




Academic support network reflects doctoral experience and productivity

Ozgur Can Seckin¹ and Onur Varol^{1,2*} 

*Correspondence:

onur.varol@sabanciuniv.edu

¹Faculty of Engineering and Natural Sciences, Sabanci University, 34956 Istanbul, Turkey

²Center of Excellence in Data Analytics, Sabanci University, 34956 Istanbul, Turkey

Abstract

Current practices of quantifying academic performance by productivity raise serious concerns about the psychological well-being of graduate students. These efforts often neglect the influence of researchers' environment. Acknowledgments in dissertation subsections shed light on this environment by providing an opportunity for students to thank the people who supported them. We analysed 26,236 acknowledgments to create an "academic support network" that reveals five distinct communities that support students along the way: Academic, Administration, Family, Friends & Colleagues, and Spiritual. We show that female students mention fewer people from each of these communities, with the exception of their families, and that their productivity is slightly lower than that of males when considering the number of publications alone. This is critically important because it means that studying the doctoral process may help us better understand the adverse conditions women face early in their academic careers. Our results also suggest that the total number of people mentioned in the acknowledgements allows disciplines to be categorised as either individual science or team science as their magnitudes change. We also show that male students who mention more people from their academic community are associated with higher levels of productivity. University rankings are found to be positively correlated with productivity and the size of academic support networks. However, neither university rankings nor students' productivity levels correlate with the sentiments students express in their acknowledgements. Our results point to the importance of academic support networks by explaining how they differ and how they influence productivity.

Keywords: Science of science; Network science; Text mining; Scientific careers

1 Introduction

In recent years, well-being and mental health concerns for PhD students have been increasing. According to a recent survey conducted in 2019 by Nature on 6300 PhD students, 36% responded that they sought help for anxiety or depression caused by their studies [1]. Another devastating fact is that doctoral students are 2.43 times more likely to have a common psychiatric disorder than the rest of the highly educated population [2]. It is therefore important to look through the journey of doctoral students not only through the lens of academic "success measures" such as publication numbers, citation counts, fellowships

© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

received *etc.* but also at their overall well-being and the quality of the environment that supports them in fulfilling their potential.

Although obtaining a doctoral degree is often viewed as an isolated process, it is a collaborative endeavor in which family, friends, colleagues, advisors, faculty, and administrative staff are directly or indirectly involved and can influence the well-being of the students. At the end of the journey, students can show their gratitude by mentioning these names in their work through “acknowledgements” section of their dissertations. Acknowledgements, even though existed before, could not be found explicitly in the academic work before 1940s, and did not become a common subsection until 1960s [3]. Hyland named acknowledgements as “Cindirella” genre because of its suffering from an undeserved neglect [4]. Through time, these sections got longer and their use have become more prevalent [5], making them more “insightful” in terms of understanding how and with whom doctoral students complete their journeys. Since there is almost no guideline or style guide to receive help when writing this section [6], students have more freedom, compared to the other parts of their dissertations. Acknowledgements also serve purposes other than expressing gratitude, such as exhibiting associations with respected academics to display a special connection to which the author has been admitted [7]. Thus, introducing their strategic decision in their professional identities by illustrating the author in a positive aspect and governing their connections with the disciplinary community [8].

Acknowledgements contain such profound details of their authors’ academic journey; however, research efforts to study how they vary concerning disciplinary and demographic differences have remained limited. Mantai and Dowling examined the type of social support that are provided for PhD students using 79 acknowledgements gathered from Australian universities [9]. Hyland examined 240 acknowledgements of MA and PhD dissertations to characterize their narrative structure [6].

Using acknowledgement sections to delve into hidden networks outlined by the gratitude and appreciation expressed by students helps drawing conclusions that cannot be obtained from measures of academic success alone. For this task, we examined 26,236 PhD dissertations, obtained from *ProQuest Open Access Dissertations & Theses* database (PQDT-Open hereafter), 99% of which are from the United States in the last 20 years. We aimed to shed light on the doctoral process by examining who is acknowledged, and how they are recognised from the perspective of students, using the tools of network science and natural language processing that enable research on large-scale data. We revealed gender based and disciplinary differences when acknowledging support providers in terms of number of people mentioned and sentiment scores. We also investigated the factors derived from academic support networks influencing productivity levels. Lastly, we point out to linguistic differences between those who are located in the extreme cases of productivity and of sentiment.

2 Methods

2.1 Data collection and information extraction

To run a large-scale analysis of dissertation acknowledgements, we retrieved data from `pqdtopen.proquest.com`, also called PQDT-Open, which provides a publicly accessible thesis dissertation data that allows our work to be reproducible. We collected the dissertations by scraping the data directly from the website using Selenium library offered in Python. We have gathered documents for 47,000 researchers, and 26,264 of them

are doctoral dissertations, written between 2000 and 2020. This collection of dissertation data also included metadata on dissertation abstract, title, author name, university, year of publication, page number, advisor name, department, subjects, and keywords. To extract the acknowledgement subsection from these dissertations, we parsed raw data obtained in Portable Document Format (PDF). We used PyMuPDF library to extract textual information for each page, and then, we utilized a rule-based approach to identify pages that are likely to belong to acknowledgement subsection – e.g., accepting the pages with the first word being “Acknowledgements,” or ignoring pages that contain text such as “Table of Contents,” “List of Figures,” or “Appendix.”

2.2 Data enrichment

Although there is a rich metadata provided by PQDT-Open, there was no gender information or discipline category given. For the former one, we inferred students’ genders using their first names. We used the online service, called `genderize.io`, which relies on a database that has over 110 million entries from 242 countries to examine whether a name is more frequently used amongst females or males. For the latter one, although dissertation subjects were given in the metadata, it was not feasible to run an analysis uncovering the disciplinary differences since we identified 572 unique subjects listed in total. Hence, to have a clearer view of disciplinary differences, we divided the subjects into 5 categories [10]. Although there is no categorization agreed in the literature and guideline or consensus on how research fields should be classified, it could be done by following previous research efforts. To assign each subject into one category, two authors separately labeled each subject with a discipline and reached to an 85% agreement and 0.75 Cohen’s Kappa [11] score for inter-annotator reliability. A detailed list of category – subject is given under Additional file 1, Sect. 4.

2.3 Bootstrapped estimates of sentiment and counts

The bootstrap method is a statistical methodology that involves averaging estimates from several small data samples to infer statistics about a population. We employed this approach to estimate mean and confidence interval of sentiment scores and number of mentions to be able to observe disciplinary and gender differences. We repeated this procedure 5000 times for each estimation. We randomly drew 5% of the population with replacement and calculated the mean of the sample. Using these mean values, we ran two-tailed T-tests to test significance of sample means ($p \leq 0.05$, $p \leq 0.01$, $p \leq 0.001$ and thresholds are denoted by “*”, “**” and “***”, respectively).

2.4 Regression analysis

Since linearity and normality assumptions do not hold in our case and our target variable (publication count of doctoral students) follows approximately an Inverse Gaussian distribution, we employed a generalized linear model with Inverse Gaussian distribution. After selecting the appropriate regression model, to detect multicollinearity and select the variables that are going to be used in regression analysis, we checked the variation inflation factor (VIF) and removed those which had higher than 10. The remaining variables were used in the regression analysis.

3 Results

3.1 Characterization of a support network

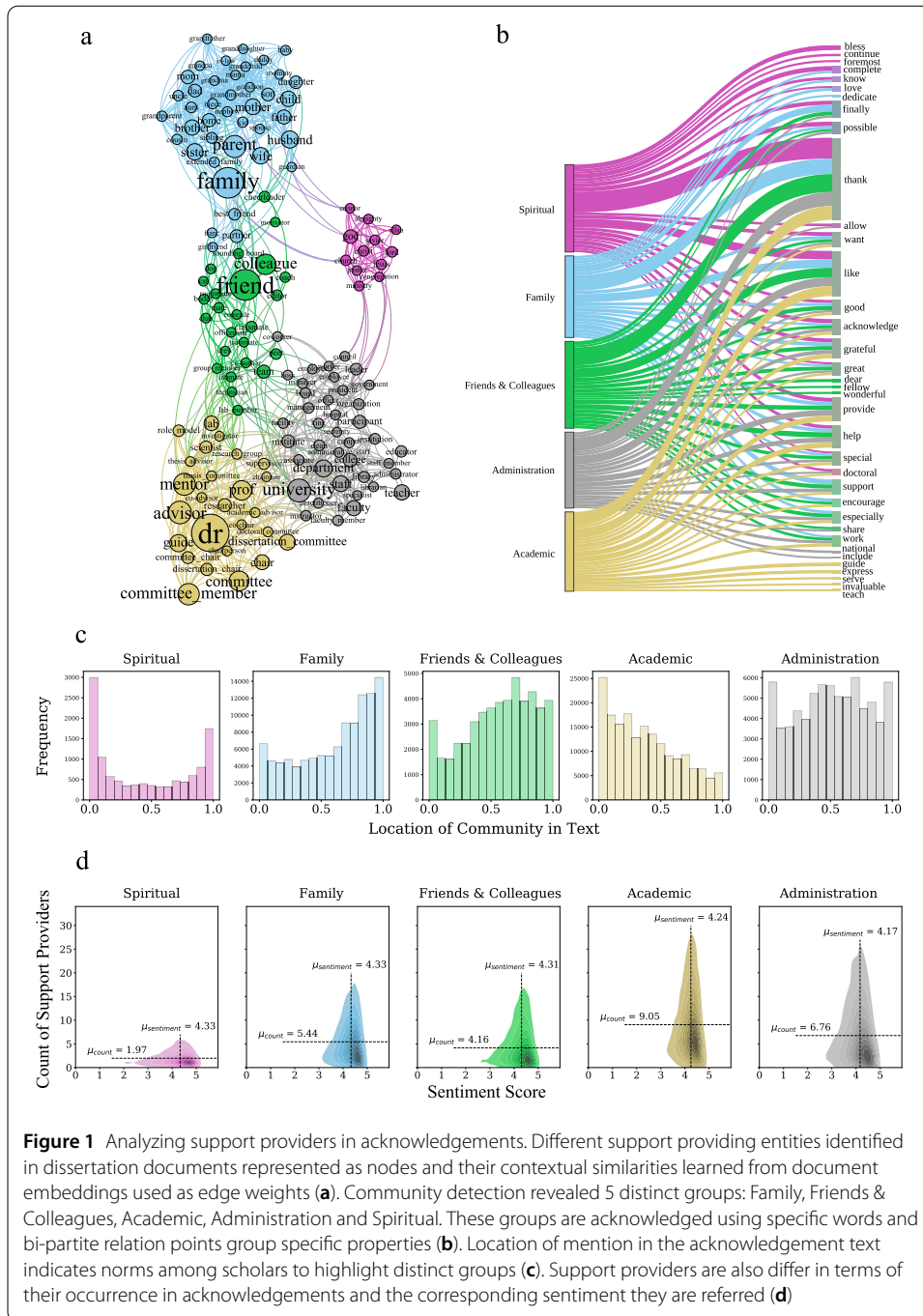
The acknowledgements section of dissertations contains statements about individuals or institutions who have provided emotional, economic, and administrative support to students on their journey towards attaining their degree. To systematically identify acknowledged individuals and institutions, we used a data-driven approach supported by manual inspection to identify distinct types of support providing entities in the acknowledgements.

To build the academic support network, we extracted different individual roles and institution types as nodes from each text and computed contextual similarities learned from the text as edge weights (Fig. 1(a)). Our entity extraction approach identified 144 support providers that were mentioned in at least 50 dissertations. We used a deep learning approach, called Doc2Vec [12], to learn embeddings for each support provider within the context they were used in the dissertation corpus. Using these embeddings, we calculated similarities as edge weights between embeddings learned for each node. We used disparity filtering and retained only the statistically significant edges (see Additional file 1, Sect. 2), therefore giving us the network that captures significant relations between these entities. Later we employed Girvan-Newman [13] algorithm for community detection to identify groups of support providers.

The network representation of all of the support providers is given in Fig. 1(a). Community detection analysis identified 5 distinct communities in this network and each of them is illustrated by a different color: Spiritual (purple), Academic (yellow), Administration (gray), Family (blue) and Friends & Colleagues (green). These communities are consistent with those identified with other clustering approaches like hierarchical clustering as well (see Additional file 1, Sect. 2). Node sizes were determined by the occurrence of support providers and the edges were weighted with cosine similarity of embeddings between node pairs.

Connectivity among these communities reveals separation between social and professional networks. Friends & Colleagues are located among Family, Academic and Administration communities. Some dissertations also refer to spiritual entities and community consisting of these entities is loosely connected with the rest of the network and has few links with the family community. Factors influencing community relations can be explained by comparing the words that are used to acknowledge these communities. By analyzing bipartite connections of support providers and prominent words, as seen in Fig. 1(b), we present the most frequent 20 words used for support providers in these communities. While four of the most widely used words for acknowledging Spiritual support providers are not linked with the other communities; words like thank, acknowledge, and grateful used approximately at the same rate for each group.

Hyland argued that the structure of dissertation acknowledgements has, in general, a “thanking move” section, in which authors start by presenting the participants, continuing next by thanking them for academic assistance (i.e. intellectual support, ideas), then for resources (i.e. technical, financial support) and lastly for moral support (i.e. friendship, patience) [6]. In our academic support network, we observed a similar narrative structure. To further support rank order in acknowledgements, we checked the locations of the support providers in the text and observed that different communities can be distinguished by analyzing the locations in which they are frequently mentioned (Fig. 1(c)). While aca-



ademic support providers are most frequently mentioned at the beginning of the acknowledgements, students tend to start talking about their families towards the end. We also observed that Friends & Colleagues and Administration are generally mentioned in the middle of the text. In contrast, Spiritual entities are mentioned either at the beginning or at the end.

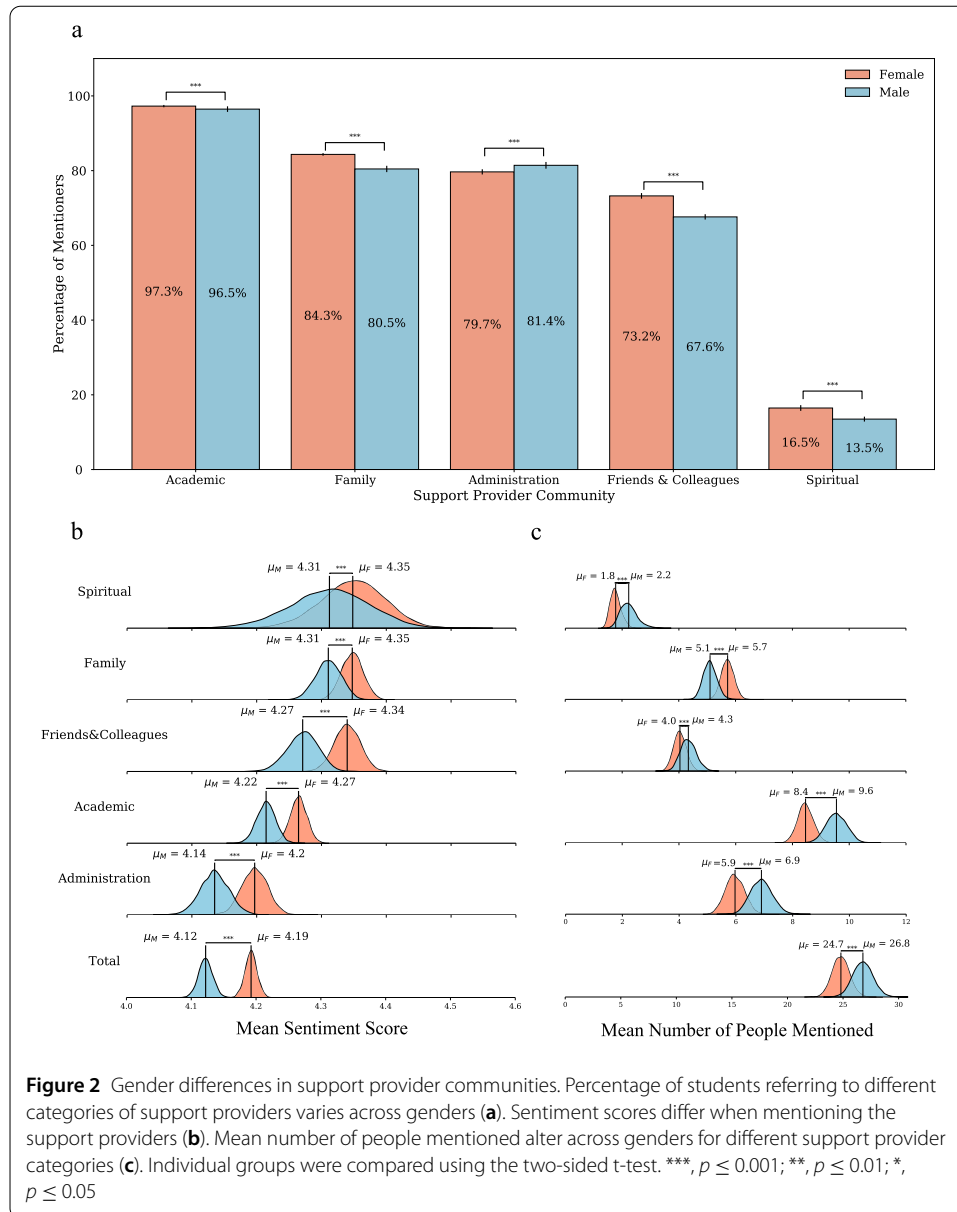
Although acknowledgements are expected to have an overall positive sentiment, certain entities receive more formal tone. To highlight these subtle differences, we explored the interplay between sentiment scores and how many times they are mentioned for each

category separately as shown in Fig. 1(d). The sentiment analysis results have shown that Spiritual, Family and Friends & Colleagues communities are being acknowledged roughly at the same level with average sentiment scores 4.33, 4.33, and 4.31 defined in $[0, 5]$ range. Academic and Administration communities have lower scores on average 4.24 and 4.17, respectively. Similarly, we analyzed the number of people mentioned from these categories. Not surprisingly, PhD students tend to acknowledge the Academic community the most (9.05 people per acknowledgement), it is followed by Administration (6.76), Family (5.44), Friends & Colleagues (4.16), and Spiritual (1.97). Families, friends, and spiritual figures generally do not involve in research as workforce; however, they provide emotional and financial support to make life easier for doctoral students and are a crucial part of the support network deserving an appropriate mention.

3.2 Gender based differences

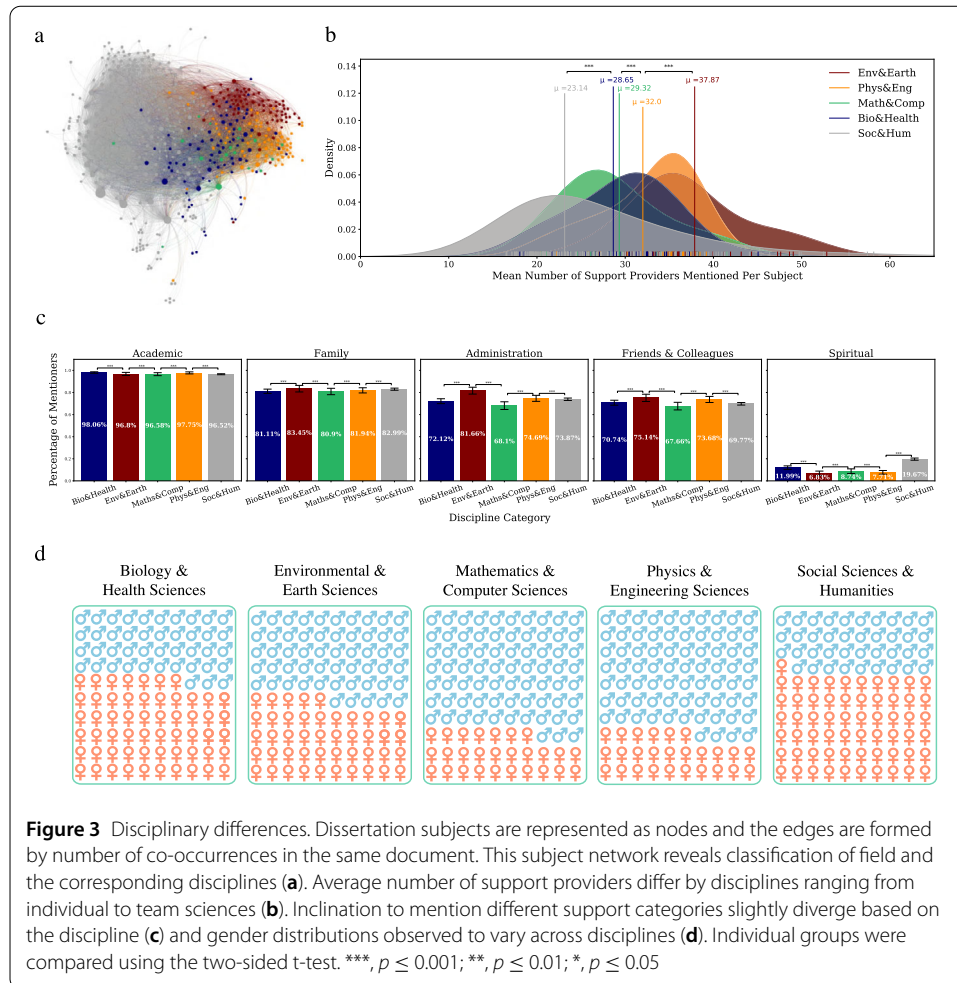
PQDT-Open provides metadata on universities, authors, and committee but lacks details about demographics of the authors such as gender of the author. We inferred this information using the names with a widely used public API and examined the differences between genders in terms of their academic support networks (see Additional file 1, Sect. 3). Previous work studied how female and male students acknowledge support providers both in quantitative and qualitative terms. Alotaibi, using Metadiscourse, studied 120 dissertation acknowledgments written by Saudi students at U.S and revealed that while all male and female students acknowledge their academic environment, there exist differences when thanking God, resources and moral support [14]. It is also shown that women in academia have less access to powerful social networks and inter-personal bounds that provide resources and create other advantages, which limits their opportunities to achieve their goals [15, 16].

Figure 2(a) shows the percentage of students who mention the respective support provider community at least one time in acknowledgements. We observed that female students are slightly more likely to thank each community at least once except for the administration. The largest gap is observed in Friends & Colleagues category where the difference is 5.6% between males and females. This is followed by a roughly 4% difference for the family members. These results differed when we examined the number of mentions instead of percentage of mentions; number of people acknowledged from each category is higher in male students except for the family members. In fact, the highest difference is observed in Academic, Administration, and Friends & Colleagues groups, suggesting more research needs to be done to better understand differences in work culture and environment for women and men. One explanation for our observation might be due to the slight differences between males and females in terms of written language. Previous work has shown that females utilize more terms referring to social and psychological processes whereas males write more about impersonal subjects and object features [17]. Another study examining the language use in argumentative essays by college students demonstrated that male students tend to utilize more nouns related to social and economic activities, while female students are inclined to use more pronouns, intensifiers, and modifiers [18]. Hence, it is possible that there exists no dissimilarity in academic support networks between female and male students in terms of their magnitude since it might be due to written communication differences. Even when we consider the possibility of difference propensity to use acknowledgements, we observe that the administration is the only support provider community where women mentions them less than men.



Besides the occurrence rates, we analyzed the sentiment of the language used for different support providers. Females are inclined to express more positive sentiment towards the ones that help them through their journey. Meanwhile, the gender gap between Friends & Colleagues community seem to be highest; the difference is narrower for the Spiritual characters, but still significant despite the large variance of the distribution.

Looking from an overall perspective, it is readily apparent that females tend to thank their families both qualitatively and quantitatively more compared to males. This is in line with the existing work on dissertation acknowledgements showing that while both men and women appreciate social support evenly, they highlight different aspects of it; men value companionship and collegiality, women note emotional support [9]. Taken together, this may be an indicator of the level of importance of families for females and lack of professional support from the other communities during their doctoral journey.



3.3 Disciplinary differences

Each academic discipline has different research practices and collaborations. These differences are also reflected by mentorship styles and academic environment. To analyze these disciplinary differences, we manually categorized the subjects given in the PQDT-Open metadata into five main category of academic disciplines (Biology & Health Sciences, Environmental & Earth Sciences, Mathematics & Computer Science, Physics & Engineering Sciences, Social Sciences & Humanities) (see Additional file 1, Sect. 4). Figure 3(a) shows the subject co-occurrence network and categories labeled as different disciplines.

We argued that while dissertations on certain subjects may be considered as individual work and require less academic collaboration and administrative support, other subjects might require cooperation, teamwork, and access to resources and field work. We calculated the number of support providers mentioned for each subject and presented in Fig. 3(b). While Social Sciences & Humanities students mention the least number of people with 23.14 on average, this number is the highest with 37.87 for Environmental & Earth Sciences students. Number of support providers mentioned for each discipline aligns with academic norms of individual and team science as shown in the literature [19]. Here, we measured not only size of academic groups, but also other support provider categories.

Moreover, it is also reasonable to presume that different disciplines might have different preferences in terms of acknowledging the support provider categories. While the results

show that there is a small gap between occurrence rates, it varies the most for Administration and Spiritual communities (see Fig. 3(c)). Mathematics & Computer Science students seem to mention their families, friends, colleagues, and the administration less than other disciplines. Additionally, almost one fifth of Social Sciences & Humanities students acknowledge spiritual characters, which may be explained by dissertation studies in religion and relevant fields (e.g. “Biblical Studies”, “Islamic Studies”) covered under this discipline. We also investigated the percentage of genders in these disciplines and, consistent with the past work [20, 21], we observed that female students are underrepresented in STEM fields. Percentage of females is the lowest for Physics & Engineering Sciences with only 26%, which is followed by 27% in Mathematics & Computer Sciences. However, the majority of students in Social Sciences & Humanities are women, with a rate of 71%. These outcomes are also supported by previous work on intersectional inequalities, where it is shown that there is homophily between identities and subject of research [22].

3.4 Social determinants of the academic productivity

Research on academic performance and success focuses on metrics that are easy to quantify, accessible for research, and standardized across disciplines [19]. Efforts on quantifying academic performance at individual and group levels use productivity measures such as number of publications and impact indicators like citation counts [23, 24]. Impact of academic mentorship and institutional quality for academic growth have been recently studied by using these bibliographic indicators [21, 25–27].

We want to investigate academic productivity by utilizing the social aspect of doctoral studies. We investigated the publication records of students obtained from an online service, called Dimensions, for 2824 former doctoral students [28]. By conducting a regression analysis, we analyzed the role of academic support network by considering number of mentions and sentiment scores of acknowledgement to estimate the correlation with the publication count by taking their productivity during the doctoral studies as our target variable while controlling for disciplinary differences and gender. We employed an Inverse Gaussian regression model with a log link function to estimate the parameters and their significance, since the target variable is the publication count and it approximately follows an Inverse Gaussian distribution. In this model, regression coefficients and confidence intervals should be interpreted as multiplicative terms. For instance one unit change on a variable with an estimated coefficient of β affect the target variable as multiplication of e^β . We show positive effect when $e^\beta > 1$ (see Additional file 1, Sect. 5 for details). To capture the productivity during doctoral studies, we used number of publications as a measure and considered the period of doctoral studies four years before graduation and four years after to account for the work in submission or in progress during the thesis defense. Results of the regression analysis are summarized in Table 1. By analyzing the regression coefficients and the significant variables, we assessed the factors influencing the academic productivity and the social determinants of the doctoral students performance.

We investigated the regression analysis to validate our earlier observations about the gender and disciplinary differences. It was shown in the literature on research outcomes that women have slightly less publication rates than men while the difference can be attributed to various systematic biases in academia [29–31]. Especially for STEM fields, empirical data reveals considerable gender variations in number of citations, publication counts and the impact of their academic careers [20, 32]. This phenomenon can be explained by several factors; it is possible to consider that women are underrepresented in

Table 1 Regression results. Inverse Gaussian model for explaining productivity by gender, discipline and textual features extracted from dissertation text

	Publication Count		
	All Students	Female Students	Male Students
gender	-0.1839*** (0.038)		
env_earth	-0.1218* (0.058)	-0.0943 (0.086)	-0.1544 (0.079)
math_comp	-0.0032 (0.063)	0.0420 (0.123)	-0.0318 (0.077)
phys_eng	0.2468*** (0.057)	0.1171 (0.104)	0.2623*** (0.069)
soc_hum	-0.3968*** (0.048)	-0.3907*** (0.066)	-0.4404*** (0.073)
family_sentiment	-0.0375 (0.047)	-0.0426 (0.073)	-0.0566 (0.061)
academic_count	0.0044* (0.002)	-0.0011 (0.003)	0.0071* (0.003)
family_count	0.0004 (0.004)	0.0044 (0.006)	-0.0027 (0.005)
friends&colleagues_count	0.0037 (0.003)	0.0029 (0.005)	0.0073 (0.005)
administration_count	0.0018 (0.002)	-0.0024 (0.003)	0.0048 (0.003)
spiritual_count	0.0032 (0.210)	-0.0506 (0.040)	0.0083 (0.025)
Constant	2.1818*** (0.210)	2.0914*** (0.329)	2.2291*** (0.276)
Observations	2824	1099	1725
AIC Score	16,206	5831	10,383

Standard errors in parentheses * ($p < 0.05$); ** ($p < 0.01$); *** ($p < 0.001$)

scientific cooperation and publishing and struggle with implicit biases since they are more likely to play a significant role in parenting [33], obtain less institutional assistance and have more service duties [34], or the systematic undervaluation of women's involvement and their invisibility in scientific research, known as the "Matilda Effect" [35]. Consistent with the literature, our model have demonstrated that female productivity is lower than that of males when considering simply the number of publications ($M = -0.1839$, 95% CI $[-0.258, -0.110]$), in other words, 16.8% less than males ($e^{-0.1839} = 0.832$, see Additional file 1, Sect. 5 for detailed explanation) and leading on average 10.4% ($e^{-0.110} = 0.896$) to 22.7% ($e^{-0.258} = 0.773$) fewer publications. These gender differences imply that studying the doctoral process may help to better understand the above mentioned adverse conditions.

Another important aspect explaining the productivity is the academic discipline because publication counts vary from one field to another [36], which is a key indicator of quality in higher education since the research performance has an influence on rewards, tenure, promotion decisions and staff recruitment [37–39]. Therefore, it is essential to demonstrate and explain the alterations between scientific fields. When Biology & Health Sciences is taken as the reference group, our model indicates that while the Physics & Engineering students are associated with 28% more publications ($M = 0.2468$, 95% CI $[0.136, 0.358]$), Environmental & Earth Sciences ($M = -0.1218$, 95% CI $[-0.235, -0.009]$) and Social Sci-

ences & Humanities ($M = -0.3968$, 95% CI $[-0.492, -0.302]$) students are affiliated with 11% and 38% less number of papers, respectively. Becher's work on disciplinary differences [40] made similar observation about productivity of disciplines.

However, when we controlled for the gender variable, our findings showed that the magnitude of these effects change. For example, being a Social Sciences & Humanities ($M = -0.3907$, 95% CI $[-0.520, -0.262]$) student is associated with 32% less number of papers for females while this is 36% for males. On the other hand, while male Physics & Engineering ($M = 0.2623$, 95% CI $[0.126, 0.398]$) students are associated with 30% more publications, this is not a statistically significant variable for female students. These results might also indicate the under-representation of female doctoral students in Physics & Engineering fields.

Aside from the demographic aspects, our results demonstrated that the number of people from academic network mentioned in acknowledgements is associated with 0.44% more publications for each person acknowledged ($M = 0.0044$, 95% CI $[0.000, 0.009]$). However, when controlled for genders, the regression analysis suggested that this statement holds only for male students with approximately just 1% more publications ($M = 0.0071$, 95% CI $[0.002, 0.013]$).

Our models do not suggest a statistically significant relationship between the rest of the variables and the number of publications; however revealed the influence of gender and discipline on productivity. Therefore, we normalized sentiment scores and publication counts between zero and one at the individual level by taking into account gender and discipline of a student. More clearly, we filtered out each gender-discipline pairs from our data and normalized publication counts by min-max scaling. These values are then subtracted from the group mean to center around zero. Distributions of normalized sentiment and productivity scores are shown in Fig. 4(a).

Empirical and visual evidence shows no sign of significant links between sentiment and productivity levels. Additionally, we compared the language characteristics of extreme cases for both productivity and sentiment levels to help us understand the mindsets of people from upper and lower quantile of the distributions. To achieve this, we inspect the word usage differences in two groups by using Jensen-Shannon (JS) divergence for words that are used more than 10 times in each group. These words are then represented as word-shift graphs as shown in Fig. 4(b) for sentiment and Fig. 4(c) for productivity [41].

Sentiment scores depend on content and context of texts. Hence, there expected to be certain alterations between relatively more positive and negative acknowledgements. As seen on Fig. 4(b), we observed that most contented 25% of PhD students emphasize gratitude by giving more space in their narratives to words such as *grateful*, *gratitude*, and *thankful*. These results also conformed to the past work which suggests that expressing gratitude helps to increase well-being [42, 43]. Our results also demonstrate that both family and the academic environment are more frequently mentioned in the narrative of the most contented 25%. Figure 4(c) illustrates the JS divergences of words across most and least performing doctoral students. It is apparent that those who over-perform their counterparts emphasize more endeavour related concepts such as *productive*, *busy*, *internship*, and *article*.

3.5 Institutional ranking and student performance

Since the most well-known university ranking organizations such as Quacquarelli Symonds (QS), Times Higher Education (THE), and Center for World University Rank-

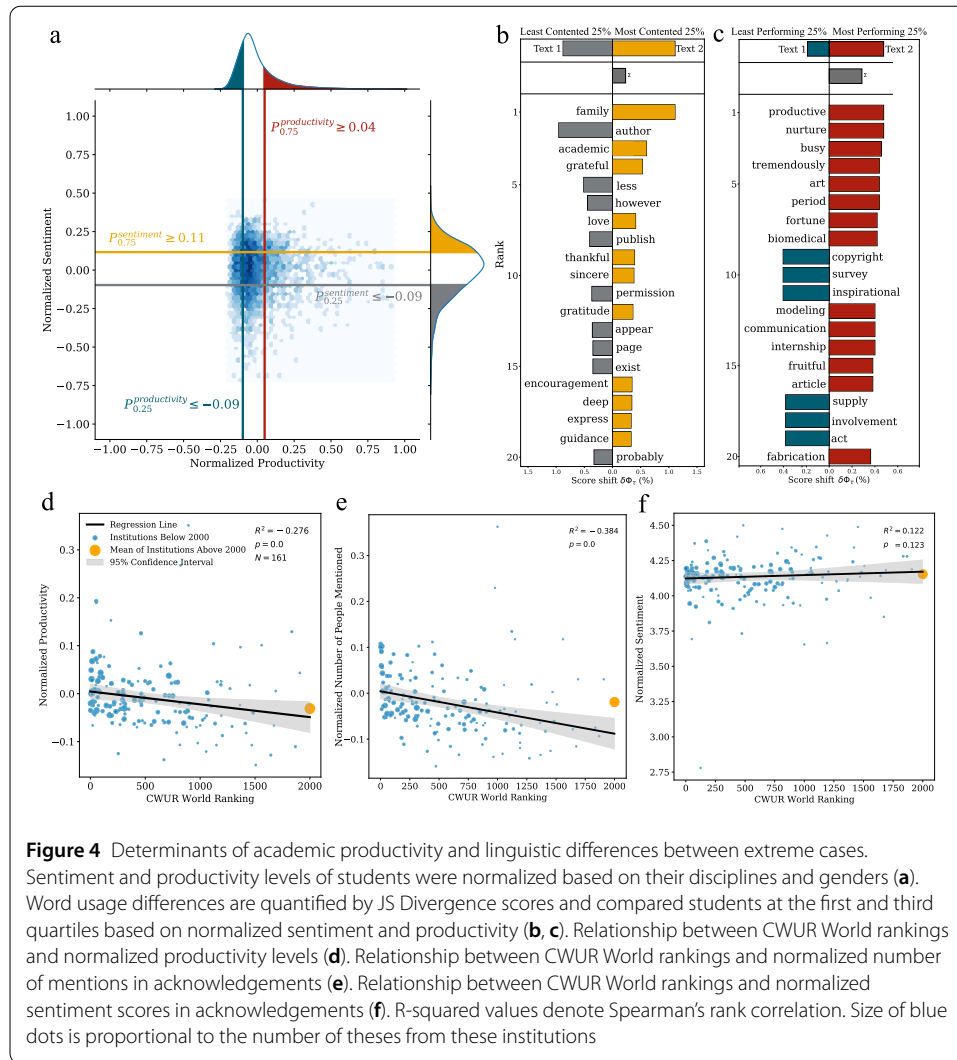


Figure 4 Determinants of academic productivity and linguistic differences between extreme cases. Sentiment and productivity levels of students were normalized based on their disciplines and genders (a). Word usage differences are quantified by JS Divergence scores and compared students at the first and third quartiles based on normalized sentiment and productivity (b, c). Relationship between CWUR World rankings and normalized productivity levels (d). Relationship between CWUR World rankings and normalized number of mentions in acknowledgements (e). Relationship between CWUR World rankings and normalized sentiment scores in acknowledgements (f). R-squared values denote Spearman’s rank correlation. Size of blue dots is proportional to the number of these from these institutions

ings (CWUR) employ “number of research papers published” as a factor in their ranking, we assumed that productivity levels of doctoral students may have associations with the success of their institutions. We present analysis on CWUR ranking, since it provides a more granular and longer list, but our results are consistent for other ranking systems (see Additional file 1, Sect. 6).

We investigated the relationship between university rankings and productivity of graduate students. We found that university rankings have significant positive correlation with the number of publications (Fig. 4(d)). Research environments in these institutes provide more opportunities to publish and introduce them a broader collaboration network as well, partially observed by total number of people mentioned (Fig. 4(e)). Number of people mentioned in dissertations have a higher correlation with institute ranking than the productivity levels, suggesting environment cultivate institutional success more than publications alone. However, there is no associations between sentiment of a doctoral student with respect to the ranking of their institutions (Fig. 4(f)) meaning that the top-ranked institutes provide advantage in professional growth while well-being of the doctoral students mostly determined by their academic support networks.

4 Discussion

Our research uncovered the network of support providers, assisting doctoral students in achieving their goals. We showed that there exist gender and disciplinary differences in acknowledging support providers and sentiment scores when mentioning different communities. Since acknowledgements often appear as the sole section in which students talk about their experience as doctoral candidates, it is noteworthy to observe that the link between productivity levels and their academic support networks can be revealed.

Our results showed that the number of publications among doctoral students varies by academic discipline, with students in social sciences and humanities publishing the least and students in physics and engineering publishing the most. Our data also suggested that productivity is positively correlated with the number of people mentioned from academic environment when publication counts are normalized with regard to gender and discipline. We showed that female students are more likely to acknowledge each support provider group in a more positive sentiment. They did, however, mentioned fewer people from their workplace and published fewer academic publications. Gender differences of support network size can be explained by differences in perceived support or ease of accessing these support providers. Future research can conduct surveys to untangle role and effectiveness of different support providers. Our results also demonstrated that schools with higher rankings provide PhD students wider networks, in which academic environment is significantly bigger and productivity levels as a result are higher. In fact, it is shown in the literature that as the number of writers rises, so does the impact of the research [44], highlighting the importance of young scholars' academic networks.

Quantitative analysis of acknowledgement texts provided a deeper insight on social interactions and experiences of doctoral students as well. Our results suggested that the narrative of the most performing 25% is more centralized on endeavour-related content compared to the least performing 25%. Similarly, most contented 25% show more gratitude towards their family and academic environment compared to the least contented 25%.

It is crucial to note that while support providers from academic communities have a positive influence on productivity, overall well-being of a student require contributions from social interactions with family and friends and administrative support from their institute. Our results showed that higher university rankings or productivity levels do not lead to a higher sentiment reflected towards doctoral experience, but positive influence in their professional growth.

Analyzing thousands of acknowledgement sections, we created an alternative angle reflecting social aspects of the doctoral studies where friends, families, colleagues, and administrative staff have different roles to play ensuring utmost performance and well-being of the student. Therefore, instead of directly analyzing publication counts or number of citations to explain doctoral studies, it may be better to embrace a new approach where students' well-beings and academic support networks are also put forward. People compare themselves to those who are similar to them with regard to demographic and social proximity and how individuals evaluate their own subjective well-being and happiness depends on those of others [45, 46]. It is also known that "success narratives" have an impact on the reader's judgements and decisions [47], which may imply that when doctoral students compare themselves with their counterparts, it would decrease their subjective well-being. Future work may contribute to a profound understanding of how support net-

works influence productivity in late career stages and researchers' overall well-being by reaching out to people and possibly conducting a survey. It is also important to collect theses published all around the world to improve the representativeness of the data and observe how cultural aspects influence the way of doctoral students acknowledge their support providers.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1140/epjds/s13688-022-00369-z>.

Additional file 1. Supplementary information (PDF 6.5 MB)

Acknowledgements

We would like to thank VRL Lab members, Qing Ke, and Nur Mustafaoglu for their feedback and fruitful discussions. O.V. also thanks his academic support network challenging him to do his best.

Availability of data and materials

The dataset generated and analysed during the current study is not publicly available due to the presence of personal information such as student name, surname, school, fundings etc. but are available from the corresponding author on reasonable request after filtering out the mentioned information.

Declarations

Competing interests

The authors declare no competing interests.

Author contribution

OV, and OCS contributed equally on research design, data analysis and manuscript writing. All authors read and approved the final manuscript.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 24 May 2022 Accepted: 14 November 2022 Published online: 22 November 2022

References

1. Woolston C (2019) Phds: the tortuous truth. *Nature* 575(7782):403–407
2. Leveque K, Anseel F, De Beuckelaer A, Van der Heyden J, Gisle L (2017) Work organization and mental health problems in phd students. *Res Policy* 46(4):868–879
3. Swales J (1988) Shaping written knowledge: the genre and activity of the experimental research article in science: Charles Bazerman. University of Wisconsin Press, Madison, pp. 356. *English for Specific Purposes* 9(1), 98–101 (1990) [https://doi.org/10.1016/0889-4906\(90\)90032-8](https://doi.org/10.1016/0889-4906(90)90032-8).
4. Hyland K (2003) Dissertation acknowledgements: the anatomy of a cinderella genre. *Writ Commun* 20(3):242–268
5. Cronin B, McKenzie G, Stiffler M (1992) Patterns of acknowledgement. *J Doc* 48(2):107–122
6. Hyland K (2004) Graduates' gratitude: the generic structure of dissertation acknowledgements. *Engl Specific Purposes* 23(3):303–324
7. Scrivener L (2009) An exploratory analysis of history students' dissertation acknowledgments. *J Acad Librariansh* 35(3):241–251
8. Ben-Ari E (1987) On acknowledgements in ethnographies. *J Anthropol Res* 43(1):63–84
9. Mantai L, Dowling R (2015) Supporting the phd journey: insights from acknowledgements. *Int J Res Dev* 6(2):106–121
10. Lamers WS, Boyack K, Larivière V, Sugimoto CR, van Eck NJ, Waltman L, Murray D (2021) Measuring disagreement in science. [arXiv:2107.14641](https://arxiv.org/abs/2107.14641)
11. Cohen J (1960) A coefficient of agreement for nominal scales. *Educ Psychol Meas* 20(1):37–46
12. Le QV, Mikolov T (2014) Distributed representations of sentences and documents. [arXiv:1405.4053](https://arxiv.org/abs/1405.4053)
13. Girvan M, Newman ME (2002) Community structure in social and biological networks. *Proc Natl Acad Sci USA* 99(12):7821–7826
14. Alotaibi HS (2018) Metadiscourse in dissertation acknowledgments: exploration of gender differences in efl texts. *Educ Sci: Theory Pract.* 18(4):899–916
15. Casad B, Franks J, Garasky C, Kittleman M, Roesler A, Hall D, Petzel Z (2021) Gender inequality in academia: problems and solutions for women faculty in stem. *J Neurosci Res* 99:13–23. <https://doi.org/10.1002/jnr.24631>
16. Collins R, Steffen-Fluhr N (2019) Hidden patterns: using social network analysis to track career trajectories of women stem faculty. *Equal Divers Incl* 38:265–282. <https://doi.org/10.1108/EDI-09-2017-0183>
17. Newman ML, Groom CJ, Handelman LD, Pennebaker JW (2008) Gender differences in language use: an analysis of 14,000 text samples. *Discourse Process* 45(3):211–236

18. Ishikawa Y (2015) Gender differences in vocabulary use in essay writing by university students. *Proc, Soc Behav Sci* 192:593–600
19. Fortunato S, Bergstrom CT, Börner K, Evans JA, Helbing D, Milojević S, Petersen AM, Radicchi F, Sinatra R, Uzzi B et al (2018) Science of science. *Science* 359(6379):eaao0185
20. Huang J, Gates AJ, Sinatra R, Barabási A-L (2020) Historical comparison of gender inequality in scientific careers across countries and disciplines. *Proc Natl Acad Sci USA* 117(9):4609–4616. <https://doi.org/10.1073/pnas.1914221117>. <https://www.pnas.org/content/117/9/4609.full.pdf>
21. Way SF, Larremore DB, Clauset A (2016) Gender, productivity, and prestige in computer science faculty hiring networks. In: Proceedings of the 25th international conference on world wide web, pp 1169–1179
22. Kozłowski D, Larivière V, Sugimoto CR, Monroe-White T (2022) Intersectional inequalities in science. *Proc Natl Acad Sci USA* 119(2):e2113067119
23. Sinatra R, Wang D, Deville P, Song C, Barabási A-L (2016) Quantifying the evolution of individual scientific impact. *Science* 354(6312):596
24. Wu L, Wang D, Evans JA (2019) Large teams develop and small teams disrupt science and technology. *Nature* 566(7744):378–382
25. Sekara V, Deville P, Ahnert SE, Barabási A-L, Sinatra R, Lehmann S (2018) The chaperone effect in scientific publishing. *Proc Natl Acad Sci USA* 115(50):12603–12607
26. Way SF, Morgan AC, Larremore DB, Clauset A (2019) Productivity, prominence, and the effects of academic environment. *Proc Natl Acad Sci USA* 116(22):10729–10733
27. Ma Y, Mukherjee S, Uzzi B (2020) Mentorship and protégé success in stem fields. *Proc Natl Acad Sci USA* 117(25):14077–14083
28. Hook DW, Porter SJ, Herzog C (2018) Dimensions: building context for search and evaluation. *Front Res Metr Anal* 3:23. <https://doi.org/10.3389/frma.2018.00023>
29. Lee S, Bozeman B (2005) The impact of research collaboration on scientific productivity. *Soc Stud Sci* 35(5):673–702. <https://doi.org/10.1177/0306312705052359>
30. Fox MF, Faver CA (1985) Men, women, and publication productivity: patterns among social work academics. *Sociol Q* 26(4):537–549
31. Larivière V, Ni C, Gingras Y, Cronin B, Sugimoto CR (2013) Bibliometrics: global gender disparities in science. *Nat News* 504(7479):211
32. Abramo G, D'Angelo CA, Caprasecca A (2009) Gender differences in research productivity: a bibliometric analysis of the Italian academic system. *Scientometrics* 79(3):517–539
33. Kyvik S, Teigen M (1996) Child care, research collaboration, and gender differences in scientific productivity. *Sci Technol Human Values* 21(1):54–71
34. Duch J, Zeng XHT, Sales-Pardo M, Radicchi F, Otis S, Woodruff TK, Nunes Amaral LA (2012) The possible role of resource requirements and academic career-choice risk on gender differences in publication rate and impact. *PLoS ONE* 7(12):51332
35. Rossiter MW (1993) The Matthew Matilda effect in science. *Soc Stud Sci* 23(2):325–341
36. Sabharwal M (2013) Comparing research productivity across disciplines and career stages. *J Comp Policy Anal* 15(2):141–163
37. Costas R, Van Leeuwen TN, Bordons M (2010) A bibliometric classificatory approach for the study and assessment of research performance at the individual level: the effects of age on productivity and impact. *J Am Soc Inf Sci Technol* 61(8):1564–1581
38. Bland CJ, Center BA, Finstad DA, Risbey KR, Staples J (2006) The impact of appointment type on the productivity and commitment of full-time faculty in research and doctoral institutions. *J High Educ* 77(1):89–123
39. Sonnert G (1996) Faculty at work: motivation, expectation, satisfaction. *J High Educ* 67(6):716–718. <https://doi.org/10.1080/00221546.1996.11774822>
40. Becher T (1994) The significance of disciplinary differences. *Stud High Educ* 19(2):151–161
41. Gallagher RJ, Frank MR, Mitchell L, Schwartz AJ, Reagan AJ, Danforth CM, Dodds PS (2021) Generalized word shift graphs: a method for visualizing and explaining pairwise comparisons between texts. *EPJ Data Sci* 10(1):4
42. Emmons R, McCullough ME (2003) Counting blessings versus burdens: an experimental investigation of gratitude and subjective well-being in daily life. *J Pers Soc Psychol* 84(2):377–389
43. Killen A, Macaskill A (2015) Using a gratitude intervention to enhance well-being in older adults. *J Happ Stud* 16(4):947–964
44. Larivière V, Gingras Y, Sugimoto CR, Tsou A (2015) Team size matters: collaboration and scientific impact since 1900. *J Assoc Inf Sci Technol* 66(7):1323–1332
45. Posel DR, Casale DM (2011) Relative standing and subjective well-being in South Africa: the role of perceptions, expectations and income mobility. *Soc Indic Res* 104(2):195–223
46. De la Garza AG, Mastrobuoni G, Sannabe A, Yamada K (2012) The relative utility hypothesis with and without self-reported reference wages
47. Lifchits G, Anderson A, Goldstein DG, Hofman JM, Watts DJ (2021) Success stories cause false beliefs about success. *Judgm Decis Mak* 16(6):1440