Gender Differences in Science and Engineering: A Data Mining Approach

Shivangi Chopra, Melicaalsadat Mirsafian, Abeer Khan, Lukasz Golab
University of Waterloo
Waterloo, Ontario, Canada
{s9chopra,mmirsafi,a383khan,lgolab}@uwaterloo.ca

ABSTRACT

In this paper, we describe a data-intensive approach to study gender differences in Science, Technology, Engineering and Mathematics (STEM). We apply deep learning, text mining and statistical methods to unique academic datasets, including undergraduate admission data, co-operative job descriptions and student entrepreneurship data. Our results show that women have different reasons than men for applying to an engineering program, that women tend to fill slightly different co-operative positions during their undergraduate studies, and that women are less likely to be interested in entrepreneurial activities.

KEYWORDS

data mining, STEM education, gender gap, STEM pipeline

1 INTRODUCTION

The gender gap in Science, Technology, Engineering and Mathematics (STEM) is well documented. For example, according to a report on the gender distribution in STEM employment in Canada[1], as of 2016, only 20% of the employees are women. Numerous studies have considered the different steps in the STEM educational and professional pipeline, from high school to post-secondary education and beyond, to understand why fewer women enrol in STEM programs and pursue STEM careers [17, 21, 35, 38]. A major focus has been to identify retention problems, or “leaks” in the pipeline [9].

Figure 1 illustrates the STEM educational and professional pipeline, along with the gender issues that have been studied (details in Section 2). Background refers to primary and secondary education, where differences in interests and aptitude have been studied. We then divide undergraduate education into classroom learning and work-integrated learning, also known as co-operative education, which is now part of many science and engineering programs worldwide (co-operative programs include both on-campus study terms and co-operative work terms). Here, gender differences have mainly been studied in the context of satisfaction with the academic and work environments. Finally, in the context of post-graduate careers, there have been various studies investigating career preferences, workplace experiences and biases, advancement opportunities, and salary differences.

In this paper, we present our data-intensive research to understand gender differences in STEM, including the methods we used and the insights we have obtained. Using real datasets from a large North American undergraduate institution, combined with deep learning, text mining and statistical methods, our goal is to measure the gender gap and suggest how to close it. Access to unique datasets combined with state-of-the-art data science methods allows us to obtain new insights compared to previous work. Our contributions and the paper outline are as follows.

- In Section 3, we summarize our recent work [10] on text mining of undergraduate admission data. We use question-answering methods, word embeddings and text clustering to understand the differences in the *reasons* why young men and women want to study engineering.
- In Section 4, we present new research on gender differences in undergraduate co-operative education. Using topic modelling and document clustering, we analyze the differences in the descriptions of co-operative jobs held by male and female students. This is an important problem in the context of gender studies because, as noted by Kauhenen et al. [23], these early work experiences can greatly affect subsequent career choices.
- Section 5 builds on our recent statistical analysis of student entrepreneurship [3], and presents insights on gender differences in entrepreneurship interests and outcomes.
- Finally, we outline directions for future data-intensive work on gender differences in Section 6.

2 RELATED WORK

We start by reviewing prior work on data analysis (usually statistical analysis including distributions, regression, and ANOVA) to study gender issues in STEM education and STEM careers.

In the context of differences in interests, Sadler et al. [33] noticed that men’s interest in engineering was stable over their high school years but women’s interest declined near graduation. Some work suggests that men gravitate toward things-oriented careers and women towards people-oriented careers, even within STEM. Further, women show more artistic and social interests [1, 36]. Some studies show that mathematical abilities are not sufficient to encourage more interest in STEM [32], using methods such as clustering and association rule mining for their analysis. Kauhenan et al. [23] noted that individuals with high mathematical and verbal abilities preferred non-STEM careers while those with high mathematical but moderate verbal ability were more likely to pursue STEM. Some work indicates that women’s lack of interest may be related to the perceived mismatch of STEM careers with their career goals [12, 40]. Finally, Bystydzienski et al. [7] found that an intervention program targeting high-achieving female high school students helped develop an interest in engineering, but some participants decided against pursuing it due to lack of financial and social support, and fears of failure.

Some works investigate student experiences with STEM education. Amelink et al. [2] noted that perceptions of being respected by course instructors positively influence students’ intent to continue in engineering studies and also engineering careers in the case of female students. Espinosa et al. [14] found that women of colour who were academically engaged outside of the classroom, had altruistic ambitions, and attended institutions that

[1] http://wiseatlantic.ca/wp-content/uploads/2018/03/WISEReport2017_final.pdf

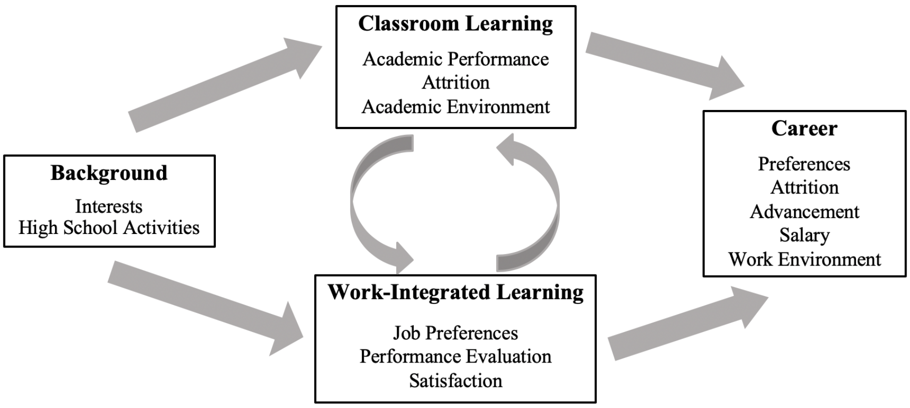

Figure 1: The STEM pipeline

were not highly selective with a robust student community, were more likely to persist in STEM. Griffith et al. [18] found no evidence that having more female faculty members increases the likelihood of women's persistence. Rosenthal et al. [31] found that single-sex programs within STEM helped women achieve a greater sense of belonging at their university, due to the perceived identity compatibility and perceived support derived from these programs. Some qualitative work found that women may face negative experiences in school, in the form of implicit or overt bias from their professors or peers [34].

There is a wealth of literature on understanding men's and women's STEM careers, including hiring practices, workplace evaluations and attrition.

There are conflicting reports on gender differences in hiring. Some show a bias in favour of women when hiring teachers or university faculty [6, 41]. Ceci et al. [8] found that women were preferred over identically qualified men, but not over better qualified men, for tenure track assistant professorship. Others find a bias towards hiring men [27, 30] for laboratory manager or other positions. Moss et al. [27] further found that female applicants were given lower starting salaries and less mentoring by the hiring faculty.

Some works show a gender difference in salaries, with female professors receiving lower salaries than male professors [5], even with equal likelihood of negotiation [28]. Hu et al. [20] discovered that men who were academically engaged during college, and women who were socially engaged, had better early career earnings. Berheide et al. [5] also found that among the associate professors who served as department or program chairs, women were promoted a year later on average. Focus groups further revealed that a lack of feedback and mentoring decreased the likelihood of women applying for promotion to full professor.

Workplace evaluations also show gender differences [24]. Reilly et al. [29] studied workplace evaluations and advice given to technology interns experiencing difficulties in the workplace. They found that women with ability issues were viewed as having lower field aptitude than men with ability issues, when judged by individuals holding both hostile and benevolent sexist beliefs. Men and women with interpersonal issues had similar aptitude ratings, but men were dissuaded from seeking help while women were expected to find mentors and control their emotions. Dutt et al. [13] conducted text analysis on recommendation letters and discovered that female applicants are only half as likely to receive excellent letters versus good letters compared to male applicants. Male and female evaluators were equally likely to display this bias. Lee et al. [25] studied how entrepreneurial ventures (and the entrepreneurs themselves) are evaluated by venture capitalists. They found that women without technical backgrounds were evaluated as having less leadership ability than similar men. They also received less capital investment than technical women, technical men, and non-technical men. Terrel et al. [37] found that on the open source software website Github, women's contributions tend to be accepted more often than men's, but for contributors whose gender is identifiable and who are outsiders to a project, men's acceptance rates are higher. The authors suggest that although women on GitHub may be more competent overall, bias against them may exist nonetheless. On the other hand, Van et al. [39] examined applications for a research grant by early career researchers in STEM disciplines in the Netherlands, and found that men and women received similar evaluations and had similar success rates.

There is also some qualitative research on other aspects of the workplace environment. Thakkar et al. [38] found that in India, although computer science was gender balanced at the bachelor's level, marriage and childcare norms, family influence, and finances drove women away the field at later stages. Some work found overt and implicit sexism, gendered expectations and a lack of professionalism as some of the challenges women face in the STEM workplace [16, 34, 35].

Several works observed that the attrition rate for women in STEM is higher than for men. Some find that family related constraints are not the primary reason for this [17, 21]. Hunt et al. [21] observed that dissatisfaction over pay and promotion opportunities is the main problem, with working conditions, unavailability of a job in the field, changes in professional interests, and job location playing statistically significant but secondary roles. Glass et al. [17] found that having an advanced degree increases the odds of women leaving STEM employment, suggesting that the STEM jobs held by advanced-degree holders are less satisfying than those held by bachelor's degree recipients. Kaminski et al. [22] found comparable retention rates for men and women among science and engineering faculty members, but higher attrition rates of women in mathematics. Both qualitative and quantitative work has found that women who *do* stay in engineering receive better workplace support [4, 15].

In contrast to prior work, we use unique datasets and state-of-the-art data science methods to obtain new insight into the reasons why men and women want to study engineering, the co-operative jobs held by undergraduate male and female students, and the gender differences in student entrepreneurs.

3 GENDER DIFFERENCES IN ENGINEERING APPLICANTS

3.1 Motivation

It is well known that women are underrepresented in STEM degrees: only 23% of women with high mathematics scores pursue STEM degrees compared to 45% of men with the same scores [19]. To understand why this is the case, we analyze gender differences in high school backgrounds and engineering interests of undergraduate engineering applicants. While most of the previous work on this subject has been longitudinal or survey-based, we obtain new insights using deep learning methods on a large admissions dataset.

3.2 Data and Method

Data: We analyzed over 30,000 applications – both accepted and rejected – to undergraduate engineering programs at a large North American institution from 2013 to 2016 inclusive. In their applications, prospective students describe why they are interested in engineering, and they provide background information including their reading interests, extracurricular activities, jobs they held throughout high school, programming experience (only for the Software Engineering program), and any additional information. By mining the responses to these questions, our goal was to determine whether female applicants express different reasons for applying to an engineering program, and whether female applicants have different technical and extracurricular backgrounds.

The engineering programs included in our analysis are Environmental, Biomedical, Chemical, System Design, Management, Civil, Geological, Nanotechnology, Electrical, Computer, Software, Mechanical, and Mechatronics Engineering, listed in the descending order of proportion of female students' applications in the program. In our study, we consider Environmental and Biomedical Engineering together, referred to as BEE, as the two gender balanced programs. We consider Software Engineering separately, referred to as SE, because of its unique requirement to describe the applicant's programming experience, and we consider all the other engineering programs in a single group we call OTHER.

Method: We developed a text mining method to identify the *reasons* why students apply to engineering programs based on their responses. As in other text mining applications, challenges arise due to the ambiguity of natural language. To overcome these challenges, we used word embeddings and text clustering to partition the responses into semantically meaningful groups, each group corresponding to a potential reason for applying to an engineering program. We also analyze gender differences in programming languages and extracurricular activities through classification models and word frequency analysis.

We start by entering each student's response to the question "Why are you interested in engineering?" into an open source Question Answering (QA) API. The QA method uses neural networks to extract a set of sentences within the response that best match the question. We then derive vector representations for these sentences using a Word2Vec word embedding model [26] trained on the Google News corpus. These vector representations are such that two sentences that are semantically similar have similar vectors. This enables us to cluster the sentences from all applicants into a set of ten semantically meaningful groups corresponding to different reasons. For this purpose, we use a combination of K-means clustering and Card-sorting [42]. Finally, we identify reasons that were mentioned statistically significantly more by women or men.

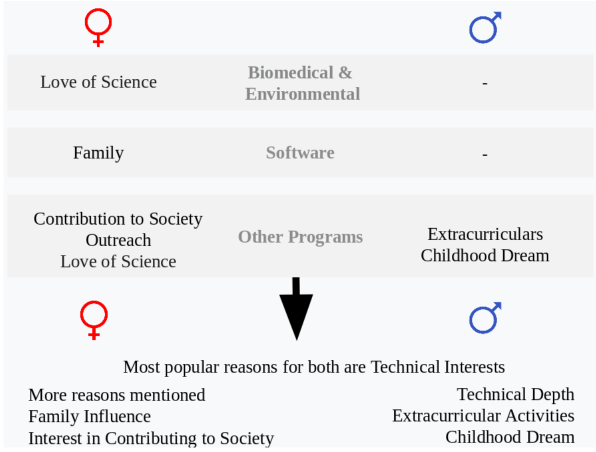

Figure 2: Gender differences in reasons for applying to engineering

We also use word frequency analysis on responses to questions regarding engineering interests and goals, extracurricular activities, job experience, reading interests, programming experience, and additional information. Subsequently, we identify words mentioned statistically significantly more by women or by men. We also show Venn diagrams to illustrate the overlap among the top 100 frequent words used by men and women. Each word is *stemmed* for frequency analysis to ensure that similar words (such as "challenge" and "challenging") are counted together.

3.3 Results

We summarize the main results below and refer the reader to [10] for details of our method and findings.

3.3.1 Reasons for Applying to Engineering. When describing why they want to study engineering, men mention more technical words such as "compute", "problem", "system" and "robot". Women, while using technical words such as "science", "chemical", also use words like "people", "improve" and "health". Furthermore, we identified ten common reasons for applying to engineering: Family Influence, Contribution to Society, Outreach, Technical Interests, Love of Science, Extracurriculars, Prior Accomplishments, High School, Professional Development, and Childhood Dream. As summarized in Figure 2, depending on the program, we found that women tend to mention Contribution to Society, Family Influence, and Love of Science significantly more than men, while men mention Extracurriculars, and Childhood Dream significantly more than women. Overall, the most popular reasons are Technical Interests, Love of Science, and Professional Development.

3.3.2 Reading Interests. More men report reading technical content such as research papers, while more women report reading material with societal focus. Words chiefly mentioned by

Figure 3: Overlap between the top 100 most frequent words used by SE men and women to describe their reading interests

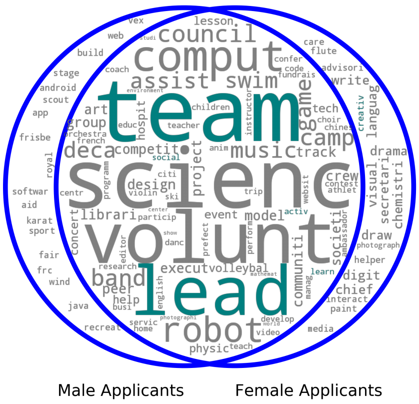

Figure 4: Overlap between the top 100 most frequent words used by SE men and women to describe their extracurricular activities

men include "article", "enjoy", "compute", and "science". Words predominantly mentioned by women include "love", "character", "women", "people", and "family". Figure 3 shows that reading interests of SE men and women include "world", "impact", and "novel", with men mentioning more "scientific", "research" and "theory" and women mentioning more "scientists" and "literature".

3.3.3 Extracurricular Activities. Male applicants' extracurricular activities tend to display a technical focus, and female applicants list a wide breadth of experience ranging from leadership to artistic pursuits. More men in different groups of programs mention "robot", "coach", and "compute", while more women mention "dance", "art", "council", "volunteer", and "lead". Figure 4 visualizes these differences for SE applicants.

3.3.4 Job Titles. When describing jobs students held throughout high school, men were more likely to mention terms that imply technical work or manual labour, whereas women were more likely to mention terms that imply customer service or caring professions. Example words in job titles for men are "referee",

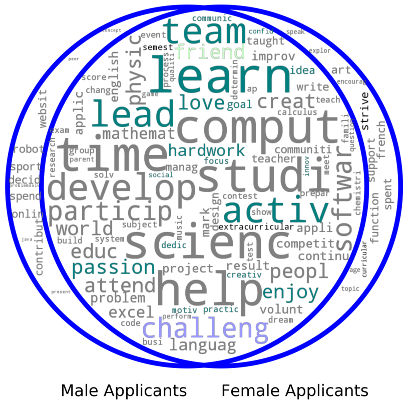

Figure 5: Overlap between the top 100 most frequent words used by SE men and women to describe additional information about themselves

"labor", and "technician". Example words for job titles for women are "cashier", "teacher", and "assist".

3.3.5 Programming Experience. In general, more women use more non-technical terms, and men use more technical terms. Words more commonly used by men include "game" and "develop", while words used more commonly by women include "mark", and "attend". Through manual inspection, we discovered that "mark" referred to earning a mark in a course and "attend" referred to attending a programming workshop or event.

3.3.6 Additional Information. We see a difference in word choice between men and women when answering a question with no restrictions on the content of the answer. Words more commonly used by men include "sport" and "compute", while words used more commonly by women include "community", and "art" (see Figure 5 for differences in SE applicants).

3.4 Insights

3.4.1 Similarities. Regardless of gender, the most commonly mentioned reason in response to "Why are you interested in engineering?" is Technical Interests. Furthermore, in SE applicants, we do not see a large gender gap in self-reported programming experience, or the number of languages known. In BEE, which is the most gender balanced group, the differences are minimal.

3.4.2 Differences. We find that men differentiate themselves through depth of experience, and women through breadth of experience. To study engineering, all applicants must demonstrate a strong background in science and mathematics through their academic work. We still see men highlight their interest in acquiring more technical skills through their writing, while female applicants mention a wider variety of topics in response to all the questions on the application form.

Women mention personal and family influences in their decision to study engineering, especially in SE where women mention it significantly more than men. Women also show a stronger desire to contribute to society and improve the world. This is evident in the OTHER group of programs where women are more likely to mention Contribution to Society, and in BEE and SE where women mention words such as "health", "improve" and "people".

We infer that to attract more women to study engineering, it must be presented as a profession that can help others and allow for a broad range of careers and learning opportunities. We believe that the message to women should not be just that they can do it, but that they should want to do it because engineering is an excellent fit for their values and priorities. Furthermore, our results suggest that in order to retain female students, engineering curricula should emphasize real-life applications and the impact of technology on society in first-year courses.

4 GENDER DIFFERENCES IN CO-OPERATIVE WORK PLACEMENTS

4.1 Motivation

Having analyzed engineering applicants, we now turn to gender differences in work-integrated learning. For many undergraduate STEM students, co-operative jobs represent their first STEM work experience, which can affect their future career choices. This motivates our study to determine if male and female students work in different types of co-operative jobs.

4.2 Data and Method

Data: We study over 17,000 co-operative jobs filled by undergraduate students from the same institution as described in the previous section in 2014. The corresponding data includes a textual job description that was created by the employer, the industry the job belongs to, and the gender of the student who obtained the job. We report results for the biggest industries in the dataset: Information Technology (IT), various Engineering fields such as Mechanical, Electrical and Chemical, Finance, Environmental Studies, Arts and Biology.

Method: Job descriptions are free text and are written directly by employers. As a result, they are not standardized or well-structured. In our prior work on job description mining [11], we developed a method to extract informative terms from job descriptions: technical skills, soft skills, perks (e.g., free food or proximity to public transit) and other terms indicating the nature of the job. To do so, we remove common English words, common misspellings, common abbreviations, and common formatting and administrative content (e.g., links to company websites, contact names and emails, timestamps and addresses). At the end of this process, each job description is converted to a binary *document term vector* indicating the presence of the (stemmed) words that were not removed.

Note that we use binary document term vectors instead of word frequency vectors. In contrast to other types of documents, job descriptions do not repeat words for emphasis. For example, the desired skills are usually only listed once.

Next, for each industry, we cluster the document term vectors to identify common types of jobs. We use the same method as in our prior work on job description mining [11]: we apply Latent Semantic Analysis (LSA) to reduce the dimensionality of the data followed by K-means clustering. To represent each cluster, we extract the most significant terms from its cluster centroid. After examining these terms, we manually assign a label to each cluster corresponding to the most likely type of job corresponding to the cluster (e.g., frequent occurrences of terms such as “HTML” and “javascript” could indicate a cluster of Web programming jobs).

Table 1: Largest clusters of IT jobs

Label	Words in cluster centroid	%All	%Higher	M/F
Start-up Culture	python, code, featur, java, scalabl, passion, languag, startup, javascript, web	21%	0%	
Programming	c#, sql, oop, java, server, net, test, code, languag, web	19%	1%	M
Web Development	javascript, css, html, web, jqueri, html5, php, framework, mysql, sql	13%	1%	M
Mobile Development	android, io, mobil, app, java, platform, iphon, devic, agil, c	13%	2%	M
Embedded Systems	c++, c, embed, hardwar, devic, debug, linux, languag, java, softwar	11%	4%	M
Business Analyst	analyst, sql, financi, bank, trade, c#, net, solut, invest, busi	11%	11%	F
System Administrator	hardwar, troubleshoot, configur, instal, server, network, desktop, user, resolut, deploy	7%	1%	M

Finally, for each industry, we calculate the percentage of jobs held by men and women in each cluster (job type) and we compare it to the proportion of men and women enrolled in the corresponding academic programs. If a cluster has a higher percentage of men (or women) than the underlying student population, we report which gender has a higher percentage and by how much. We do this only for the largest clusters to avoid drawing conclusions from small samples.

4.3 Results

We start with IT jobs, 86% of which were filled by male students. Table 1 shows the 7 largest clusters of IT jobs sorted by size; the remaining three clusters had under 2% of the total number of jobs each. Each row includes our manually-assigned label, the ten most frequent words in the cluster centroid, the percentage of jobs in this cluster out of all IT jobs, and a percentage difference of men or women having jobs in this cluster compared to the gender distribution among all students in IT programs such as Computer Science and Software Engineering. There is a negligible difference (<2%) between men and women in programming, web development, mobile development and system administrator jobs. Additionally, both men and women work equally at jobs that appear to be at technology startups. However, there is a difference in the Embedded Systems cluster where 4% more men work with the hardware and software of embedded devices; the men-women ratio in the Embedded Systems cluster is 90%-10%. Similarly, in comparison to men, 11% more women work as business analysts in the IT industry (with a men-women ratio of 75%-25%).

Table 2 shows the job clusters in the Finance industry, where half the students are men and half are women in our dataset. There are more men in financial analysis, trade, and accounting profiles, but more women in financial documentation, actuarial jobs and taxation/auditing.

Next, Table 3 shows the job clusters in Health Studies, where 68% of students are women. It appears that more men are involved in research, but more women are involved in organizing recreational and therapeutic camps for seniors and patient care.

Finally, Table 4 summarizes our results for the remaining large industries. Instead of listing all the cluster details, the table only shows the types of jobs (derived from manual labels of the clusters) which either have no difference between men and women, have more women, or have more men. The industries are sorted by a decreasing proportion of females, with Arts having 75%

Table 2: Largest clusters of Finance jobs

Label	Tokens in cluster centroid	%All	%Higher	M/F
Analyst	financi, account, analyt, busi, analyst, document, report, data, initi, financ	30%	4%	M
Trade	financ, bank, capit, risk, invest, credit, financi, deriv, riskmanag, trade	17%	9%	M
Financial Documentation	tax, bookkeep, audit, statement, incom, account, charter, file, prepar, compil	11%	9%	F
Actuarial	actuari, insur, price, exam, casualti, underwrit, reserv, valuat, financi, statist	10%	3%	F
Accounting	payabl, reconcili, account, financ, statement, invoic, journal, bank, financi, ledger	10%	3%	M
Tax Audit	audit, tax, advisori, econom, account, transcript, cpa, financi, humil, statement	9%	13%	F

Table 3: Largest clusters of Health Studies jobs

Label	Tokens in cluster centroid	%All	%Higher	M/F
Organizing Community Events	event, arrang, advertis, health, recreat, promot, communiti, organ, customerservic, educ	33%	3%	M
Research	ergonom, kinesiolog, health, literatur, statist, biomechan, conduct, assess, review, care	25%	5%	M
Therapy	physiotherapist, modal, exercis, patient, clinic, physiotherapi, treatment, rehabilit, therapi, injuri	13%	2%	M
Geriatrics	leisur, recreat, therapeut, intervent, therapi, care, cognit, health, adult, elder	9%	24%	F
Patient Care	cancer, patient, clinic, clinician, outpati, care, health, multidisciplin, journal, literatur	8%	7%	F

females and Mechanical Engineering having 11% females. Even though Arts and Biology are dominated by female students, some of the technical jobs in these fields have more men. Environmental Studies and Civil Engineering jobs show interesting differences, with men carrying out more site work. Chemical and Electrical Engineering jobs show that men and women tend to have different areas of technical work in these fields. Furthermore, different types of Mechanical Engineering jobs are held by very similar proportions of men and women yet this program has the smallest fraction of female students.

4.4 Insights

In most industries, it appears that there are some differences in the types of co-operative jobs held by men and women. Our results should be of interest to co-op employers wishing to diversify their workforce. An interesting direction for future work is to conduct interviews with a sample of co-operative students to find out more about their job search strategies. Are the differences we saw due to nature, social conditioning or a combination of factors?

5 GENDER DIFFERENCES IN ENTREPRENEURSHIP

5.1 Motivation

Finally, we examine gender differences from an entrepreneurship standpoint. Entrepreneurship can lead to job creation, innovation, and economic growth. As a result, there has been private and public emphasis on fostering entrepreneurship: examples include tax credits and establishing supporting entities such as startup

Table 4: Labels of Clusters with more men/women

Industry	No Difference	Higher %F	Higher %M
Arts	Project management, Media campaign, Legal help	Customer service	Multimedia content creation
Biology	Organizing events, Laboratory research	Healthcare	Microbiology
Environmental Studies	-	Project Management, Water, Urban planning	Field visit, Geographic Information Systems
Civil Engineering	-	Software modelling of architecture and interior	Project management, Site visit, Inspection
Chemical Engineering	-	Process improvement	Energy, Laboratory research
Electrical Engineering	Web development	Embedded systems, Circuit testing, Power	Mobile development, System administrator, Circuit design
Mechanical Engineering	Project management, Mechanical drawing, Software development, Manufacturing	Simulations	-

incubators which are often paired with universities. Our goal is to find out, using data analysis, whether there is a gender gap in entrepreneurial interests and outcomes of undergraduate students. We want to determine whether more men or women take advantage of entrepreneurial resources offered by the university, and, ultimately, whether more men or women are involved in creating startup companies.

5.2 Data and Method

Data: We used two datasets for this analysis, again, from the same institution as in the other analyses. First, we obtained the gender of each engineering student who took advantage of entrepreneurial resources offered by the university - either by taking an (optional) entrepreneurship course or by working at their own startup during a co-operative workterm (an option provided by the university to promote entrepreneurship). Second, we obtained the names and gender of 221 students who graduated from this university with an engineering degree between 2006 and 2015 and who were involved in creating at least one company.

Method: We use simple statistical methods for this analysis: we calculate and compare the fraction of men and women who took an entrepreneurship course, who worked on their own business during a co-operative work term, and who were involved in creating a startup.

5.3 Results

Entrepreneurship Courses: overall, 1965 undergraduate engineering students took at least one entrepreneurship course between 2006 and 2015. We note that these courses are not mandatory for engineering students, so we interpret enrolment in such a course as an indication of interest in entrepreneurship. Out of the 1965 students, 253 were females (i.e., about 13%). In contrast, 22% of the engineering enrollment was female, indicating that women are not electing to take entrepreneurship courses as much as men. Both men and women took these courses in their senior years.

Working in own company for a co-op work term: 139 students took this option, out of whom 12 were females (i.e., about 9%). Again, in contrast to the 22% of women enrolled in engineering, this is a low proportion.

Finally, we zoom in on the 221 student entrepreneurs who obtained an engineering degree. Only 12 of them are female (i.e., about 5%). These 221 students were involved in 242 startups, of which only 15 were started by women. Furthermore, out of the 19 "serial entrepreneurs" who started more than one company, only two were women, one of whom was involved in two companies and the other with three companies.

5.4 Insights

The main insight from this analysis is that fewer women than men in our sample chose to become entrepreneurs. Additionally, female students were less likely to take advantage of entrepreneurial resources such as taking entrepreneurship courses or spending a co-operative workterm at their own startup. Thus, one way to help close the gender gap in entrepreneurship could be to invite successful female entrepreneurs to give talks and workshops.

6 CONCLUSIONS AND FUTURE WORK

In this paper, we applied data analysis methods to study gender differences at various stages in the STEM pipeline: at undergraduate admission time, during undergraduate education (focusing on work-integrated learning) and in the context of entrepreneurship. Combining unique datasets with various deep learning, text mining and statistical methods allowed us to obtain new insights into the reasons why women want to study engineering, the types of co-operative jobs filled by women, and the gender difference in entrepreneurial activities and outcomes.

There is much more data-driven work that can be done to measure and close the gender gap in STEM. Below, we list several directions for future work.

- **Analysis of high school influence:** Do some high schools produce more successful engineering female applicants than others? If yes, why? Do those high schools have more female teachers/role models? Do they host more outreach programs? This can be addressed by combining high school data with university admissions data.
- **Analysis of classroom learning:** What kinds of courses do females choose to take? Do women switch out of engineering programs? What are the common issues that women face on campus in their undergraduate careers? This can be analyzed by combining academic records and discussions on social media channels (e.g., Reddit)
- **Analysis of work-integrated learning:** Do women receive equal opportunity in co-operative education? Do women receive equal workplace evaluations? Do women prefer certain jobs? Are women satisfied with their work experience? This can be addressed by analyzing data from a co-operative education system.
- **Analysis of career paths:** Do men and women have different career paths and opportunities? This can be addressed by mining LinkedIn data. Analyzing admission forms of women applying for Master's and Doctoral studies might also provide additional insight beyond our undergraduate admission analysis.
- **Attrition analysis:** At which point in the academic education system do we lose qualified women and what is the cause for this loss?

REFERENCES

[1] Sigal Alon and Thomas A DiPrete. 2015. Gender differences in the formation of a field of study choice set. *Sociological science* 2 (2015), 50.
[2] Catherine T Amelink and Elizabeth G Creamer. 2010. Gender differences in elements of the undergraduate experience that influence satisfaction with the engineering major and the intent to pursue engineering as a career. *Journal of Engineering Education* 99, 1 (2010), 81–92.
[3] Andrew Andrade, Shivangi Chopra, Bahra Nurlybayev, and Lukasz Golab. 2018. Quantifying the Impact of Entrepreneurship on Cooperative Education Job Creation. *International Journal of Work-Integrated Learning* 19, 1 (2018), 51–68.
[4] Mary Ayre, Julie Mills, and Judith Gill. 2013. Yes, I do belong: The women who stay in engineering. *Engineering studies* 5, 3 (2013), 216–232.
[5] Catherine White Berheide, Lisa Christenson, Rena Linden, and Una Bray. 2013. Gender Differences in Promotion Experiences at Two Elite Private Liberal Arts Colleges in the United States.. In *Forum on Public Policy Online*, Vol. 2013. ERIC.
[6] Thomas Breda and Mélina Hillion. 2016. Teaching accreditation exams reveal grading biases favor women in male-dominated disciplines in France. *Science* 353, 6298 (2016), 474–478.
[7] Jill M Bystydzienski, Margaret Eisenhart, and Monica Bruning. 2015. High school is not too late: Developing girls' interest and engagement in engineering careers. *The Career Development Quarterly* 63, 1 (2015), 88–95.
[8] Stephen J Ceci and Wendy M Williams. 2015. Women have substantial advantage in STEM faculty hiring, except when competing against more-accomplished men. *Frontiers in psychology* 6 (2015), 1532.
[9] Mitchell J Chang, Jessica Sharkness, Sylvia Hurtado, and Christopher B Newman. 2014. What matters in college for retaining aspiring scientists and engineers from underrepresented racial groups. *Journal of Research in Science Teaching* 51, 5 (2014), 555–580.
[10] Shivangi Chopra, Hannah Gautreau, Abeer Khan, Melicaalsadat Mirsafian, and Lukasz Golab. 2018. Gender Differences in Undergraduate Engineering Applicants: A Text Mining Approach. In *Proceedings of the 11th International Conference on Educational Data Mining, EDM 2018, Buffalo, NY, USA, July 15-18, 2018*. 44–54.
[11] Shivangi Chopra and Lukasz Golab. 2018. Job Description Mining to Understand Work-Integrated Learning. In *Proceedings of the 11th International Conference on Educational Data Mining, EDM 2018, Buffalo, NY, USA, July 15-18, 2018*. 32–43.
[12] Amanda B Diekman, Elizabeth R Brown, Amanda M Johnston, and Emily K Clark. 2010. Seeking congruity between goals and roles: A new look at why women opt out of science, technology, engineering, and mathematics careers. *Psychological Science* 21, 8 (2010), 1051–1057.
[13] Kuheli Dutt, Danielle L Pfaff, Ariel F Bernstein, Joseph S Dillard, and Caryn J Block. 2016. Gender differences in recommendation letters for postdoctoral fellowships in geoscience. *Nature Geoscience* 9, 11 (2016), 805.
[14] Lorelle Espinosa. 2011. Pipelines and pathways: Women of color in undergraduate STEM majors and the college experiences that contribute to persistence. *Harvard Educational Review* 81, 2 (2011), 209–241.
[15] Nadya A Fouad, Romila Singh, Kevin Cappaert, Wen-hsin Chang, and Min Wan. 2016. Comparison of women engineers who persist in or depart from engineering. *Journal of Vocational Behavior* 92 (2016), 79–93.
[16] Susan K Gardner and Amy Blackstone. 2013. Putting in your time: Faculty Experiences in the Process of Promotion to Professor. *Innovative Higher Education* 38, 5 (2013), 411–425.
[17] Jennifer L Glass, Sharon Sassler, Yael Levitte, and Katherine M Michelmore. 2013. What's so special about STEM? A comparison of women's retention in STEM and professional occupations. *Social forces* 92, 2 (2013), 723–756.
[18] Amanda L Griffith. 2010. Persistence of women and minorities in STEM field majors: Is it the school that matters? *Economics of Education Review* 29, 6 (2010), 911–922.
[19] Darcy Hango. 2013. Gender Differences in science, technology, engineering, mathematics, and computer science (STEM) programs at university. *Insights on Canadian Society* (12 2013).
[20] Shouping Hu and Gregory C Wolniak. 2013. College student engagement and early career earnings: Differences by gender, race/ethnicity, and academic preparation. *The Review of Higher Education* 36, 2 (2013), 211–233.
[21] Jennifer Hunt. 2016. Why do women leave science and engineering? *ILR Review* 69, 1 (2016), 199–226.
[22] Deborah Kaminski and Cheryl Geisler. 2012. Survival analysis of faculty retention in science and engineering by gender. *science* 335, 6070 (2012), 864–866.
[23] Antti Kauhanen and Sami Napari. 2011. *Gender differences in careers.* Technical Report. ETLA discussion paper.
[24] Silvia Knobloch-Westerwick, Carroll J Glynn, and Michael Huge. 2013. The Matilda effect in science communication: an experiment on gender bias in publication quality perceptions and collaboration interest. *Science Communication* 35, 5 (2013), 603–625.

[25] Matthew Lee and Laura Huang. 2018. Gender bias, social impact framing, and evaluation of entrepreneurial ventures. *Organization Science* 29, 1 (2018), 1–16.

[26] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*. 3111–3119.

[27] Corinne A Moss-Racusin, John F Dovidio, Victoria L Brescoll, Mark J Graham, and Jo Handelsman. 2012. Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences* 109, 41 (2012), 16474–16479.

[28] Grace Panther, Kacey Beddoes, and Cheryl Llewellyn. 2018. Salary Negotiations and Gender in Engineering Education. (2018).

[29] Erin D Reilly, Kadie R Rackley, and Germine H Awad. 2017. Perceptions of male and female STEM aptitude: The moderating effect of benevolent and hostile sexism. *Journal of Career Development* 44, 2 (2017), 159–173.

[30] Ernesto Reuben, Paola Sapienza, and Luigi Zingales. 2014. How stereotypes impair women's careers in science. *Proceedings of the National Academy of Sciences* (2014), 201314788.

[31] Lisa Rosenthal, Bonita London, Sheri Robin Levy, and Marci Lobel. 2011. The roles of perceived identity compatibility and social support for women in a single-sex STEM program at a co-educational university. *Sex Roles* 65, 9-10 (2011), 725–736.

[32] Mirka Saarela and Tommi Kärkkäinen. 2014. Discovering gender-specific knowledge from Finnish basic education using PISA scale indices. In *Proceedings of the 7th International Conference on Educational Data Mining, EDM 2014, ISBN 978-0-9839525-4-1*. International Educational Data Mining Society (IEDMS).

[33] Philip M Sadler, Gerhard Sonnert, Zahra Hazari, and Robert Tai. 2012. Stability and volatility of STEM career interest in high school: A gender study. *Science education* 96, 3 (2012), 411–427.

[34] Carroll Seron, Susan S Silbey, Erin Cech, and Brian Rubineau. 2016. Persistence is cultural: Professional socialization and the reproduction of sex segregation. *Work and Occupations* 43, 2 (2016), 178–214.

[35] Kathleen N Smith and Joy Gaston Gayles. 2018. Girl Power: Gendered Academic and Workplace Experiences of College Women in Engineering. *Social Sciences* 7, 1 (2018), 11.

[36] Rong Su, James Rounds, and Patrick Ian Armstrong. 2009. Men and things, women and people: a meta-analysis of sex differences in interests. *Psychological bulletin* 135, 6 (2009), 859.

[37] Josh Terrell, Andrew Kofink, Justin Middleton, Clarissa Rainear, Emerson Murphy-Hill, Chris Parnin, and Jon Stallings. 2017. Gender differences and bias in open source: Pull request acceptance of women versus men. *PeerJ Computer Science* 3 (2017), e111.

[38] Divy Thakkar, Nithya Sambasivan, Purva Kulkarni, Pratap Kalenahalli Sudarshan, and Kentaro Toyama. 2018. The Unexpected Entry and Exodus of Women in Computing and HCI in India. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 352.

[39] Romy Van der Lee and Naomi Ellemers. 2015. Gender contributes to personal research funding success in The Netherlands. *Proceedings of the National Academy of Sciences* 112, 40 (2015), 12349–12353.

[40] Ming-Te Wang, Jessica Degol, and Feifei Ye. 2015. Math achievement is important, but task values are critical, too: examining the intellectual and motivational factors leading to gender disparities in STEM careers. *Frontiers in psychology* 6 (2015), 36.

[41] Wendy M Williams and Stephen J Ceci. 2015. National hiring experiments reveal 2: 1 faculty preference for women on STEM tenure track. *Proceedings of the National Academy of Sciences* (2015), 201418878.

[42] Thomas Zimmermann. 2016. Card-sorting: From text to themes. In *Perspectives on Data Science for Software Engineering*. Elsevier Science, 137–141.