Community Design of a Knowledge Graph to Support Interdisciplinary PhD Students

Stanislava Gardasevic(✉) and Rich Gazan

University of Hawaiʻi at Mānoa, Honolulu, HI 96826, USA
{gardasev,gazan}@hawaii.edu

Abstract. How do PhD students discover the resources and relationships conducive to satisfaction and success in their degree programs? This study proposes a community-grounded, extensible knowledge graph to make explicit and tacit information intuitively discoverable, by capturing and visualizing relationships between people based on their activities and relations to information resources in a particular domain. Students in an interdisciplinary PhD program were engaged through three workshops to provide insights into the dynamics of interactions with others and relevant data categories to be included in the graph data model. Based on these insights we propose a model, serving as a testbed for exploring multiplex graph visualizations and a potential basis of the information system to facilitate information discovery and decision-making. We discovered that some of the tacit knowledge can be explicitly encoded, while the rest of it must stay within the community. The graph-based visualization of the social and knowledge networks can serve as a pointer toward the people having the relevant information, one can reach out to, online or in person.

Keywords: PhD students · Tacit knowledge · Knowledge graph · Data model · Interdisciplinary research

1 Introduction

Pursuing a PhD degree can be both rewarding and stressful. A significant percentage of PhD students are experiencing at least two psychological symptoms and have sought help for anxiety or depression related to their studies [1, 2]. One key factor influencing the overall experience of PhD students during their course of study is the so-called *departmental culture*, encompassing "student/faculty relationships, student involvement in academic life, student satisfaction with programs, student-to-student interactions, institutional financial assistance to students, and dissertation factors" [3]. The issue of navigating departmental cultures becomes even more problematic in interdisciplinary PhD programs, where students reported "feeling disconnected from faculty and peers, having to span boundaries between areas, departments, and knowledge bases" [4]. In this study, we describe the participative development of a knowledge base to support interdisciplinary PhD students to visualize and navigate multiple departmental cultures.

I. Sserwanga et al. (Eds.): iConference 2023, LNCS 13972, pp. 473–490, 2023.
https://doi.org/10.1007/978-3-031-28032-0_36

Science can be described as a "complex, self-organizing, and evolving network of scholars, projects, papers and ideas" [5]. As Börner [6] noted, researchers and authors can be perceived as nodes in networks of support and influence, while their selection of research topics, collaborators, students, and publication venues influence and shape their place in these networks. In these 'knowledge networks,' actors serve both as keepers of knowledge and as agents that seek out, communicate, and create knowledge [7, 8]. However, information is often not equally available to all actors [4], and the opportunities for "social" information sharing [9] diminished due to the pandemic.

Consequently, we chose a multilayer/multiplex network as the medium to address the raised issues and create a knowledge graph combining information on social and knowledge traces in a community. The aim of building this graph is to facilitate explicit and tacit information discovery and informed decision-making related to the factors identified as meaningful for this population: i) establishing collaboration with faculty [3, 4, 10], ii) information and resources available about research topic of interest [3, 7, 11], and iii) interaction with peers, including community building [3, 4]; while creating a testbed for exploring multiplex network visualization and navigation options [12].

The setting for this case study is the Interdisciplinary PhD Program in Communication and Information Sciences (CIS), at the University of Hawai'i at Mānoa. Students—especially those new to the program who have not been exposed to informal information flows—are considered the main user population of this information discovery system. Following the activity theoretical approach and requirements/milestones for the program, students were engaged in every step of the knowledge graph modeling process.

As noted by Hogan et al. [13], the definition of a knowledge graph remains contentious, but in short, it can be described as "a graph of data intended to accumulate and convey knowledge of the real world, whose nodes represent entities of interest and whose edges represent relations between these entities". Before building the graph, we created the database model to support tacit knowledge exchange-such as first-hand impressions of other students-that have been shown to be of crucial importance for this population [7, 9, 10].

Getting procedural advice or hearing other students' experiences with faculty, classes, and exams usually happens in a serendipitous way, via in-person conversations, but these opportunities for informal information exchange were severely disrupted by COVID-19. Therefore, we propose this data-driven discovery tool that combines topical research representations and tacit interpersonal relations data to identify and support impactful interdisciplinary collaborations [5] and multiple departmental cultures [7] common to iSchools.

The main contribution of this paper is the graph database model developed and presented alongside user-derived needs. We outline some of the data categories and relationships between them (including refining attributes) that can be used to encode both explicit and tacit knowledge, while the affordance of network visualization supports the discovery of people with pertinent information that is not encoded in the knowledge graph.

2 Related Work

The aim of this section is to connect the present study to previous research streams relevant to the challenges of tacit knowledge transfer, representing and supporting interdisciplinary collaborations, and to summarize and justify the affordances of a knowledge graph approach.

2.1 Tacit Knowledge Exchange for PhD Students

Personal networks and peer communities present a vital information source for doctoral students [3, 7]. So-called "insider" knowledge is something perceived as necessary for the success of interdisciplinary PhD students but is available only to students and professors who have been involved in the program for a certain amount of time [4, 7]. This knowledge is often referred to as experiential or tacit knowledge, here operationalized as the values and quality of the resource of interest, as perceived by students [14]. Resources that help students discover appropriate faculty to collaborate with are especially critical for this population [3, 4, 10], as well as information about research topics of interest [3, 7, 11].

Visualizing social networks of people in a shared academic setting could support "social" sharing [9] by providing a better picture of the common links, potentially democratizing access to useful information currently available only to those conveniently located in the social network.

2.2 User Needs for iSchools and Interdisciplinary PhD Programs

The need for a community portal to support students from diverse disciplines has been identified in previous studies as a potentially useful forum for peer information exchange [15], but it remains an elusive goal. The challenge multiplies when interdisciplinary PhD students must select appropriate faculty advisors from a rich pool of multi-disciplinary researchers, which has been a longstanding success factor in PhD student persistence [3, 16].

Choi focuses on the importance of identity formation within iSchools, where students can question dominant research trends, locate their own interests and develop their own research identity [17]. Wiggins and Sawyer propose an analysis of interdisciplinary faculty research output to outline the landscape within which iSchool students might locate their work [18].

Each of these studies distinguishes between research and community relationships but tends to focus on research topics as the basis for connection rather than social relationships. We propose a lightweight, extensible visualization, driven by student focus groups, to capture and represent community interactions and tacit information alongside research connections, so students can view paths and interactions of those who have come before.

2.3 Network-Based Representation and Discovery

Faculty profile pages can be a lightweight data source of potential connections between researchers and students from diverse fields, and community stakeholders [19], and can provide a way to identify one's place in the interdisciplinary community. However, in the case of the program studied here, students reported that faculty profiles are often not up to date, and too numerous to inspect thoroughly [10]. Also, expert finding applications in the scholarly domain are usually based on recommending acclaimed researchers based on their publication history (i.e. [20, 21]). These attempts have traditionally been based on the networks generated from publication data obtained from a single database such as Computer Science Bibliography – DBLP [21–23], Web of Science, Scopus, and PubMed [24]. However, this research, based on a convenience sample, is often addressing only a limited number of relationships, usually pertaining to co-authorship and shared research topics. Our approach aggregates different web resources (including multiple sources indicating researchers' expertise) in a single knowledge graph-since research has shown that the graph-based interface is more practical for finding specific information and simple question-answering tasks, compared to hierarchically organized information [25]. Furthermore, we expand the scope of relationships to represent the interdisciplinary domain, as three dimensions are considered as a minimum to understand the full complexity of social structures [26].

We model interdisciplinary PhD program information as a multiplex network, defined as "networks where the same set of nodes is represented in every layer, although the interaction between nodes might be different in each one. As an example, two nodes might be connected in one layer and might not in other" [27]. Analyzing and visualizing multiplex graphs are considered complex problems, as each additional relation makes the choice of an appropriate layout more challenging, even incomprehensible, as soon as it contains a few dozen nodes [28]. One previous attempt to visualize scholarly domain multiplex networks included a dataset of 61 nodes connected over five layers (work, leisure, coauthor, lunch, and Facebook) [28]. Unlike in this exploratory approach, we are grounding the graph in user needs, answering the necessity to re-frame user needs and data as multilayered networks problems, providing visualization researchers more exposure to the application domain [12].

3 Case Introduction

The CIS PhD Program was established in 1986 and has approximately 30 current students and over 100 alumni. The program is a voluntary and collaborative effort of over 40 faculty from four units housed in three colleges (as of Fall 2022):

- Communications (COM) (College of Social Sciences)
- Library and Information Science (LIS) (College of Social Sciences)
- Information and Computer Sciences (ICS) (College of Natural Sciences)
- Information Technology Management (ITM) (Shidler College of Business).

While the flexible, decentralized structure of the program provides more possible research avenues for PhD students, it also requires students to find and navigate their own path through program requirements [4, 10]. This is especially acute since the selection of a dissertation chair and other faculty mentors is the crucial factor contributing to PhD student retention and satisfaction in general [3, 31]. In addition, CIS students and alumni also serve as valuable information resources, sharing their firsthand experiences and wisdom regarding potential faculty collaborators and the practical mechanics of how to meet program requirements. However, an increasing number of students are not physically present at the university (so-called "off-island" students) and participate in joint activities such as the required weekly interdisciplinary seminar online via Zoom. This part of the population does not have the same opportunity to be part of informal interactions and information flows and gain so-called "insider" knowledge contributing to the success of interdisciplinary PhD students [4], further justifying the need for the remotely accessible technology presented here. These concerns are even more urgent considering the COVID-19 pandemic when all coursework moved online, and at this writing, the two most recent cohorts of students have yet to meet their peers or faculty in person.

Official CIS program milestones and requirements are inscribed in a Policies and Procedures document[1], which formed the basis for developing research instruments, as well as for creating the first iteration of the graph database model, later modified based on CIS student input from the workshops.

The research presented in this paper is the second stage of a larger project. The results of the first stage suggested that information resources and relevant data categories can be formalized and encoded in a form of a knowledge graph and suggested a path to capture/obtain tacit information.

The present study builds on the first, and explores the research question:

> **What tacit knowledge can be represented and discovered through a knowledge graph, and what can be only indicated?**

The final stage of this project will involve the same population in the participatory design of visualizations and seamless visual analytical approach of the multiplex graph, to examine if this artifact has utility compared to current means of information discovery, and if it is worthwhile to develop it further into an information system.

4 Method

Following a Human-Centered Design approach, upon the requirements gathering process that consisted of interviews and a website usability study engaging program students and alumni [10], we conducted three workshops over two years and developed the database model shown in Fig. 1. Students helped identify relevant people, roles, resources, actions, experiences, and relationships, and these were iteratively integrated into the database model, shared back in successive workshops and refined based on feedback. Including the community in the design of a technology to support their interactions helps reveal and unravel the underlying values of both the proposed system and the PhD program itself [29], creating a rationale and platform for further cooperation and community building.

[1] CIS Program policies and procedures- https://bit.ly/cis_policy_procedures.

4.1 Positionality and Limitations

Both authors are insiders to the CIS program, with different roles and perspectives, and with access to different kinds of knowledge about the community, which we feel provides a productive tension.

The first author is a CIS PhD candidate and must balance her own experience with those of other students with whom she is engaging in this study, including informal daily interactions. The limitation imposed by this position as a student is that graph design decisions and interpretations may be biased by personal experience. We have attempted to balance this threat to internal validity by including member checks as validation points to incorporate multiple students' perspectives.

The second author is a faculty member and former CIS Chair who teaches one of the three core courses and serves on multiple dissertation, exam, and other program committees. His perspective yields different stories about enablers, barriers, and metrics of student experience and degree progress. This limitation is that students could be understandably hesitant to share their unfiltered observations with someone in a position to evaluate them, due to courtesy bias or other factors. We address this limitation by emphasizing the anonymization component of the data analysis process and creating faculty-free workshops and spaces for data collection.

4.2 Data Collection

This research is framed by *activity theory*, as this approach has been used to inform the ways in which interactive tools should be designed to make a positive impact on human activities [30]. In this case, the overarching goal (activity) of a student is to obtain a PhD degree; to get there, students are motivated to fulfill written program requirements (actions), such as taking certain *courses* and *exams,* but also unwritten requirements such as developing and successfully navigating relationships with faculty and peers. Each step in the research design was informed by activity theory, encompassing formal program requirements and other practical and less tangible aspects of the interdisciplinary PhD student experience to help guide future students.

In the three workshops, current CIS students shared their experiences in a forum environment and reacted to the evolving graph design. Student participation was voluntary, and data was collected without faculty present and analyzed with full confidentiality and anonymity. For that reason, only the first author facilitated workshops and anonymized the raw data, then shared it with the second author in a form that honors the privacy of the participants. All workshops used a visual representation of a graph database model in a then-current version, to communicate the data modeling efforts thus far, to serve as a community member check on the "ontology" of the shared domain, and to identify areas for improvement.

The three workshops took place from December 2019 to April 2022. The first included both face-to-face and online participants, while the other two were fully online via Zoom, and included 18, 15, and 10 students, respectively. The workshops were recorded, transcribed and anonymized by the first author, while the rest of the data was captured as text files created by the participants' chat, and Google documents where students were invited to leave anonymous comments on the topics discussed.

During the first workshop, participants completed a questionnaire[2] designed to validate the findings from the previous study stage, while the poll results were used as a prompt for discussion during the workshop.

The aim of the second workshop was to gather feedback on the draft data collection form, designed with the intention to populate parts of the graph with students' personal information, based on data categories they perceived as relevant. The intent was to hone the questions and structure of the form to get the most useful information when collecting data in the future, negotiate data privacy boundaries, and get feedback and further recommendations on what data would be most relevant for this population. This workshop resulted in an elaborate data collection form[3] created in Qualtrics that was used to collect the data from the students to populate the graph. This form was designed for collecting user-generated information via pre-determined categories, as well as the free text inputs, while all the data collected corresponds with the graph model.

The third workshop invited open discussion on the students' experiences, where participants were distributed in breakout rooms depending on i) program progress stage and ii) preferred methodological approach[4], to discuss and exchange information about what helped and hampered their progress.

4.3 Data Analysis

The data consist of anonymized i) transcription of workshops, ii) Google documents, iii) researcher notes, and iv) pertinent questionnaire/form inputs. The data analysis was conducted iteratively. After each workshop, both authors conducted the first cycle of descriptive coding of student responses, and the categories that emerged in all three workshops were pertinent to 'flow of information in the community', 'lack of opportunities for in-person contact with peers', and 'obtaining information on program requirements'. Upon the discussion about results from this phase, researchers created the utility-driven set of codes, to conduct the structural data coding [31] with the previously elicited and potential new concepts that could be represented via a graph model. Upon the second round of coding, we discussed revisions to the data model based on the data analysis insights as we outline in several examples below. To perform the participant' checks, each workshop started with an overview of the graph model in its current state, to gather further feedback and reflect on the accuracy of data representation for their information needs. Participatory research relies on participant engagement with the data collection and coding processes, which helps create a sense of ownership of the data reflecting their community, a "safe space" where subjects can interact and reflect, and an understanding of how their actions and interactions are represented within the graph [32].

[2] Questionnaire with results- https://bit.ly/cis_worskhop_1.

[3] Data collection form - https://bit.ly/student_data_collection_cis.

[4] Breakout room prompts for Workshop 3 https://bit.ly/cis_workshop_3.

5 Results: User-Driven Graph Modeling Decisions

In this section, we discuss the themes that emerged as important for the community during the workshops, and how they influenced the graph modeling and design decisions. In Fig. 1, we presented the CIS domain graph database model in its current iteration and the nodes and relationships between them. The model or its parts can be adapted and reused in other interdisciplinary environments.

Since we use the Neo4J graph database for building this knowledge graph, we utilize the native Cypher language syntax when referring to specific parts of the model, i.e.: (Node)—[RELATIONSHIP]—> (Node); in this case, (Node) represents a class/category in the model, and not pertinent instances.

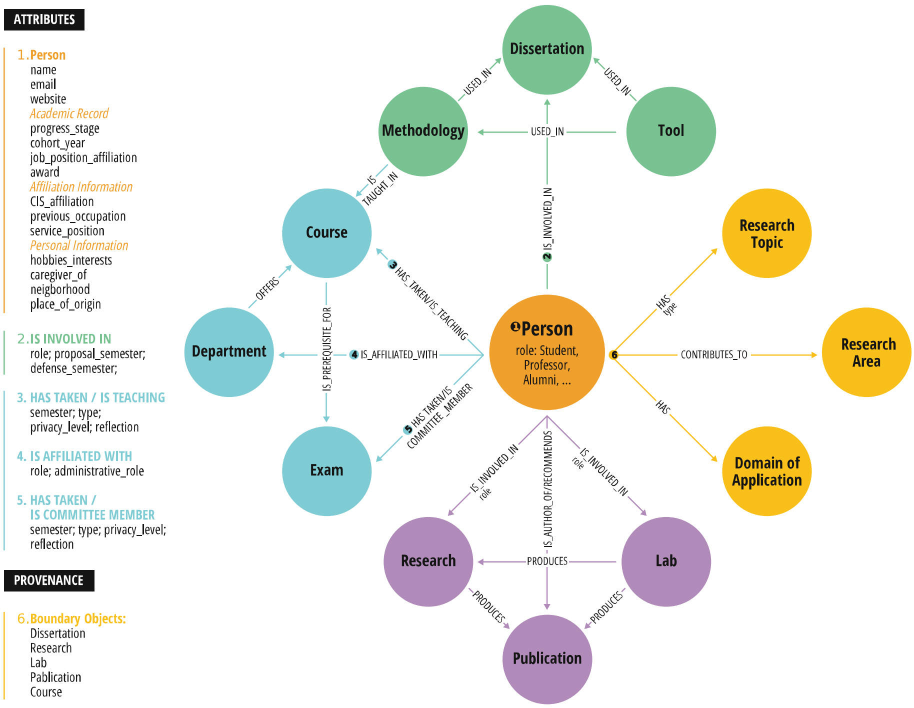

Fig. 1. CIS graph database model (Color figure online)

Considering our aim to facilitate the active exchange of tacit information, (Person) occupies the central place in the model and is connected to most of the surrounding 12 nodes/classes, presented through 4 color-coded groupings. Those nodes present the affiliations (layers) via which the actors in this multiplex network may be connected. The data comprising the graph was gathered from different web locations, normalized and ingested into the Neo4J, and is available for download and reuse[5], while the student-generated data is not included in this dataset. The graph created according to this model

[5] The data 'dump' of CIS Neo4J database-https://doi.org/10.6084/m9.figshare.21663401.v2.

represents a picture of a domain, as it is captured in Spring 2022, while the publications in the corpus represent the sample of publishing activity of faculty in the last 10 years. We acknowledge the challenge of maintaining updated data, but we offer this model as a useful snapshot to elicit student reaction and engagement.

Following the activity theory approach, the categories that were initially included by default in the graph were the CIS degree requirements that every student must fulfill to advance to ABD (all but dissertation) status. Those categories are (Course), (Exam), (Dissertation), and (Publication).

5.1 (Person) Node and Its Attributes (Fig. 1, Central Orange Node)

The (Person) node, naturally, has the most attributes, as it is designed to capture information primarily about other students and outline some interesting, relevant, or potentially shared interests and experiences.

There are three subsections of attributes; *Academic Record* attributes representing information such as *cohort_year*, indicating the strength of ties with others that share the same value attribute since participants stated that taking classes together with cohort peers makes them closer, as they can talk about challenges and struggles (noted as pre-COVID practice). In the *Affiliation Information* subsection, we capture the CIS affiliation, with a Boolean value, to distinguish CIS-affiliated people from others (such as co-authors or faculty that left the university). The *Personal Information* section contains categories that capture free-text information that may serve as potential connection points, such as shared hobbies or neighborhoods. This was intended to support community-building, since students reported feeling lonely, especially those at the dissertation research and writing stage and due to pandemic isolation [4, 10].

Another attribute that evolved based on participant suggestions was *caregiver_of*. Initially, the category was named *family_status*, where students could share data such as the ages of their minor children, to potentially connect for play dates. Participants broadened our conception of this category to include other senses of caregiving, which also provided deeper insights into the range of students' life situations:

> *For the number of 'minor children'…some of us have adult children. Also consider changing this to 'are you a caregiver' as some need to care for elderly parents, etc.*

5.2 (Dissertation), (Methodology), (Tool) Nodes (Fig. 1, Top Three Green Nodes)

As mentioned, (Dissertation) is crucial for this population, therefore separated from the (Research) as a category and forms the basis of one of the most important networks—the dissertation mentorship network.

In this program, there are over 40 faculty members from which students can choose i) a dissertation chair, ii) three dissertation committee members (at least two of which must be affiliated with different units (Department) to help ensure an interdisciplinary research approach), and iii) an "external" member, which further increases the disciplinary range of potential collaborators. Based on the inputs from participants, asking to see what faculty served as "external" members in previous committees, we modeled the relationship by adding the *role* of faculty involved in the previous dissertations. Upon applying social network analysis and centrality metric, visualizing this network can quickly show the most active dissertation chairs and committee members; seeing their previous collaborations might help students decide who to invite to serve on their committee. In Fig. 2 we demonstrate a perspective of the dissertation mentorship network, where faculty node colors represent their departmental affiliations (LIS-purple, COM-orange, ITM-yellow, ICS-blue); green nodes are dissertations; while the number of dissertation engagements is shown in node size, and dissertation chairing via red edges.

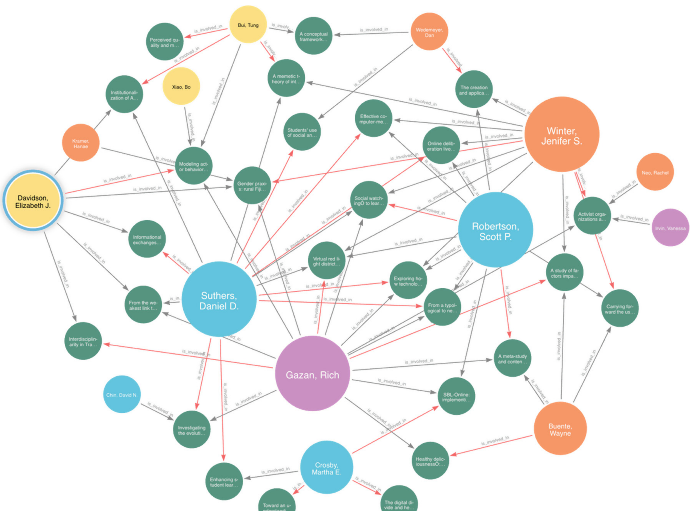

Fig. 2. Dissertation-mentorship network showing collaboration between faculty from different departments. (Color figure online)

Considering the importance of the relationship with the supervisor for the overall PhD experience [3, 31], one workshop participant wrote succinctly and directly about the information they seek from other students:

I would ask them-is your supervisor a good mentor?

This statement demonstrates that some of the most-sought information is not easily encodable, but the graph supports the option to indicate a person that can share it with you, directly.

The (Methodology) node reflects the program requirement of taking a research methods course, and the common but informal practice of including a methodology expert in the dissertation committee. With the information encoded this way, students can get i) a quick overview of faculty who use a certain methodological approach in their research (Fig. 3); and ii) discover appropriate methodology classes based on the comments of other CIS students-a requirement expressed during the first workshop.

The (Methodology) node relates to the (Tool) node, as students are interested in finding out methodology classes that use particular tools. This information was lacking from the current information space and was indicated by study participants at the third workshop as much needed. Students who populated the data collection form indicated tools they used in classes they attended. Furthermore, the graph also contains data on the people who used a tool to perform research, so a student can identify and directly consult that person on the issue.

5.3 (Course), (Department), (Exam) Nodes (Fig. 1, Left Three Blue Nodes)

On the left side of the model, we formalized the departmental affiliation of people, and the attribute named *role* is intended to capture data that would help distinguish/filter actors based on their affiliation to different departments ("Student", "Professor", "Alumni", "TA", etc.). This modeling decision helps students see the interdepartmental collaborations (interdisciplinarity indicator), and current students' graduate assistantship affiliations (insider knowledge indicator).

When it comes to encoding tacit knowledge, some community members were understandably hesitant to leave written traces of their experience with a professor, as opposed to their experience regarding courses and exams taken. During the first workshop, more students expressed support for the latter, both via the questionnaire and discussion. A comment illustrating this point is:

Social rating is toxic.

As a result, we included the space for *reflection* as an attribute of the relationship (Person)—[HAS_TAKEN]—>(Course/Exam), where students can note their experiences if they feel they should share them with their peers, allowing for tacit knowledge encoding and sharing. This way, students can comment on professors if they wish, albeit indirectly. To manage the privacy settings, upon discussion with students, we allowed three options of the public for the reflection *attribute* value: "public," "CIS students", and "private". For those who indicated their observations as "private", we anonymized the data by connecting it to an *Anonymous* person instance-so the comments are still available to other CIS students, without clear implications of who might have left them. Via graph visualization, students will be able to see others who took a course/exam of their interest and reach out to them for a private discussion, if sufficient reflection input is missing.

Some of the reflections left by students for courses are:

This class was meant to prepare us for [a qualifying] exam, but it did not.

Way too much reading.

I enjoyed this class, it helped me write dissertation proposal- without this class I wouldn't be able to write it that fast. It gives you the basic structure and what you can expect from the journey. And I liked the professor- she's very encouraging and inspirational.

5.4 (Research Topic), (Research Area), (Domain of Application) Nodes (Fig. 1, Right Three Yellow Nodes)

Considering the case of an interdisciplinary PhD program, facilitating communication across disciplines [33] is a core aim of this work. The three yellow nodes on the right side of the model, namely (Research Topic), (Research Area), and (Domain of Application) are intended to host vocabularies that will serve as boundary objects, which "inhabit several communities of practice and satisfy the informational requirements of each of them" [34]. With these concepts in mind, we have chosen controlled vocabularies for pertinent nodes that would be understandable across the four constituent disciplinary areas of CIS. This contribution helps bridge gaps in cross-disciplinary understanding among diverse units that have been identified in prior studies (e.g. [35, 36]). As one of the participants stated:

> *The program's strength — the cross-discipline nature — is also its weakness. It would be interesting to find the "e Pluribus Unum" that makes the "one from the many," some set of unifying principles to rally around.*

Our approach was to involve people educated in the different disciplines encompassed by the program, to serve as translators or mediators, and create vocabularies to improve cross-disciplinary communication [33].

In the first version of the modeling effort, *Discipline* was not included as a separate node in the graph, since it was potentially repetitive with the (Department), and we did not wish to reify disciplinary divisions. However, the workshops again yielded valuable insights from community members, and we introduced a (Research Area) as a node in response to comments in line with:

> *I think because there is interdisciplinarity, people tend to gravitate to others who are doing a similar type of research.*

Research Area is a loosely defined term, often used to refer to discipline-like structures, yet targeting smaller units that often span disciplinary borders [37]. This term is used throughout the CIS website to describe areas of focus, such as Human-Computer Interaction, Data Science, and Health Informatics, providing a warrant from the community to include the term and node in the model.

We operationalize Research Topic entries as uncontrolled, free-response words and phrases that allow students and researchers to describe their research interests in their own words. Finally, we operationalize the Domain of Application as the particular setting, technology, and/or community with which their research engages.

For the sake of the readability of the model, links from other nodes were omitted but are originating from (Publication), (Dissertation), (Research), (Lab), and (Course) instances. The controlled vocabularies were used to manually index (Publication) and (Dissertation) instances and are subsets of the Australian and New Zealand Standard Research Classification (ANZSRC). Pertinent classification systems were complemented by domain-specific inputs. The same corpus was automatically indexed by Computer Science Ontology [38].

Students consider information about the research nuances of CIS faculty particularly valuable, and a participant in the third workshop expressed their issue as follows:

I talked to [CIS chair] and I was like, who else studies [Topic X]? And l they were like "Oh, well you just have to go through everyone's CV." And I'm like,

"I don't have time to go through how many different CVs to find out potentially someone who might be able to help me!" And maybe it's just because I'm introverted too. But I don't know how to start those conversations with professors like, "What do you study? Can you help me?"

Indexing the work of faculty and students with familiar terms would allow not only for quick discovery of potential collaborators, but also provide potential inputs for conversation starters for students to approach faculty, as they reported a lack of opportunities to communicate with them.

While the boundary object nodes may appear to be similar categories with subtle distinctions, within this interdisciplinary environment researchers often have the same Domain of Application, but different lenses for conceptualizing and studying it. For example, "Cybersecurity" might encompass various disciplinary topics such as algorithms, system architecture, policy, or ethics.

Potentially, these nodes could be entry points to which graphs from other departments and programs could be "attached", to help identify shared topics of interest across the university system.

5.5 (Research), (Lab), (Publication) Nodes (Fig. 1, Bottom Three Purple Nodes)

The nodes on the bottom of the model are yet another three categories that indicate valuable affiliation information, allowing for visualization of collaboration networks. The current graph has a corpus of 260 publications that served as a basis for the co-authorship network as well as input for populating the boundary object nodes.

6 Discussion

In this paper, we have proposed a graph model to support interdisciplinary Ph.D. students in their academic journey by encoding both explicit and tacit/experiential knowledge to make 'insider' information crucial for this population [4, 7] available to newcomers. Upon demonstrating the current knowledge graph and querying options, the majority of participants in the third workshop had positive attitudes toward the perspective of this technology being implemented, emphasizing its usefulness for newly admitted students.

We therefore suggest that the approach of engaging community members to identify and help aggregate into a knowledge graph both formal and informal information they find relevant to success in the domain is a promising avenue for other groups and knowledge management applications. Some of the comments on the utility of having data from multiple web locations aggregated in the knowledge graph versus the online search were:

"*Direct, in one platform*" and "*More easily accessible information"*

6.1 Transfer of Tacit Knowledge

As mentioned, one of the most important factors in PhD student retention is related to their satisfaction with their research supervisor and faculty collaborations [3, 31]. Even though some of the students expressed their hesitance to leave written traces of their experiences with supervisors, the data (encoded in the *reflection* attribute) gathered via the form captured the experience of students who took directed readings courses with two faculty members, clearly indicating the differences in their styles of supervision:

E.g. 1 - Very helpful, because [they] give you exactly what to do, specific advice-tools, websites, etc.

E.g., 2 - They are really hands-off, which can be a problem if you want direction.

The student-provided data support the decision-making process, and upon inspecting *reflections* of students on different exams they have taken, the latest cohort student was able to make a more informed decision when choosing a focus area exam:

Reading through what you've collected from students was helpful. I'm all for data-driven decisions:).

At the same time, the graph-based technology allows for recording the collaborations of current students with faculty, and better possibility to exchange tacit knowledge and student reflections upon establishing relationships with peers. However, for the proposed approach to be successfully implemented, some of the motivators (e.g. self-efficacy, self-enjoyment, reciprocity, and rewards) [39] need to be enacted, within the system or the program policy, to entice the continuity in knowledge sharing practices among peers.

Finally, the format of workshops to discuss these issues was also proven to make a difference as the community building effort, considering they reported lack of opportunities to exchange information with peers [10]. Participating students labeled the experience as "insightful" and "enlightening".

6.2 Affordances of Graph Modeled Approach

The graph supports both specific curiosity (e.g. Fig. 3) as well as exploratory search [40] this user population considers useful.

The current iteration of the created knowledge graph has 5,365 node instances (about 122 are CIS-affiliated people) and 6,116 relationships among them, making it a small-scale but rich-in-context dataset, created to explore visualization and navigation options. This dataset has double the layers and sample size of the graph created by [28] and the model was designed by researchers immersed in the domain created with the end users' involvement - all of which are considered necessary to tackle the complex problem of multiplex visualizations [12] - issue that will be addressed in the future participatory design workshops. We provided this dataset to other researchers who are interested in examining the stated problem.

Aside from serving the PhD student population, this research aims to tackle challenges encountered by iSchools faculty involved in interdisciplinary research [36] with a domain analytical angle to build the vocabularies and capture disciplinary norms via

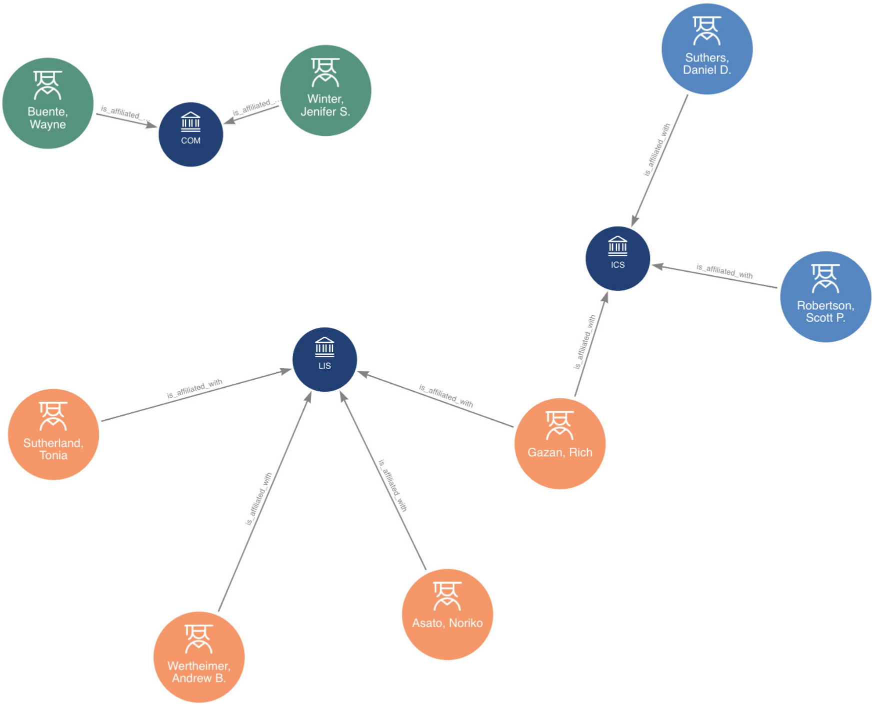

Fig. 3. Example query showing faculty from different departments conducting qualitative research

'boundary object' nodes, and possibly facilitate the collaboration of researchers with different disciplinary traditions looking at the same domain of application.

7 Conclusion

Anyone who remembers their first days in a graduate program can appreciate the inadequacy of relying solely on formal program documentation as a guide to the overall experience. By engaging students through iterative workshops, we propose a network-based visualization and discovery tool that integrates topical research data, degree requirements, and the tacit knowledge by which a program's culture–and its members' lived experiences–are communicated.

With it, students can capture their progress, plan/project future steps and collaborations, and see themselves as part of the community, subcommunity, and/or an invisible college, while providing valuable information and tacit knowledge traces for the other students.

This approach emphasizes an active role for both information seekers and sharers, by putting the people within the community in the center of the information retrieval loop, to allow a direct exchange of trusted tacit information. We discovered that some of the tacit knowledge can be explicitly encoded, while most of it must stay within the community. The graph-based visualization of the social and knowledge network can serve as a pointer toward the people having the relevant information, one can reach out to, online or in person.

Preserving and aggregating these traces via a knowledge graph and presenting them as a contrast with the common model of a generic student progressing through generic degree milestones, can yield a valuable visualization of the traces of departmental culture, and provide a more useful, accurate, and compassionate roadmap through an interdisciplinary graduate program.

Acknowledgment. We would like to wholeheartedly thank all CIS students who participated in this research.

References

1. Levecque, K., Anseel, F., De Beuckelaer, A., Van der Heyden, J., Gisle, L.: Work organization and mental health problems in PhD students. Res. Policy **46**, 868–879 (2017). https://doi.org/10.1016/j.respol.2017.02.008
2. Woolston, C.: PhDs: the tortuous truth. Nature **575**, 403–406 (2019). https://doi.org/10.1038/d41586-019-03459-7
3. Bair, C.R., Haworth, J.G.: Doctoral student attrition and persistence: a meta-synthesis of research. In: Smart, J.C. (ed.) Higher Education: Handbook of Theory and Research, pp. 481–534. Kluwer Academic Publishers, Dordrecht (2005)
4. Gardner, S.K.: 'A Jack-of-All-Trades and a Master of Some of Them': successful students in interdisciplinary PhD programs. Issues Integr. Stud. **29**, 84–117 (2011). http://hdl.handle.net/10323/4465
5. Fortunato, S.: Science of science. Science. **359**, 6379 (2018). https://doi.org/10.1126/science.aao0185
6. Börner, K.: Atlas of Science: Visualizing What We Know. MIT Press, Cambridge (2010)
7. Boden, D., Borrego, M., Newswander, L.K.: Student socialization in interdisciplinary doctoral education. High Educ. **62**, 741–755 (2011). https://doi.org/10.1007/s10734-011-9415-1
8. Phelps, C., Heidl, R., Wadhwa, A.: Knowledge, networks, and knowledge networks: a review and research agenda. J. Manag. **38**, 1115–1166 (2012). https://doi.org/10.1177/0149206311432640
9. Talja, S.: Information sharing in academic communities: types and levels of collaboration in information seeking and use. New Rev. Inf. Behav. Res. **3**, 143–160 (2002)
10. Gardasevic, S.: User-driven efforts in creating knowledge graph information system. Digit. Libr. Perspect. **36**(2), 97–111 (2020). https://doi.org/10.1108/DLP-12-2019-0043
11. Catalano, A.: Patterns of graduate students' information seeking behavior: a meta-synthesis of the literature. J. Documentation **69**, 243–274 (2013). https://doi.org/10.1108/00220411311300066
12. McGee, F., Ghoniem, M., Melançon, G., Otjacques, B., Pinaud, B.: The state of the art in multilayer network visualization. Comput. Graph. Forum **38**, 125–149 (2019). https://doi.org/10.1111/cgf.13610
13. Hogan, A., Blomqvist, E., Cochez, M., d'Amato, C., de Melo, G., Gutierrez, C., et al.: Knowledge graphs: a comprehensive introduction. arXiv:2003.02320 [cs] (2020)
14. Seufert, A., von Krogh, G., Back, A.: Towards knowledge networking. In: Österle, H., Back, A., Winter, R., Brenner, W. (eds.) Business Engineering — Die ersten 15 Jahre, pp. 289–308. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-642-18542-7_14

15. Naughton, R., Hall, C., Zhao, H., Lin, X.: PhD portal: developing an online iSchool doctoral student community. In: iConference, Urbana-Champaign, IL, USA (2010)
16. van Rooij, E., Fokkens-Bruinsma, M., Jansen, E.: Factors that influence PhD candidates' success: the importance of PhD project characteristics. Stud. Contin. Educ. **43**, 48–67 (2021). https://doi.org/10.1080/0158037X.2019.1652158
17. Choi, H.: An academic nomad? Identity formation of iSchool students. In: iConference, Newport Beach, CA, USA (2015). https://www.ideals.illinois.edu/items/73928
18. Wiggins, A., Sawyer, S.: Intellectual diversity and the faculty composition of iSchools. J. Am. Soc. Inf. Sci. **63**, 8–21 (2012). https://doi.org/10.1002/asi.21619
19. Archer, M., Zytko, D.: Social matching systems for research collaboration: a profile page design for university faculty. In: Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing, pp. 146–150. ACM, Austin, TX, USA (2019)
20. Javed, M., Payette, S., Blake, J., Worrall, T.: Viz–vivo: towards visualizations-driven linked data navigation (2016). https://doi.org/10.6084/M9.FIGSHARE.3545663.V1
21. Osborne, F., Motta, E., Mulholland, P.: Exploring scholarly data with rexplore. In: Alani, H., et al. (eds.) ISWC 2013. LNCS, vol. 8218, pp. 460–477. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-41335-3_29
22. Cabanac, G.: Accuracy of inter-researcher similarity measures based on topical and social clues. Scientometrics **87**, 597–620 (2011). https://doi.org/10.1007/s11192-011-0358-1
23. Chen, S., Arsenault, C., Gingras, Y., Larivière, V.: Exploring the interdisciplinary evolution of a discipline: the case of biochemistry and molecular biology. Scientometrics **102**(2), 1307–1323 (2014). https://doi.org/10.1007/s11192-014-1457-6
24. Liu, J., Tang, T., Wang, W., Xu, B., Kong, X., Xia, F.: A survey of scholarly data visualization. IEEE Access. **6**, 19205–19221 (2018). https://doi.org/10.1109/ACCESS.2018.2815030
25. Sarrafzadeh, B., Vtyurina, A., Lank, E., Vechtomova, O.: Knowledge graphs versus hierarchies: an analysis of user behaviours and perspectives in information seeking. In: Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval - CHIIR 2016, pp. 91–100. ACM Press, Carrboro, North Carolina, USA (2016)
26. Dickison, M.E., Magnani, M., Rossi, L.: Multilayer Social Networks. Cambridge University Press, New York (2016)
27. Aleta, A., Moreno, Y.: Multilayer networks in a nutshell. Annu. Rev. Condens. Matter Phys. **10**, 45–62 (2019). https://doi.org/10.1146/annurev-conmatphys-031218-013259
28. Rossi, L., Magnani, M.: Towards effective visual analytics on multiplex and multilayer networks. Chaos, Solitons Fractals **72**, 68–76 (2015). https://doi.org/10.1016/j.chaos.2014.12.022
29. Carroll, J.M., Rosson, M.B.: Wild at home: the neighborhood as a living laboratory for HCI. ACM Trans. Computer-Human Interact. **20**, 1–28 (2013). https://doi.org/10.1145/2491500.24915042
30. Kaptelinin, V., Nardi, B.: Activity Theory in HCI: Fundamentals and Reflections. Morgan & Claypool Publishers, San Rafael (2012)
31. MacQueen, K., McLellan-Lemal, E., Bartholow, K., Milstein, B.: Team-based codebook development: structure, process, and agreement. In: Handbook for Team-Based Qualitative Research, pp. 119–135 (2008)
32. Bergold, J., Thomas, S.: Participatory research methods: a methodological approach in motion. Forum Qualitative Sozialforschung/Forum: Qual. Soc. Res. **13**, 191–222 (2012). https://doi.org/10.17169/fqs-13.1.1801
33. Schummer, J.: Science Communication Across Disciplines. Practising Science Communication in the Information Age: Theorising Professional Practices, pp. 53–66. Oxford University Press, Oxford (2008)

34. Star, S.L., Griesemer, J.R.: Institutional ecology, ‘translations’ and boundary objects: amateurs and professionals in Berkeley’s museum of vertebrate zoology, 1907–1939. Soc. Stud. Sci. **19**, 387–420 (1989). https://doi.org/10.1177/030631289019003001
35. Firpo, D., Zhang, S., Olfman, L., Sirisaengtaksin, K., Roberts, J.: System design for an online social networking app with a notification and recommender system to build social capital in a university setting. In: Proceedings of the 52nd Hawaii International Conference on System Sciences (2019)
36. Luo, L.: Being interdisciplinary : a look into the background and experiences of iSchool faculty members. LIBRES **23**(2), 1–20 (2013). https://doi.org/10.32655/LIBRES.2013.2.3
37. Hammarfelt, B.: Discipline. In: Hjørland, B., Gnoli, C. (eds.) Encyclopedia of Knowledge Organization. ISKO (2016)
38. Salatino, A.A., Thanapalasingam, T., Mannocci, A., Osborne, F., Motta, E.: The computer science ontology: a large-scale taxonomy of research areas. In: Vrandečić, D., et al. (eds.) ISWC 2018. LNCS, vol. 11137, pp. 187–205. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00668-6_12
39. Nguyen, M., Malik, A., Sharma, P.: How to motivate employees to engage in online knowledge sharing? Differences between posters and lurkers. J. Knowl. Manag. **25**(7), 1811–1831 (2021). https://doi.org/10.1108/JKM-08-2020-0649
40. Sarrafzadeh, B., Vechtomova, O., Jokic, V.: Exploring knowledge graphs for exploratory search. In: Proceedings of the 5th Information Interaction in Context Symposium on - IIiX 2014, pp. 135–144. ACM Press, Regensburg, Germany (2014)