User-driven efforts in creating knowledge graph information system

Stanislava Gardasevic

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

Abstract

Purpose – This paper presents the results of a qualitative study that involved students of an interdisciplinary PhD program. The study objective was to gather requirements to create a knowledge graph information system. The purpose of this study was to determine information-seeking practices and information needs of this community, to inform the functionalities of a proposed system, intended to help students with relevant resource discovery and decision-making.

Design/methodology/approach – The study design included semi-structured interviews with eight members of the community, followed by a website usability study with the same student participants.

Findings – Two main information-seeking styles are recognized and reported through user personas of international and domestic (USA) students. The findings show that the useful information resides within the community and not so much on the program website. Students rely on peer communication, although they report lack of opportunities to connect. Students' information needs and information seeking are dependent on their progress through the program, as well as their motivation and the projected timeline.

Practical implications – Considering the current information needs and practices, a knowledge graph hosting both information on social networks and the knowledge produced by the activities of the community members would be useful. By recording data on their activities (for example, collaboration with professors and coursework), students would reveal further useful system functionalities and facilitate transfer of tacit knowledge.

Originality/value – Aside from the practical value of this research that is directly influencing the design of a system, it contributes to the body of knowledge on interdisciplinary PhD programs.

Keywords PhD students, Information seeking, Information needs, Interviews, Knowledge graph, Tacit knowledge, Interdisciplinary education, User personas

Paper type Research paper

Introduction

While the PhD student population is growing exponentially (Cyranoski *et al.*, 2011), attrition rates from PhD programs remain high owing to a variety of factors. Some of the central

issues preventing students retention in PhD programs are attributed to departmental culture (interactions of students with colleagues), degree and quality of the relationship between doctoral students and their advisors or choosing an appropriate topic (Bair and Haworth, 2005). In this article, I examine a way to address aforementioned issues faced by PhD students by offering them a systematic solution, allowing them to make better-informed decisions by uncovering the latent relationships to information resources available to them, as well as the academic paths of their predecessors and peers. This paper reports on a study that will inform the design and development of a knowledge graph-based information system, intended to support data exploration and decision-making. The case taken for this research is an interdisciplinary PhD program and the main user population is students of that program.

Research purpose

The problem that this study addresses is facilitating the discovery of not only relevant information on resources that are considered as necessary for the success of a PhD student (finding an appropriate supervisor, thesis committee members and collaborators) but also information on relevant courses, projects, conferences, labs, grants and other resources that might help with promoting good scholarship. This problem is going to be addressed by designing a prototype of an information system based on a knowledge graph, intended to support students' decision-making processes in choosing among the available resources. The term *knowledge graphs*, has been closely associated with Semantic Web technologies and was popularized by Google in 2012 (Ehrlinger and Wöß, 2016). According to Paulheim (2016), a "knowledge graph (i) mainly describes real world entities and their interrelations, organized in a graph, (ii) defines possible classes and relations of entities in a schema, [and] (iii) allows for potentially interrelating arbitrary entities with each other; and (iv) covers various topical domains." (p. 2)

Overall, the goal of this study is to improve the quality of the PhD academic and research experience for the students in the interdisciplinary PhD program taken as the case study, through the proposed information system. Before moving on to this larger goal, some preliminary information is required. For that reason, in this first part of the study the requirements' gathering was performed to inform the database design and the potential functionalities of that information system.

The research questions this qualitative study seeks to answer are:

RQ1. What are the current information-seeking practices and information needs of the program's PhD students?

RQ2. What type of functionalities should the proposed information system support?

Literature review

Considering that the research reported here is a part of the larger study, the relevant literature draws from three pertinent areas:

(1) issues causing distress in PhD population;

(2) information-seeking practices of PhD students, in general; and

(3) efforts addressing information discovery in academia and using the graph technology.

In a study by Levecque *et al.* (2017), authors report that over half of the PhD student population in their sample is experiencing at least two psychological symptoms, while almost a third of the students in the sample are at risk of developing a psychiatric disorder, especially depression.

The same issue was raised by a recent study featured in a Nature (2019) editorial, stating that over a third of the sample "had sought help for anxiety or depression related to their PhD." To identify possible factors influencing this phenomenon, Bair and Haworth (2005) conducted a meta-synthetical analysis focused on doctoral student attrition and retention. Their work highlights some of the most important issues found across 118 different studies, such as:

- the relationship between doctoral student and advisor;
- students' interaction with peers and faculty;
- their involvement in various programmatic/departmental activities;
- financial support (for example, research/teaching assistantship); and
- general satisfaction with the program.

Another interesting area of research pertains to whisper networks in academia, which are informal chains of information passed privately mainly between women, used to warn about sexual harassers or abusers (Meza, 2017). The information that is transferred through a whisper networks is made available only to those that are part of it, leaving others as excluded, unless an initiative finds its way online, an approach often considered controversial and criticized (dos Santos Bruss, 2019; Jackson, 2018).

The system proposed here will be designed to address most of these issues, especially by facilitating the discovery of relevant connections in one's surroundings and the communication of tacit knowledge and information not available online (for example, quality of course or adequacy of faculty to perform mentorship tasks). Furthermore, the graph-based system will include data on the mentioned issues that Bair and Haworth (2005) found relevant for PhD student retention, with the intention to allow for interactive and intuitive visualization of the categories and their interconnectedness. This may be accomplished by aggregating data on the dissertation mentorship activities of the faculty or their affiliations and collaborations within a project or lab data that is currently available only via numerous websites. This graph exploration would enable users to make sense of dynamically changing domain data, as well as enhance discovery and decision-making (Rogers *et al.*, 2011).

Finally, the study presented here was conducted to determine whether the findings from the previous research applies to the population of PhD students' this case study, and if so, how can the functionalities of the information system support these and other needs of PhD students?

The area of research looking into PhD student information needs and seeking practices is also relevant in this case. Research by Catalano (2013) shows that even though the internet is one of the first destinations for information seeking, people still play a large role in helping graduate students (master and PhD) begin their research. After going to faculty (especially academic supervisor), students consulted librarians and peers. Research also shows that personal networks present an important information source for doctoral students (Barrett, 2005; Vezzosi, 2009). Barrett (2005) further demonstrates that doctoral students are often exposed to educational experiences (conferences, courses, workshops and recommended readings), which helps them discover a gap in research related to their topic. These findings indicate categories that should be included in the design of the information system. Most of the studies dealing with information seeking and information needs of PhD students consider their research-related activities, whereas in this study, I was exploring how this happens in a broader context of fulfilling the program requirements and completing a particular PhD program, which will be discussed in the following section introducing the case study.

An additional relevant research domain is related to the development of graph-based models and systems that tackle the issue of information discovery in academia. Considering

that "science can be described as a complex, self-organizing, and evolving network of scholars, projects, papers and ideas" (Fortunato *et al.*, 2018, p. 2), there is a plethora of studies using graph technology and social network analysis (SNA) approaches, for example, those based on co-authorship and/or publishing venues graphs (Cabanac, 2011; Makarov *et al.*, 2018; Tchuente *et al.*, 2013) in efforts to facilitate resource or expert discovery. A body of research based on Semantic Web principles has produced some noteworthy models of the scholarly domain. One such example is VIVO Ontology for Research Discovery [1] implemented on the OpenVIVO platform, which can be used by individual researchers of institutions to upload the data and explore/navigate author-topic connections or co-authorship networks (Ilik *et al.*, 2017). Another comprehensive effort to map the scholarly practices is presented by Scholarly Ontology, aimed at encoding the activities taken by digital humanities researchers, while being potentially applicable to other domains (Pertsas and Constantopoulos, 2017). Similarly, a prototype named *Rexplore* (Osborne *et al.*, 2013) develops the possibility of exploring research-related data through a graph-based interface even further, by combining semantically conceptualized topics of research with number of SNA metrics to interact with the graph. The aforementioned projects are focused on addressing general and global needs of scholars. The research presented in this paper is more locally focused. Being strongly grounded in a particular place (geolocation, implying an organization, even a unit within), entails having local norms and requirements related to the research topic and practices, while keeping the focus on the needs of the student population. Furthermore, discussed technologies as well as other relevant research regarding expert recommender systems in academia (for example, STEP methodology by Ziaimatin *et al.* (2013) are designed to suggest experts in a given field based on their research (usually faculty or researchers). In the case where PhD students are the user population, peers might serve as equally valuable information sources (Catalano, 2013). Therefore, the recommender system proposed in this research will also include people that are in different career stages, even though they are not considered to be field experts yet.

Introduction to the case
The case taken for this research is an interdisciplinary PhD program located in the USA. Currently, there are about 30 PhD students and candidates (out of which 5 are distant learners), as well as over 100 alumni.

This program makes a particularly compelling case. Rather than being part of a single school or department, this is an interdisciplinary program, comprising four different departments administered across three schools, each of which have their own professors, research disciplines, jargon and norms. In total, there are over 40 faculty members that might serve as dissertation supervisors and committee members for the students of the program. This means that students have a lot of freedom to pursue and develop special custom lines of research. This is the opposite of more prevalent models – especially those in the natural sciences – where affiliation with a lab or particular professor influences one's research direction and further collaborations.

All students in this program are subject to the rules and procedures for this degree, outlined in the Policies and Procedures [2] document, summarized in the Table I. Over the course of their degree, students move through three stages (pre-proposal, dissertation proposal preparation and "ABD" or all-but-dissertation stage) and each of which has important milestones and requirements. One such requirement applicable to each stage is participation in weekly interdisciplinary seminar. The timeline outlined in Table I was taken as a basis for developing the data collection instruments for the analysis of the collected data and outlining the relevant categories of the database model.

Stage	Deadline	Program requirements
1st Pre-proposal	End of 3rd year	Pass three core courses with a grade of B or better Pass at least one approved research methods or data analysis course with a grade of B or better Pass one primary area examination (comprising two exams) and two secondary area examinations Publish or obtain approval of a research paper under the guidance of a research mentor with whom student has enrolled directed reading credits
2nd Dissertation Proposal Preparation	End of 4th year	Choose a dissertation committee chair and work with that chair to select other committee members and write the proposal (total of five members from at least two different departments) Pass the proposal defense
3rd "ABD" – All but Dissertation	End of 7th year	Pass the dissertation defense and turn in the final revised dissertation with signatures

Theoretical implications

I am using activity theory as a grounding framework, as this study seeks to understand the actions of information seeking for fulfilling program requirements, which is related to the overall goal of obtaining a PhD degree. The activity theory perspective has been used in human computer interaction (HCI) research with the primary concern on the tool mediation, while investigating the ways in which technologies are appropriated by individuals and groups. Activity theory has also been used to inform how interactive tools should be designed to make a positive impact on human activities (Kaptelinin and Nardi, 2012). Five main constructs that are an integral part of this theory are *activity*, *object-orientedness*, *tool mediation*, *internalization/externalization* and *development*. In the case of this study, students (*actors*) are motivated to fulfill each of the program requirements (*objects*), for example, take courses/exams, publish a paper and choose committee chair and members. For that reason, they are taking certain *actions* toward achieving these goals. The progress in the program is considered as a *developmental* process, while the ultimate *object* of the study (*activity*) is to finish the dissertation and leave the program with a PhD degree. The information system that is going to be designed will be a technical *mediation/tool* intended to support decision-making (*internalization*) for each of these steps, that would result with the decisions and actions taken toward the goal, resulting in the changes in the environment and consequently the representation of the setting in the graph (*externalization*).

Study design

This study was motivated by an interest in understanding the personal perspectives and experiences of being a PhD student at the program in question. Described above, this particular program is fairly unique and based on their interests' students may navigate across four unrelated departments, providing a particularly compelling case study to investigate information seeking for decision-making related to their studies.

Participants for the interviews were chosen as a purposeful sample from different categories. The cases/categories of recruited students are as follows:

- latest cohort students/pre-proposal phase (four participants);
- dissertation proposal preparation students (two participants); and
- alumni (two participants).

General demographics of the participants include six females and two males, out of which four are international (non-permanent residents) and four are from the USA. Recruiting participants at varying stages of progress allowed for insights about key junctures in the program to be revealed. To establish a baseline about how new students begin their search for information, I invited students from the newest PhD student cohort to participate in the study, owing to their assumed lack of initiation into the informal flows of information in this degree program, especially when compared with their more established peers. To gain the perspectives of students further along in the program and of those who had managed to complete the program, another four participants were invited to discuss their information needs and information-seeking activities.

The data discussed herein was collected via semi-structured interviews. Current student participants (total of six) were also asked to perform a usability study of the program website that was based on a script with set of tasks, as well as a cognitive walkthrough with "think aloud" protocol (Reeves *et al.*, 2005). The website is currently the only online information resource for the program. This way, students demonstrated some of their information-seeking actions, while commenting on the usage of the information currently available to them.

As mentioned in the Research Purpose section, the study reported here is part of a larger project, and other methodologies will be used to reach the overall goal and to design an information system prototype. The most prominent planned methodology is participatory design, where all of the current students of the program will be invited to contribute to system design efforts. These user-driven research approaches will be complemented with several data-driven approaches, and multiple algorithmic operations will be applied to the data considered as relevant by the study participants. Some of the examples of the operations are:

- SNA algorithms (to extract and visualize the social current and latent connections of the people in the community based on their activities); and
- topic modeling (Blei, 2012) to extract the research topics and approaches based on publications of people in the community.

Data analysis

Interview and usability study data were transcribed and coded in NVivo based on the categories that are considered relevant to the research questions, as well as program requirements outlined in the program policy document. Furthermore, exploratory coding identified the categories that emerged as relevant to this sample that might not have been anticipated beforehand.

"User personas" were used to ensure participant anonymity – a domestic (USA) student and an international student. The persona usage is a common occurrence in HCI research and the requirements gathering stage of system design, where potential users are modeled based on their demographic and behavioral data obtained from real user observation (Pruitt and Grudin, 2003). Following are the descriptions of the user personas.

Domestic (US) student

Brian is a domestic (USA) student. Coming to the program, he was looking for the best fit for him and his family, both regarding academics and lifestyle. He has a strong background in a particular domain and has been working in higher education, but he wants to expand and develop expertise in at least one more discipline. He is self-driven, forward-thinking, and motivated to complete the program as fast as possible. Therefore, he reads the policies and requirements first,

then goes to people (mainly the program chair or academic supervisor) if clarification is needed. Currently, he has two part-time jobs while studying and is taking a mixture of courses that are required and/or of interest, but most important to his decision is that the course is conveniently scheduled. Brian does not have a well-defined research topic and is looking for a personality fit with the supervisor, with whom he will negotiate the appropriate direction.

International student

Petra is an international student from the Czech Republic. She has always wanted to study in the USA, so she found a program in a great location that fits her interdisciplinary background. She is there to learn, explore different areas and courses, and expand her experience- both academically and personally. She is single. Also, she is certain about the topic that she wants to do for her research and would like to find a professor to support her. The bureaucracy and the customs in this foreign land are still strange to her, sometimes overwhelming. Petra is not on the fast track and is not in a hurry to finish the program. While working part time as a graduate assistant, she is taking it one step at a time and doing her best to fulfill the program requirements. She is very reliant on people (especially other students) when it comes to information discovery and looks for a leader in a supervisor that will help her with her goal.

Here I note that not all participants fit exactly into the two personas, but rather they reflect the general traits of the students in this research sample, while still allowing for some degree of anonymity.

I, now turn my attention to themes that were present across the interviews, elaborating on information-seeking practices and some of the most common information resources, but also giving an overview of issues of importance for the study participants, such as social connections, time as a dimension and privacy.

Information seeking and information sources

The research information seeking practices of PhD students is a topic that has been thoroughly covered (Catalano, 2013; Spezi, 2016). My research suggests that domestic students tend to start the search for sources relevant to their topic with the library website, while using Google to refine search terms. On the other hand, three out of four international students went straight to Google Scholar.

On a different note, in this section, I examine another type of information-seeking activities, those that are based on needs related to fulfilling the requirements and successfully completing this specific program. My first pass through the data was to determine what source(s) of information my participants relied upon and the results, together with the number of occurrences in the analyzed text, are shown in Table II. *People* were the most mentioned information resource, whereas other resources included the program website, the weekly interdisciplinary seminar, other classes and websites, as well as a *Facebook group* that connects all of the program-affiliated people, including alumni.

Type of information source	No. of occurrences
People	118
Program website	30
Interdisciplinary seminar	9
Other classes	8
Other websites	7
Facebook group	4

Table II.
Number of interview references to information sources

People as information sources
As mentioned previously, of all possible information sources, *People* were by far the most used, with a total of 118 references across the eight interviews. Specifically, 78 mentions were referring to other students and 40 to professors. Seven out of eight participants pointed to the program chair as the main source of information, yet two students noted a discrepancy in information flow when there is a rotation in the position. Furthermore, there were also references to a dissertation supervisor, especially when it came to research related issues (possible publication venue). Still, ten references were about a lack of contact that is considered necessary, whether because of inability to reach to other students or insufficient responsiveness on the part of professors. This lack of opportunity to communicate with faculty was noted by half of study participants.

Students mostly exchange information within their cohort, which has proven difficult in smaller ones (in some years only two to three students may join the program). They also pointed out the necessity of connecting with students from other cohorts and the lack of opportunity to do so, a phenomenon that will be further discussed in the next section.

In large part owing to the unique nature of this specific PhD program, it makes sense that students will seek help from other people in their surroundings, whether those people are students or faculty. This is especially the case with international students that are not too familiar with the academic and bureaucratic processes and norms pertinent to the USA, which may differ significantly from their home countries. When asked about finding relevant information for international students, Petra replied:

> I don't believe that's something the website can answer- it's like levels- the first level I got the information from the website, until there was no information to mine from there; and then I moved on to talking to whoever it was in the faculty who could help me. But, at some level, when you need workarounds, you have to talk to students. You have to talk to other people who are international and have been here for a while and have gone through this, and that's where you really get the information.

Brian will also go to other students for advice:

> I've needed information about other [non-core] faculty or about the program, I've gone to senior students simply because they have the experience.

Although the composite US student is driven, self-guided and has read the policy several times before coming to the program, he will still to go to the academic supervisor or program chair for clarification.

Furthermore, PhD students who previously attended or are in some way are affiliated with the university in question, often serve as information sources to other students. These students feel more comfortable with independent information seeking and do not rely as much on other people in the program. On the other hand, international students will often go to peers and senior students for clarification, and they have expressed an interest in finding out more about the topics they could potentially discuss with these students. Out of eight students that participated in this research, seven have had positive experiences communicating with others to get needed information, whereas one student felt that the information was being withheld and people were not willing or were hesitant to share.

The program website
During the interviews, six out of eight students stated that they have barely used the website as a source of information, except when they have a particular need at a given point in time, for example, looking for exam areas and their committee members. These time-related information

needs and seeking patterns are discussed further in the next section. In addition, these claims were further examined during the program website usability study, and the following section elaborates more on the findings from that part of the research.

International students were less acquainted with the contents of the program website than the US students. Interestingly, during the interview stage, two students that claimed they knew the website content well were surprised with the information found on it during the usability exercise performed. This point proves the necessity of complementing the findings from the interviews with observations of the actual interaction with the program website as the main online information source for this program (Nielsen, 2001).

Three students were critical of the design of the website, its usability and the way the information was presented ambiguously. Additionally, the lack of regular updates was noted as needing improvement. Brian commented on the activity of looking for professors for collaboration based on their website profiles:

> It was pretty much going through research of everybody, and even when they say they have an interest, it's not necessarily their interest. And then, things change.

The website usability also raised questions of issues of interactivity and personalization of the information for individual students. Several interviewees mentioned the need to have more feedback on their standing in the program as well as what is expected/recommended for them, especially considering the number of possible paths one might take. At the moment, the program chair is the only one who can give such feedback and no system is in place to track the requirements, especially in the first stage of the program. Three students noted that the program chair at the time was not sufficiently responsive to that matter, which left them worried. To that Brian added:

> I feel it'd be nice if we also had the ability to customize our timeline for ourselves and have something to check off, so we know that we've done the things that we really need to do.

The importance of time

Another theme that emerged as significant from the data is the difference in perceiving time-pertinent issues. This influences perceptions and behavior that brings about information needs and seeking. For students to remain in good standing, it is crucial for them to fulfill the program requirements in the given timeframe, as shown in Table I. Therefore, information needs and seeking activities are dependent on the point in time or the stage of progress in the program, as well as motivation of the students and their projected timelines. On the one hand, Brian was looking at the requirements for at least one semester ahead. He said:

> I did a lot of research ahead of time to figure out what are the exams, what are the exam requirements, what are the courses that are affiliated with exams.

On the other hand, Petra is taking one step at a time, focusing only on imminent stage and short-term goals. When asked about her plans for future semesters, she said:

> We'll cross that bridge when we get to it.

Regarding seeking relevant information about the program, the more informed students pointed to the mentioned policy document and the timeline information on the program webpage, outlining expected results.

Time as a dimension is interesting in other senses. While Brian hopes to finish the program in as little as three years, Petra is on a loose five-year timeframe, recommended as

optimal for this program. This might have influenced her information-seeking style, which is more casual than Brian's. Petra stated:

> I don't gather information that is not necessary to me in the immediate future, that's just because I would have to re-gather it, so I just don't.

The importance of social connections

One of the surprising issues that emerged from the data, and noted by all of the interviewees, is the need for more opportunities to connect with other people from the program. Scholarly community plays an important role in the socio-psychological well-being of PhD students, and they depend on support from the faculty and their peer network (Gardner, 2007; Stubb, *et al.*, 2011). Similarly, the results of this study show that community is vital as the means to get needed information. Findings discussed in this section could potentially be valuable for other interdisciplinary programs (i.e. iSchools) where several departments are contributing to the education of interdisciplinary researchers.

Participants suggested social activities ranging from meeting in informal setting (e.g. in a campus bar) or more formal organized events and seminars that would allow for such interactions, as current events are not sufficiently supporting this need. About this phenomenon Petra noted:

> I didn't get a chance to really get to know students from the cohorts in the years prior. I already met some, but we don't have a lot of chances to interact, only the [required] seminar and if we are in the same class.

As mentioned above, the social connections within one's cohort are the most prevalent. Considering the interdisciplinary nature of the program and the freedom to choose and shape their research direction, each student carves their own path by choosing courses (for example, a research methodology course) or collaborations pertinent to their line of research, which reduces their chances of meeting other students regularly. Even though there is the requirement of attending a weekly 1h interdisciplinary seminar, when Petra was asked about the ability to ask other people for advice on courses they took, she commented that there is a lack of opportunity to start this discussion face-to-face at the seminar, because:

> [. . .] people would usually come at 4:29 PM and leave at 5:31 PM. So, there's no chance to talk to anyone after that.

Thus, findings suggest that students generally wish to have more socialization opportunities with not only other students but also with professors. About this phenomenon Brian commented:

> You kind of feel alone, or at least I had felt that way, in this program. That's been a bit of a difficult side to me. But I guess the benefit is that we all get to go in the directions that we want to.

Another way of meeting and connecting to people noted both by Brian and Petra is introduction, whether by faculty members or by other students. While Petra said that one of the professors was making a lot of efforts to introduce her to other professors that might be helpful in her research, Brian gave an example of being referred to another student, which resulted in a relationship based on more than shared academic interests:

> When I came in my first year, there was a fourth-year student who the chair connected me with, because [that student] had kids like me, and lived down the same street. So, we spent time together and I would ask [that student] about requirements, progress, and experience. I think in that aspect I was a bit of an anomaly because I think first year students don't talk to fourth years a lot because they are in such different places in the program.

Finally, social dynamics emerged as an interesting phenomenon when forming the dissertation committee, as a point of potential distress. About this Petra said:

> You want to be careful how you chose them [five people on the committee]. It is more important that they get along with each other, than how you get along with them!

Privacy concerns

As mentioned in the Introduction, this study was conducted to gather requirements for designing an information system that would serve this community. So, when the participants were asked about different functionalities of the system and their preferences, four students (three of who were international students) expressed their concern about privacy issues. The program in question is a small one and the students do not always feel safe expressing their honest opinions on the courses or professors in writing because of the possibility of identification. The same goes for the aggregated ratings (i.e. star rating) or reporting on exams, where identities could still be assumed. When asked about aggregated statistics regarding success rates of particular exam areas, Petra commented the idea as:

> [. . .] not good, because a person who failed might be self-conscious about being identified.

Participants mentioned that students would feel safer sharing such information if faculty were not able to access the system. As Brian said after being asked about his willingness to leave such comments:

> [. . .] It would be lovely. My concern is about the anonymity of i [. . .] I would love to be able to share that in a systematic- everybody can know this- kind of way, without it being attributed to me.

Denying faculty access to this part of the system was suggested by Petra as well, when discussing the previously mentioned Facebook group for all of the people in the program, as an information source and communication channel:

> There is a hierarchy problem in the group- we are with the faculty in the group. There needs to be separation. There needs to be another group without faculty.

To this, she added more concerns:

> I doubt how much a system could protect the students. Because, I think each student has things to say [. . .] but, I definitely see that some professors have tendency to play the power against students.

Discussion and implications for the system design

The preliminary qualitative study reported here was performed to identify user needs and current information-seeking practices, as well as study participants' interaction with the main online source of the program related information – the program website. All of this was with the goal to address those needs with particular features of the proposed information system. Not only was this study conducted with the objective to gather requirements to inform the aforementioned information system design but also to inform the modeling of the graph database.

The findings show that most of the useful information in this case resides within the people in the community and not on the program website or other web resources. Thus, the proposed knowledge graphs technology seems appropriate, through which both social and knowledge networks can be modeled. The special emphasis would be on visualizing social connections between people based on their collaborations and the outputs of their activities (such as co-authorship, shared research projects, labs, workshops and publishing venues).

Other connections that should be recorded and visualized are those pertinent to education as the subdomain of science as a domain (coursework, examinations and dissertation mentorship). In addition to aggregating the information that is currently available via disparate web locations or not available at all, the graph would support recommendation based on latent collaborations based on shared research topics and research methodologies extracted from the publications of the people in the community. This way access to the useful information that is available only to those conveniently located in the social network (as in the case of whisper networks) would be democratized. Based on the themes that emerged and that are described earlier, the following section will discuss some of the functionalities that should be supported by the proposed system.

Information on people

As shown earlier, *People* are by far the most important information resource in the case of this PhD program. For that reason, one of the main features of the system should support the discovery of both students and faculty that could be helpful. Still, the information needed is dependent on the progress in the program, because there is a significant difference in the resources considered helpful for those who just joined the program and those who are preparing to defend the dissertation. Thus, the recommendation system should include a feature of recommending other students that are either in the same stage in the program (i.e. in cases where we want to form a study group) or, those in the more advanced stage, that could serve as a model or mentor users.

International students are interested in knowing more about other students in the program and the system should support exploration based on their research and personal interests, but also allow for their online communication. Presenting personal social networks could be beneficial for several reasons:

- looking for advice about potential research supervisors;
- looking for a way to be introduced to others in the community; or
- as another way of creating social connection.

All of these features would directly answer to the needs outlined as helpful by study participants. Still, the privacy concerns are valid, so the system should also support the connection with other students in a face-to-face or personal messaging manner, to avoid leaving a written trail on sensitive matters.

This feature of pointing to the potentially helpful person in one's environment is crucial, especially when tacit information is what needs to be conveyed. Much of the information noted as useful is tacit, especially for international students (for example, information related to health insurance and benefits is often puzzling for newcomers). However, students would first have to volunteer to be contacted and/or populate the relevant categories, as the relevant data in this case is protected by the Family Education Rights and Privacy Act [3]. This feature would support establishing missing social connections within the community, for those not only physically present and attending weekly seminars but also distance learning students. Belonging to a scholarly community is an issue playing a key role in how students experience their doctoral journey (Bair and Haworth, 2005), and participants of this study expressed there is a lack of community support in case of the program investigated in this study. Therefore, the feature of linking people in the community, it is intended to address this issue by strengthening the social capital in the program and by supporting retribalization. Visualizing the social network would also facilitate the discovery of the current collaborations between students and faculty, information that is not currently

available and hopefully promote the flow of information that may be helpful when choosing a research advisor and committee members, another crucial factor influencing PhD students' satisfaction (Bair and Haworth, 2005).

On the other hand, domestic students would like an information system to support their further explorations of the faculty, their research, publishing collaborations with other students and the courses they teach. Luckily, most of the data on faculty and their activities are already publicly available online, yet are widely dispersed. This information on faculty (for example, courses and exams they chair, their research topics, collaborations, mentorship activities, labs, publication venues and methodologies they use) may inform PhD students on helpful resources. Therefore, aggregating all of that data in a single knowledge graph may facilitate information discovery and decision-making. For example, visualizing the graph of previous dissertation mentorship activities of faculty would make the faculty collaborations more apparent, a feature, which may be useful when assembling one's dissertation committee.

Program support
To address the need for personalized feedback and support for students during the program, the proposed system should allow students to maintain personal profiles and record the steps taken toward fulfilling program requirements, which should result in the functionality of a progress check. Furthermore, as discussed previously, in case they opt for this data to be visible to other students, others would be able to reach out for advice based on populated categories.

The system should be made to accommodate both of the information-seeking styles demonstrated through the user personas. For example, when providing model timelines by visualization of the aggregated achievements of the previous students, both the absolute minimum time required to achieve them, as well as the average/more typically expected time needed to meet the milestone would be helpful. This would allow the system to provide appropriate prompts and suggestions for the next possible steps.

Future steps
It is crucial to implement the viewpoints of the user group when designing the system, so the findings of this study may be validated during future participatory design workshops. In this approach, participants' interpretations are taken into consideration throughout the project, to envision better outcome, while reshaping and transcending the current situation (Spinuzzi, 2005). Also, it is crucial to include the values and ethical norms of the group, to be in line with the value-sensitive design principles (Centivany, 2016).

In this context, it is important to investigate the willingness of students to share information and help the community with their actions by populating the system with data that can serve as information points for the seeker. This could be encodable data (for example courses and exams taken) or other information points which signal opportunities to contact particular people to obtain tacit information (for example experiences of international students with different bureaucratic issues). Therefore, the participatory design research method will be harnessed in examining this aspect, as well as negotiating the ontology representing the domain and visualizing knowledge graph.

Notes
1. https://bioportal.bioontology.org/ontologies/VIVO
2. www.hawaii.edu/cis/wp-content/uploads/2016/10/CIS_Policies_June_2014.pdf
3. www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html

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, pp. 481-534, available at: https://doi.org/10.1007/1-4020-2456-8_11

Barrett, A. (2005), "The information-seeking habits of graduate student researchers in the humanities", *The Journal of Academic Librarianship*, Vol. 31 No. 4, pp. 324-331, available at: https://doi.org/10.1016/j.acalib.2005.04.005

Blei, D.M. (2012), "Probabilistic topic models", *Communications of the Acm*, Vol. 55 No. 4, pp. 77-84. available at: https://doi.org/10.1145/2133806.2133826

Cabanac, G. (2011), "Accuracy of inter-researcher similarity measures based on topical and social clues", *Scientometrics*, Vol. 87 No. 3, pp. 597-620, available at: https://doi.org/10.1007/s11192-011-0358-1

Catalano, A. (2013), "Patterns of graduate students' information seeking behavior: a meta-synthesis of the literature", *Journal of Documentation*, Vol. 69 No. 2, pp. 243-274, available at: https://doi.org/10.1108/00220411311300066

Centivany, A. (2016), "Values, ethics and participatory policymaking in online communities", in *Proceedings of the Association for Information Science and Technology*, Vol. 53 No. 1, pp. 1-10, available at: https://doi.org/10.1002/pra2.2016.14505301058

Cyranoski, D., Gilbert, N., Ledford, H., Nayar, A. and Yahia, M. (2011), "Education: the PhD factory", *Nature*, Vol. 472 No. 7343, pp. 276-279, available at: https://doi.org/10.1038/472276a

dos Santos Bruss, S.M. (2019), "Naming and shaming or 'speaking truth to power'? On the ambivalences of the Indian 'list of sexual harassers in academia' (LoSHA)", *Ephemera: theory and Politics in Organization*, Vol. 4, pp. 721-743.

Ehrlinger, L. and Wöß, W. (2016), "Towards a definition of knowledge graphs", *in Ceur,* Vol. 1695, Joint *Proceedings of the Posters and Demos Track of 12th International Conference on Semantic Systems – SEMANTiCS2016 and 1st International Workshop on Semantic Change and Evolving Semantics (SuCCESS16)*, *Leipzig*, available at: http://ceur-ws.org/Vol-1695/paper4.pdf

Fortunato, S., Bergstrom, C.T., Börner, K., Evans, J.A., Helbing, D., Milojević, S., . . . Barabási, A.L. (2018), "Science of science", *Science*, Vol. 359 No. 6379, available at: https://doi.org/10.1126/science.aao0185

Gardner, S.K. (2007), "'I heard it through the grapevine': doctoral student socialization in chemistry and history", *Higher Education*, Vol. 54 No. 5, pp. 723-740, available at: https://doi.org/10.1007/s10734-006-9020-x

Ilik, V., Conlon, M., Triggs, G., Haendel, M.A. and Holmes, K. (2017), "OpenVIVO: transparency in scholarship", *Frontiers in Research Metrics and Analytics*, Vol. 2, available at: https://doi.org/10.3389/frma.2017.00012

Jackson, A. (2018), "Women at Yale say they developed a secret way to protect themselves from dangerous men because the school keeps failing them", Business Inside, available at: www.businessinsider.com/yale-sexual-assault-allegations-2018-1 (accessed 15 February 2020).

Kaptelinin, V. and Nardi, B. (2012), *Activity Theory in HCI: Fundamentals and Reflections*, Morgan and Claypool Publishers, CA.

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, available at: https://doi.org/10.1016/j.respol.2017.02.008

Makarov, I., Gerasimova, O., Sulimov, P., Zhukov, L.E. (2018), "Co-authorship network embedding and recommending collaborators via network embedding", in van der, A. (Ed.), *Analysis of Images, Social Networks and Texts, Lecture Notes in Computer Science*, Springer International Publishing, Cham, pp. 32-38, available at: https://doi.org/10.1007/978-3-030-11027-7_4

Meza, S. (2017), "What is a whisper network? How women are taking down bad men in the #MeToo age", Newsweek, available at: www.newsweek.com/what-whisper-network-sexual-misconduct-allegations-719009 (accessed 20 February 2020).

Nature (2019), "The mental health of PhD researchers demands urgent attention", Vol. 575, pp. 257-258, available at: www.nature.com/articles/d41586-019-03489-1 (accessed 14 February 2020).

Nielsen, J. (2001), "First rule of usability? Don't listen to users", from Nielsen Norman Group website, available at: www.nngroup.com/articles/first-rule-of-usability-dont-listen-to-users/ (accessed 28 November 2019).

Osborne, F., Motta, E. and Mulholland, P. (2013), "Exploring scholarly data with rexplore", in Alani, H., Kagal, L, Fokoue, A, Groth, P, Biemann, C, Parreira, J. X. and Janowicz K. (Eds), *The Semantic Web – ISWC 2013*, Vol. 8218, pp. 460-477, available at: https://doi.org/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, available at: https://doi.org/10.3233/SW-160218

Pertsas, V. and Constantopoulos, P. (2017), "Scholarly ontology: modelling scholarly practices", *International Journal on Digital Libraries*, Vol. 18 No. 3, pp. 173-190, available at: https://doi.org/10.1007/s00799-016-0169-3

Pruitt, J. and Grudin, J. (2003), "Personas: practice and theory", Proceedings of the 2003 Conference on Designing for User Experiences - DUX '03, p. 1, available at: https://doi.org/10.1145/997078.997089

Reeves, T.C., Apedoe, X. and Woo, Y.H. (2005), "Evaluating digital libraries: a user-friendly guide", available at: www.academia.edu/2799062/Evaluating_digital_libraries_a_user-friendly_guide (accessed 12 February 2020).

Rogers, Y., Sharp, H., Preece, J. (2011), *Interaction Design: Beyond Human-Computer Interaction*, 3rd ed., Wiley, Chichester, West Sussex.

Spezi, V. (2016), "Is information-seeking behavior of doctoral students changing? A review of the literature (2010-2015)", *New Review of Academic Librarianship*, Vol. 22 No. 1, pp. 78-106.available at: https://doi.org/10.1080/13614533.2015.1127831

Spinuzzi, C. (2005), "The methodology of participatory design", *Technical Communication*, Vol. 52 No. 2, pp. 163-174.

Stubb, J., Pyhältö, K. and Lonka, K. (2011), "Balancing between inspiration and exhaustion: PhD students' experienced socio-psychological well-being", *Studies in Continuing Education*, Vol. 33 No. 1, pp. 33-50, available at: https://doi.org/10.1080/0158037X.2010.515572

Tchuente, D., Canut, M.-F., Jessel, N., Peninou, A. and Sèdes, F. (2013), "A community-based algorithm for deriving users' profiles from egocentrics networks: experiment on Facebook and DBLP", *Social Network Analysis and Mining*, Vol. 3 No. 3, pp. 667-683, available at: https://doi.org/10.1007/s13278-013-01

Vezzosi, M. (2009), "Doctoral students' information behaviour: an exploratory study at the University of Parma (Italy)", *New Library World*, Vol. 110, pp. 65-80, available at: https://doi. org/10.1108/03074800910928595

Ziaimatin, H., Groza, T. and Hunter, J. (2013), "Semantic and time-dependent expertise profiling models in community-driven knowledge curation platforms", *Future Internet*, Vol. 5 No. 4, pp. 490-514, available at: https://doi. org/10.3390/fi5040490

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