi: 10.1615/ JWomenMinorScienEng.2016014000.

Woolston, C. (2019), "PhDs: the tortuous truth", *Nature*, Vol. 575 No. 7782, pp. 403-406, doi: 10.1038/ d41586-019-03459-7.

Young, S.N., Vanwye, W.R., Schafer, M.A., Robertson, T.A. and Poore, A.V. (2019), "Factors affecting PhD student success", *International Journal of Exercise Science*, Vol. 12 No. 1, pp. 34-45, available at: www.ncbi.nlm.nih.gov/pmc/articles/PMC6355122/

Further reading

Jockers, M.L. (2023), "Syuzhet: extract sentiment and plot arcs from text", available at: https://github. com/mjockers/syuzhet (accessed 13 November 2023).

Appendix 1. Voting results on 20 queries on CIS knowledge graph3

Queries	Student progress in PhD program		
	Avg_New	*Avg_Mid*	*Avg_ABD*
Student pathways			
Q1 How long did it take students to publish a paper?	*4.75*	*5*	*4*
Q2 Students' timeline – to propose and defend	*4.75*	*4.5*	*4.5*
Q3 Venues where students published papers	*4.25*	*4.5*	*5*
Faculty and their dissertation involvements			
Q4 Number of dissertations all faculty chaired	*3.5*	*4*	*4*
Q5 Number of dissertations faculty were involved in	*4.25*	*4*	*3*
Q6 Number of dissertations Professor X chaired on Area Y	*4*	*3.5*	*5*
Q7 Who are university representatives on dissertations?	*3.25*	*3.5*	*4*
Q8 Dissertations faculty chaired on research area/research topic/domain of application	*3.75*	*4*	*4*
Q9 Faculty dissertation collaborators of Prof X	*3.75*	*4*	*4*
Q10 Current students of Professor X	*4.75*	*5*	*4.5*
Q11 Dissertation committee involvement of Prof X	*4.75*	*5*	*5*
Q12 Dissertation collaborators of Prof X on topic Y	*4.25*	*4*	*3.5*
Courses/exams info			
Q13 Comments of people who took directed readings with Prof X	*4.75*	*4*	*4*
Q14 Students who took course/Exam X and their comments	*4.75*	*4*	*4*
Q15 Courses teaching Methodology X, and students who took them	*4.75*	*5*	*4*
Faculty activities (other than dissertation)			
Q16 Co-authorship network within CIS	*4*	*3.5*	*4.5*
Q17 People who share research topics with Professor X	*4.25*	*3.5*	*4*
Q18 People and publications on research area/Topic X	*4.5*	*4.5*	*5*
Q19 Researchers from the different departments who look at the domain of Application X	*4*	*4.5*	*4.5*
Q20 Expand the previous query – show methodologies/tools they are using	*3.25*	*4.5*	*5*

Source: Table created by authors

The table shows the voting results on the queries (on the five-point Likert scale where five is "most useful"), averaged for *new*, *mid* (second year) and *ABD* students who participated in the evaluation study. The queries are grouped into categories (e.g. *Student pathways*). All of these Cypher queries are available in the published data set, together with the layout for visualizing them.
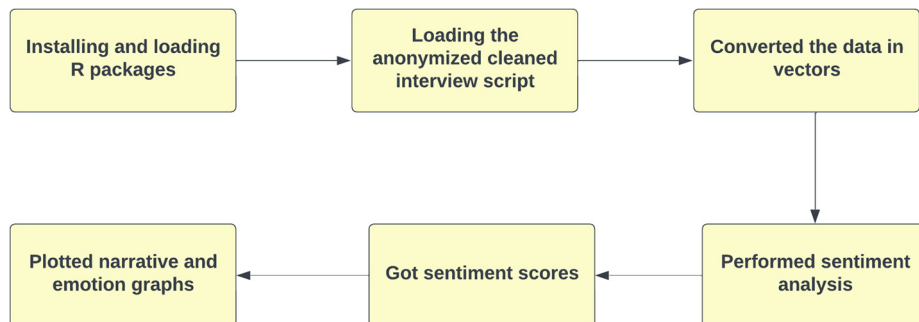
Appendix 2

Source: Figure created by authors

Corresponding author
Stanislava Gardasevic can be contacted at: gardasev@hawaii.edu