Student Success in Co-operative Education: An Analysis of Job Postings and Performance Evaluations

by

Veronica Salm

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Mathematics
in
Computer Science

Waterloo, Ontario, Canada, 2023

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Co-operative education (co-op) programs combine coursework and work internships and have become popular worldwide. In this analysis, we use two separate co-op datasets to understand employer expectations and factors that contribute to student success.

First, we analyze over 13000 unique filled job postings from work terms in 2021. We group skills using k-means analysis and frequency counting to characterize the types of co-op jobs available to students, finding that co-op students are frequently required to possess both technical skills (such as knowledge of specific tools) and soft skills (such as communication). Next, we construct two separate weighted bipartite graphs linking the groups of academic programs advertised to by employers to either the required skills or titles of each job. By using community detection to co-cluster the nodes in each graph, we determine the types of skills and roles expected by employers for students in different programs. We find significant differences in the expectations of employers for students in each program, including the importance of soft skills for arts students and the prevalence of data science and artificial intelligence skills in many academic programs.

Second, using over 45000 performance evaluations collected separately for in-person (2019) and remote (2021) internship positions, we uncover the characteristics of successful co-op students. Each evaluation includes an overall performance rating and written comments and recommendations provided by the supervisor. By using logistic regression and word frequency counting to analyze supervisors' general and recommendation comments, we find the most successful students to be excellent leaders and innovators, with remote students also being praised for their independence. Supervisors encourage remote students to be innovative and learn technological skills, while the supervisors of in-person students recommend improving oral communication and presentation abilities.

By identifying the job roles and required skills expected by employers for students in different academic programs, institutions can better prepare students for appropriate jobs. By understanding the skills that contribute to student success in remote and in-person contexts, students can focus on developing the most important skills for their intended work environment. Together, these findings highlight important skills that students should acquire in their early careers.

Acknowledgements

This work would not have been possible without the assistance of many talented people.

To my supervisor, Dr. Lukasz Golab, thank you for always pushing me in the right direction and asking the important questions. Before, I used to think about *what* analysis I should do; now, I think first about *why* I am doing the analysis, to focus on the actionable insights from and positive impact of the results. Your flexibility and guidance have been invaluable and I am a better researcher after working with you.

For their insights into the University of Waterloo co-op program, ideas, advice, and constructive criticism, I would like to thank all members of the Work-Learn Institute, especially Dr. Judene Pretti, Dave Drewery, Robert Craig, and Joanna Wajda.

For providing assistance with her job description parser, invaluable contributions to the literature review, and patience during many weekly meetings, I extend special thanks to Shivangi Chopra.

Finally, to my family and especially to my fiancé, Jason, thank you for supporting me every step of the way. I could not have done this without you.

Dedication

This thesis is dedicated to my fiancé and my parents.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Co-operative education (co-op) is a type of work-integrated learning in which students alternate between classroom study terms and paid work terms. Co-op programs have become popular worldwide as a way for students to acquire practical experience, a talent pipeline for employers, and a recruiting tool for universities [27, 66]. Co-op programs provide students with opportunities to apply the knowledge they have learned and gain real-world experience. Furthermore, they allow students to assess the skills they have been taught in their classes and learn which skills are most valuable for their future career.

Students in the University of Waterloo co-operative education program alternate between study terms, during which they take courses on campus, and work terms, in which they are hired by employers for four-month or eight-month consecutive positions. Students are required to complete between four and six work terms to complete their degree. Before each work term, students use WaterlooWorks, an online database and job matching system, to find and apply to jobs for the upcoming work term. After students have applied and been interviewed, employers rank their preferred candidates on a scale from 1-10, while students rank jobs in a similar fashion. The system then matches students with potential employers by taking the lowest sum of student and employer rankings, attempting to maximize overall satisfaction as much as possible while ensuring that there is a match for every open job [46]. At the end of each work term, employers complete performance evaluations for each student employee, to assess their work and recommend areas for improvement.

According to recent work, the co-op job market is competitive [33, 47, 67], especially since the COVID-19 pandemic, when many co-op positions were cancelled [35]. Furthermore, many students count on being hired permanently by a former co-op employer after graduation [2]. It is therefore important for students to make a good first impression in their early careers.

To be successful in co-operative education programs, it is important for students to meet the expectations of their employers. Employers expect to hire qualified students for each co-op position who possess the required skills, abilities, and attributes. They expect students with different academic backgrounds to possess skills that may make them uniquely suited for specific roles. By understanding what employers expect from co-op students, institutions can ensure that the most valued skills are being emphasized in academic courses. Current co-op students can use this information to determine what to improve to be more successful in future work terms. Furthermore, prospective students can use information about the skills and roles expected by employers to decide which program to pursue. This motivates the following broad research questions:

1. What skills, abilities, and attributes make students successful in co-operative education programs?
2. What roles and skills do employers expect to be most appropriate for students with different academic backgrounds?
3. How do the most valued skills differ in remote and in-person contexts?

To answer these questions, we study two separate secondary datasets collected from the University of Waterloo co-op program:

- First is a set of job descriptions representing jobs posted on WaterlooWorks that were later filled by students in 2021. Each job description is paired with information describing the student hired for that job and a list of co-op job posting clusters used by employers to advertise that job to students from specific academic backgrounds.
- Second is a collection of performance evaluations completed by workplace supervisors for their employees at the end of work terms in 2019 and 2021. Each evaluation consists of an overall performance rating on a 7-point scale, written comments to explain the choice of performance rating, and optional recommendation comments to specify areas of improvement.

Permission for this secondary data analysis was granted by the university's office of research ethics (application number 43970) on January 11, 2022. All analyses were completed on a laptop with an Intel Core i7-10510U CPU @1.8GHz, with 16 GiB of RAM. Most analyses finished quickly (within a few minutes), but some parameter tuning took up to several hours to complete.

1.1 Outline

In Chapter 2, we describe prior work on co-clustering, bipartite graph community detection, and graph methods in job description datasets. Using the list of academic programs targeted by employers, we construct two bipartite graphs. First, to understand which skills are most valued for students in each academic program, we construct a graph mapping the academic programs targeted by each job to the skills required by that job. Second, to understand what roles employers expect to be filled by students with different academic backgrounds, we map job titles to the list of targeted academic programs. Using k-means clustering, frequency analysis, and community detection in the two constructed bipartite graphs, we identify types of jobs available to co-op students and the skills and roles expected for students with different academic backgrounds.

In Chapter 3, we review existing literature on remote and in-person co-operative work positions. Next, we use logistic regression to determine which skills, abilities, and attributes predict outstanding students and how this differs in remote and in-person contexts. Finally, we use frequency counting and logistic regression to understand what supervisors recommend that their remote and in-person students improve to be more successful in future placements.

Chapter 4 summarizes our findings and presents avenues for future work.

Chapter 2

Clustering and Community Detection Analysis of Required Skills in Job Descriptions

2.1 Introduction

Each term at the University of Waterloo, co-op students from a variety of programs search for positions advertised by employers on WaterlooWorks. Those who have not yet secured a position for an upcoming work term will likely be concerned about the skills required to land a job in their field.

Students and institutions may have some idea of the roles available to and best suited for students in each academic program. However, it is not immediately obvious whether these match the expectations of employers. By examining the roles advertised to students in different programs and the skills required for those roles, we can assess the gap between employer expectations and the skills taught to students studying in academic institutions. If students are aware of the skills and roles that employers believe are best suited to their program background, they can be better prepared for future employment in their field. Similarly, institutions can ensure that they are focusing on teaching the specific skills that employers expect most specifically for students in each academic program grouping. With this motivation in mind, we aim to answer the following three questions:

1. What skills, abilities, and attributes are required for co-op jobs in 2021?

2. What skills do employers expect from students with specific academic backgrounds?
3. What job roles do employers expect to be most appropriate students with specific academic backgrounds?

To answer question 1, we perform dimensionality reduction using latent semantic analysis and clustering using k-means to report representative skills from each cluster centroid. We also report the skills mentioned in the largest number of unique job descriptions, to identify popular skills, abilities, and attributes for 2021 jobs.

To answer question 2, we create a weighted bipartite graph. When posting a job, employers must specify a list of co-op job posting clusters used to advertise their jobs to specific student groups. Employers can use these advertising clusters to target the type of student they would like to hire (academic clusters) or to summarize the job function (thematic clusters). This bipartite graph maps academic advertising clusters to skills from the corresponding job's required skills section. We use only academic clusters for this analysis to focus on the skills expected from students in each program. For each job description, the graph is constructed by adding edges between academic clusters and skill tokens. An edge is added between an academic cluster and skill token if a job targeting that cluster mentions the given token in its required skill section, so the edges represent cases where employers expected a specific skill from a student of the given background. Because jobs can target multiple advertising clusters, each cluster is linked to each skill in the job description. The weight of the edge between a cluster and a skill is the number of times that skill is mentioned in a job description targeting that cluster, so that skills and clusters that appear more frequently together are more strongly linked in the graph. We perform community detection on this graph to reveal skills that are strongly linked to specific academic clusters; the skills employers expect those students to have.

To answer question 3, we construct a second weighted bipartite graph, similar to the first, but instead linking academic co-op job posting clusters to job titles (used to represent job roles). Performing community detection on this graph reveals roles that are more strongly linked to one particular academic background than any other.

2.2 Background

2.2.1 Using Graphs to Model and Analyze Job Markets

Many real-world datasets can be modeled as graphs, including social networks, protein interactions, and job markets. Analyzing the structure of these graphs often provides useful insights and reveals relationships that may not be discovered using other methods. In particular, identifying modules (groups of simiilar nodes) in the graph and their boundaries helps to classify nodes and demonstrate how communities differ and interact. A large body of previous work has developed standard techniques for community detection [24]. The Louvain method [8] is a popular community detection algorithm that aims to maximize connections within communities while minimizing connections between them. In this algorithm, each node is initially assigned to its own cluster and then iteratively assigned to or moved to different clusters, selectively merging them with the goal of maximizing a modularity function. Louvain has become common for analyzing graph structure in job market applications. For example, Choi et al. use Louvain to analyze communities in a graph of employees to understand how they are connected based on the similar tasks they perform [12]. Samek et al. construct a graph of skills for artificial intelligence (AI) jobs, linked if they appear in the same job posting, and perform Louvain community detection to find related skills [56]. However, Louvain may generate poorly connected (or even disconnected) communities. More recently, the Leiden algorithm has been introduced to solve this problem and is faster, uncovers better partitions, and guarantees well-connected communities [69]. Louvain is a greedy algorithm, meaning that it always makes the change that results in the best possible increase. This can sometimes result in sub-optimal clusters. The Leiden algorithm provides better guarantees by periodically randomly breaking down communities to ensure that they are always well-connected.

Community detection has also been used to analyze competition in the co-op workforce. Jiang & Golab [33] use interview data to characterize competition between students from different academic programs in the co-op job market. After modeling the interview data as a graph (where each node is an academic program and the edge weights denote the percentage of jobs interviewing at least one student in each of those programs), they perform Louvain community detection to understand the interview relationships between different academic programs. Similarly, Toulis & Golab [67] construct two graphs of jobs and employees, where jobs in the job graph are connected by an edge if they interview at least one student in common, and employees in the employee graph are connected if they interview for the same job. They apply Louvain community detection to identify the top employee and employer communities and identify missed opportunities in the hiring process. However, these studies both focus on characterizing competition using interview data rather than identifying important skills that are expected by employers.

Other graph analyses have been used to study job markets, including graph representations for job recommendation [71, 60, 25] and the use of knowledge graphs to understand online job postings [17]. However, these analyses typically require access to additional datasets with user or employer information that is not available in our dataset.

2.2.2 Using Job Advertisements to Understand Required Skills

Previous work has analysed job postings to better understand the types of skills required by employers. Lyu & Liu extract skill keywords from an online database of job postings and find that soft skills (such as social skills and people management) are increasingly required in the energy sector [40]. Sodhi & Son identify the skills that employers expect from graduates with operational research backgrounds, using frequency and correlation of skill keywords with the specific degree background of employees. Brüning & Mangeol analyse over 9 million job postings from four US states using frequency to understand how employer demand of skills from graduate students varies for different states and occupations [10]. However, these content-based analyses do not leverage relationships between skills and other attributes of a job posting. They also do not use a parser designed to extract skills from job postings and rely on the manual development of skill frameworks.

In contrast, other studies have constructed knowledge graphs to understand the relationships between jobs and skills. After constructing a knowledge graph of skills and occupations, de Groot et al. [18] use shortest path analysis to find similar occupations and find the most relevant skills per occupation group using TF-IDF. However, this analysis does not show the skills that are more strongly linked to one occupation over any other. It is possible for two occupations to have the same top skill. For example, "Electronic Communication" is in the top five skills for three occupation groups: managers, service and sales workers, and technicians and assocate professionals. This analysis does not show the skills that most strongly differentiate each occupation and does not consider the academic or work history background of the employee. Finally, Jia et al. [32] use a long short-term memory (LSTM) network to extract named entities from artificial intelligence (AI) positions. Using rule-based methods, they identify skill entities and then construct and visualize a knowledge graph of skills and occupations to find the most in-demand skills. However, this analysis only focuses on AI positions from mostly Asian recruitment websites and the researchers do not specify how they grouped jobs into categories. Finally, both of these knowledge graph analyses focus on comparing skill requirements for different job types and do not focus on important skills for employees with different academic or work experience backgrounds.

2.2.3 Clustering Job Descriptions

K-means clustering has been previously used to identify the types of co-op jobs available to students. In 2018, Chopra & Golab [13] used a custom job description parser, k-means clustering, and frequency analysis to identify frequent skills and skill clusters from information technology jobs at the University of Waterloo. Although our approach for identifying skill clusters is similar, we use a more recent dataset (2021 jobs) and analyze required skills from all jobs, not only information technology jobs. We are also able to leverage the structure of recent job descriptions, which now have a specific required skills field, instead of needing to extract skills from the full description. As a result, we can be more confident in our results because the risk of extracting non-skill terms is minimized.

2.2.4 Co-Clustering and Bipartite Graph Community Detection

In this analysis, we aim to determine the required skills and job roles that are most important for students with specific academic backgrounds. When constructing graphs of this data, we end up with two classes of node: advertising clusters (which employers can use to target students with specific academic backgrounds) and either job titles (representing job roles) or skills. There are connections only between the two classes, with no direct links between skills/titles or between advertising clusters, so the result is a bipartite graph with two separate classes of nodes. The problem of finding clusters of two classes of nodes simultaneously is known as co-clustering. In co-clustering, data can be stored as a co-occurrence matrix where the rows and columns represent each of the two data types or as a bipartite graph where the data types are modeled as each of the two sets of nodes [54]. Co-clustering is appropriate for many real-world data clustering problems that require capturing the relationships between two types of objects. In 2001, Dhillon proposed spectral co-clustering, a novel approach that solves the problem of simultaneously clustering documents and words. After modelling the documents and words as a bipartite graph, spectral co-clustering is used to identify bipartitionings [20]. Previous work has shown improved results when co-clustering either documents [7] or sentences [11] and words together, compared to the quality of clusters generated by clustering on a single dimension. Co-clustering has been used to understand the qualities users care about most strongly in video games by co-clustering adjectives and context features [53], to cluster sentences and words simultaneously to improve sentence clustering for theme-based summarization [11], to find teams of experts whose combined skills fulfill a given task [22], and to co-cluster user sessions and pageviews for web log analysis [72]. These approaches better reveal the underlying relationships between the two dimensions, instead of relying on one dimension alone.

Many existing solutions to this problem project the bipartite graph to a one-mode network where there is only one type of node, with edges constructed based on relationships with the other type [73]. For example, in our case, we might create a graph of skill nodes where two skills are linked if they appear in documents targeting the same academic clusters. Although doing this can allow us to use powerful tools designed for one-mode networks, we lose some of the crucial information stored in the two-mode version of the network [39]. For example, with our two-mode network, we can see which academic clusters link specific skills and the strength of each individual connection. In the projection, the correspondence between the classes is lost and we cannot easily extract the most important academic advertising clusters that caused the overlap. Furthermore, projection can lead to extremely dense networks with many more edges, making the graph more costly to analyse [39]. In our case, co-clustering enables us to answer our research questions by determining not (for instance) how academic clusters are related to each other, but instead how they are related to job titles or skills.

One method of co-clustering is community detection, but this is non-trivial for bipartite graphs. Most algorithms for community detection, including both Louvain [8] and Leiden [69], are designed for one-mode networks [64]. Some previous work has designed or adapted specialized methods for bipartite community detection [64, 65, 55], but many implementations are not readily available. Fortunately, the Python package Leidenalg [68] implements the Leiden algorithm efficiently in C++ and exposes it to Python. The package supports modified community detection for undirected weighted bipartite graphs. It presents a method for community detection using the Constant Potts Model (CPM), a standard quality function for the Leiden algorithm [69], and defines three different resolution parameters: one for within each class of node and one for the links between the nodes. By emphasizing the links between the classes and de-emphasizing links within the classes, we are able to perform community detection effectively on undirected bipartite graphs without producing poorly connected communities. Finally, community detection forces the assignment of each node to a unique cluster, and both classes of node are clustered together. This enables us to find the skills or roles that are more linked to each academic cluster than to any other cluster.

2.3 Data Overview

Our job description dataset consists of 19083 pairs of job description (including employer information) and student information, describing the student hired for that job. These jobs were posted over one year between the winter and fall terms in 2021. Many jobs hire only one student (73.47%), but the rest have multiple positions available. In the case that more than one student is hired for the same job, that job appears multiple times in the dataset (once for each position that is filled). For each occurrence, the job and employer information is identical, but the student information is specific to each hired co-op student.

Because of this, the 19083 job description pairs represent 13377 unique jobs filled in 2021. When performing the graph analyses described below, we consider unique jobs instead of positions. This is to avoid biasing the analysis towards large companies who offer the same job many times.

In the following sections, we provide a summary of the important attributes of the job descriptions and student information in this dataset.

2.4 Job Descriptions

Each job description consists of a title, three required unstructured text fields, and several optional fields for special requirements. The required fields are:

- **Job Title:** the name of the role or position being advertised, such as "Software Developer". There are 5836 unique job titles in our 2021 dataset.
- **Job Summary:** Typically contains information about the employer, what they are looking for, and why a prospective employee should work for them.
- **Job Responsibilities:** Lists the main tasks and responsibilities of the job.
- **Required Skills:** Lists the required skills of the job.

Employers can also provide additional information about the job and special requirements for prospective employees using a variety of optional fields. This information includes transportation and housing allowances, compensation and benefits, and special requirements for the job and interview process. Unfortunately, these optional fields are often left blank, making it difficult to extract insights from this information. For example, when considering filled positions in 2021, only 43.78% of the compensation and benefits fields and 21.76% of the transportation and housing fields were completed by employers.

There are also a variety of fields to indicate information about the job outside of the job description itself.

- **Work Term:** This field indicates the term during which the job was filled, using a string such as "2019 - Winter" that is specific to a given term and year.

- **Job Posting ID:** A number that uniquely identifies a job posting. Because each row in the spreadsheet associates a job description with a student who filled that job, the same Job Posting ID may appear multiple times (because that job was filled by multiple students). This can be used to remove duplicate rows if only one copy of each unique job is needed.
- **Number of Positions:** The number of available positions for this job, with 70.84% of all jobs having one position only.

2.4.1 Student Information

Each row contains the following information about the student who was hired for the given job (if the job was filled):

- **Academic Level:** Indicates a student's progress through their current degree with a combination of year and semester. Undergraduate students have values ranging from "1A" (first semester) to "5B" (final semester), while master's students always have "M". This field is helpful for identifying a student's level of prior academic experience, but not their level of previous work experience as is indicated by the next field.
- **Work Term Number:** The number of work terms (including this one) completed by this student. For example, "W-4" indicates that a student is on their fourth work term. This field allows researchers to differentiate junior and senior students based on prior work experience. For jobs in 2021, the breakdown of students in each work term was as follows:
  - "W-1": 20.88%
  - "W-2": 20.58%
  - "W-3": 19.04%
  - "W-4": 17.76%
  - "W-5": 13.64%
  - "W-6": 8.1%

  In this analysis, we use work terms to group students by experience: junior students (work terms 1 and 2), intermediate students (work terms 3 and 4), and senior students (work terms 5 and 6). We excluded approximately 191 jobs with work terms higher than 6, as these do not represent typical co-op students and may include master's students (while this analysis focuses on undergraduate students primarily).

- The following fields provide information about a student's major and program.
  - **Faculty:** The employed student's faculty is always one of Engineering (40.02%), Mathematics (26.31%), Arts (14.06%), Science (9.09%), Environment (5.58%), or Health (4.94%). It should be noted that the Mathematics faculty at the University of Waterloo is unique because it includes programs related to computer science that may be included in the Science faculty at other universities.
  - **Primary Plan Code:** A series of letters and numbers that indicate specific information about a student's program. For example, a code featuring -H (e.g., "ZOOH") indicates an honours program, while -HC means honours co-op (e.g., "ZOOHC").
  - **Program Grouping:** This field groups similar primary plan codes and can be used to report findings by program (e.g., "Biology").

2.4.2 Co-op Job Posting Clusters

Employers must select advertising clusters to target the students whose backgrounds they believe would be best suited to the given position. These clusters reflect the academic programs offered by the University of Waterloo, but enable the employer to advertise to groups of students easily without having to understand institution-specific programs. There are two types of clusters:

1. Academic clusters target groupings of academic programs (such as "MATH Applied Mathematics" or "SCI Pharmacy"). These clusters allow the employer to specify the type of student they believe would be the best fit for the role. Each academic cluster is related to at least one of the six faculties. In the dataset, each of these groupings is preceded by the faculty code(s) to which it is related, e.g. "ARTS Business" or "ARTS/SCI Psychology". Only three academic clusters, "ARTS/MATH Finance", "ARTS/MATH/SCI Chartered Professional Accounting" and "ARTS/SCI Psychology" are related to multiple faculties. Finally, each academic cluster is related to one or more real programs from the University of Waterloo and it is possible for a real academic program to be related to multiple clusters. For example, "ARTS Humanities" is related to both Classical Studies and French, while French is also related to the cluster "ARTS Languages and Cultures".

2. Thematic clusters reflect the function of the job (such as "Theme Finance and Investment"). They allow the employer to indicate the type of work that a future employee might perform in the given role.

In total, there are 86 unique clusters represented in the dataset: 41 thematic and 45 academic.

While every job posting must specify at least one advertising cluster, employers can choose freely between academic and thematic clusters. However, the vast majority of employers do specify at least one academic advertising cluster. In 2021, 13088 unique jobs used at least one academic cluster when advertising to students, while only 289 (2.16%) specified thematic clusters only.

A list of all academic clusters in the dataset is provided in Table 2.1. The clusters are grouped by faculty (including a separate section for three academic clusters that correspond to multiple faculties). We also include the number and percentage of unique jobs that advertise to each academic cluster, with the most popular being "MATH Computer Science" (39.9% of unique jobs), "ENG Software Engineering" (38.3%), and "ENG Electrical and Computer Engineering" (34.8%).

Table 2.1: A list of all academic co-op job posting clusters with the number and percentage of unique jobs advertising to that cluster. Each cluster is linked to at least one of the six faculties (ARTS, ENG, ENV, AHS, MATH, and/or SCI).

Co-op Job Posting Clusters	**# of Unique Jobs**	**% of Unique Jobs**
AHS Public Health and Health Systems	561	4.2%
AHS Kinesiology	542	4.1%
AHS Recreation and Leisure Studies	529	4.0%
ARTS Business	2426	18.1%
ARTS Economics	1222	9.1%
ARTS Social Sciences	729	5.4%
ARTS Global Business and Digital Arts	654	4.9%
ARTS English Language and Literature	598	4.5%
ARTS Humanities	484	3.6%
ARTS Sociology and Legal Studies	472	3.5%
ARTS Political Science	442	3.3%
ARTS Languages and Cultures	285	2.1%

ARTS Fine and Performing Arts	252	1.9%
ARTS/MATH Finance	1308	9.8%
ARTS/MATH/SCI Chartered Professional Accounting	883	6.6%
ARTS/SCI Psychology	441	3.3%
ENG Software Engineering	5117	38.3%
ENG Electrical and Computer Engineering	4649	34.8%
ENG Mechanical and Mechatronics Engineering	2679	20.0%
ENG Systems Design and Biomedical Engineering	2617	19.6%
ENG Civil, Environmental and Geological Engineering	1541	11.5%
ENG Management Sciences	1522	11.4%
ENG Nanotechnology Engineering	1176	8.8%
ENG Chemical Engineering	1095	8.2%
ENG Architectural Engineering	595	4.4%
ENG Architecture	508	3.8%
ENV Business, Enterprise and Development	1165	8.7%
ENV Geography and Environmental Management	751	5.6%
ENV Environment, Resources and Sustainability	721	5.4%
ENV Planning	658	4.9%
ENV Geomatics	415	3.1%
MATH Computer Science	5343	39.9%
MATH Business	2402	18.0%
MATH Computing and Financial Management	1768	13.2%
MATH Statistics and Actuarial Science	1623	12.1%
MATH Applied Mathematics	1421	10.6%
MATH Pure Mathematics	940	7.0%
MATH Combinatorics and Optimization	918	6.9%
MATH Teaching	368	2.8%
SCI Business	1532	11.5%
SCI Biological Sciences	825	6.2%
SCI Earth, Environmental and Geological Sciences	729	5.4%
SCI Chemical Sciences	710	5.3%
SCI Physics	700	5.2%
SCI Pharmacy	479	3.6%

Finally, given the faculty of the student hired for each position, we can determine the percentage of employers hiring students from one of the academic program groupings to which they advertised. Out of 19083 total positions, 15888 hired a student whose faculty matched one of the academic clusters used to advertise the job (83.26%). As stated above, a small percentage of jobs (2.16%) did not advertise using academic clusters at all. Considering only cases where an academic cluster was mentioned, there were 15888 matches out of 18136 positions (87.6%). This supports the idea that employers are using advertising clusters accurately (i.e., they are satisfied with hiring students from the academic advertising clusters they target) the majority of the time.

2.4.3 Opportunities and Challenges

From job description information, researchers have the opportunity to understand how employers advertise themselves to and what they expect from prospective student employees. In January 2017, WaterlooWorks replaced JobMine, the University of Waterloo's previous system for hiring co-op students [45]. As part of the new system, a more comprehensive job posting format was introduced. Job descriptions that previously consisted of a title and a text description were further subdivided into five main unstructured text fields, each with a specific purpose: job summary, job responsibilities, required skills, compensation and benefits, and transportation and housing. Assuming that the employer has not made any errors or ignored instructions, we can leverage the new structure by focusing on one specific field when extracting each type of information.

However, this data presents many challenges. Many of the fields in this dataset contain unstructured text that was written freely by the employer. This natural language is difficult to parse because it is prone to exceptions, mistakes, and other idiosyncrasies. For example, words with similar meanings may not have the same representation in the text. Punctuation, special characters, and capitalization make semantically identical words such as "résumé" and "resume." appear to be distinct. Verb tenses, plurality, abbreviations, slang, domain-specific vocabulary, and words with multiple meanings provide additional challenges. For example, when students are required to be familiar with a tool such as Microsoft Office, a wide variety of phrases such as "Microsoft Office", "MS Office", "MSOffice", and "MS Office365" may be used for the same tool. The challenge lies in making these dissimilar phrases appear identical to a computer.

Another challenge is the lack of structure within fields of a job description, leaving employers to individually interpret the best use for each field. For example, there is no clear format for listing responsibilities and skills. Some employers include headers for different sections, but it is not always easy to identify these headers in the text because of the number of possible headers that could be used (e.g., “You will need to:”, “Responsibilities”, “Your tasks will be as follows.”, etc.). Some list responsibilities and skills using numbers or bullet points, with each on a separate line, while others use free-form paragraphs for the same purpose.

When considering how to parse unstructured text fields (such as required skills), we can begin by using standard natural language processing techniques. We can convert the text to lowercase, remove punctuation, and normalize special characters. We can split the text into individual words (called tokens) and use a stemmer to convert words with similar meanings to the same representation. The job description parser developed by Chopra & Golab [13] is designed to extract required skills from job descriptions using a combination of these basic methods and additional domain-specific rules, including pattern matching using regular expressions to identify multi-word skills that should be combined into a single token (such as “problem solving”). Although we now have access to a specific required skills field that was not available when that parser was designed, it is still unstructured text and prone to many of the same issues and challenges. This makes Chopra & Golab’s parser an appropriate starting point for parsing the required skills sections in our dataset.

2.5 Methods

For the k-means clustering, frequency, and graph analyses, we use all 13377 unique jobs filled in 2021. Notably, in contrast to Chapter 3, we do not repeat this analysis to compare to previous jobs in 2019. Initial investigations found that job descriptions have not changed significantly during this period. For example, there are 6708 unique jobs in 2021 for which there is a corresponding job in 2019 with the same title and employer. After parsing using the job description parser (below), we found that of those, 3486 (51.97%) had the same required skills sections as a corresponding 2019 job after parsing. This is quite high considering that any change in wording (e.g., change to the order of skills) would cause the match to fail. Further manual inspection found that many of the jobs that were not an exact match were still very similar (e.g., specifying a newer version of a tool or adding a single required skill). Therefore, to provide the most up-to-date results, we consider only 2021 job descriptions for the following analyses.

2.5.1 Preprocessing

For each analysis described below, we use all 13,377 unique job descriptions from the dataset as described in the data overview section. We begin by extracting the following fields from each job description.

Required Skills

For all analyses involving the required skills field, we apply the job description parser by Chopra & Golab [13]. This parser is designed to extract required skills directly from the unstructured text of job descriptions. However, we found that some tokens related to artificial intelligence (AI), machine learning (ML), and data science were missing or incorrectly parsed. For example, the term “AI” was removed and “neural network” was being incorrectly converted to simply “network”, losing important information. To solve this problem, we use a glossary of AI related terms was developed by the ISO (the International Organization for Standardisation) and IEC (the International Electrotechnical Commission) [29]. The glossary includes terms related to artificial intelligence, machine learning, neural networks, trustworthiness, and natural language processing. After enumerating all forms of each term mentioned in the document, this list produced 164 AI-related terms (including abbreviations). We updated the parser to ensure that all terms are included in the output and that terms consisting of multiple words (such as “neural network”) were joined to be treated as one token. After preprocessing, the result is a list of 6313 unique tokens representing required skills in our dataset.

Co-op Job Posting Clusters

For each job, all advertising clusters are stored as a single string, separated by semicolons. However, two of the thematic clusters contain embedded semicolons (‘Theme Digital and Graphic Media; Web Site Design and Development’ and ‘Theme Visual, Urban & Industrial Design’). Therefore, we first use regular expressions to search for and remove these semicolons. Next, we split the string on the semicolon into a list of clusters for each job. Finally, we filter out thematic clusters (by removing any that begin with “Theme”) to produce a list containing only the academic clusters for each job. After processing, we find 45 unique academic clusters.

Job Titles

No preprocessing is performed on job title text because the number of unique job titles is small and preprocessing might remove important information (for example, stopword removal or stemming might result in abbreviations like “QA” or “CS” being incorrectly removed). Because there are only 5836 unique job titles for 13377 unique jobs, multiple unique jobs may have the same title (such as "Software Developer"). For jobs with duplicate titles, we aggregated the list of co-op job posting clusters for each job by concatenating them together before performing further analyses.

2.5.2 K-Means Clustering and Frequency Analysis

We use clustering to identify the different types of available co-op jobs within a discipline. Following previous work on text clustering, we begin by applying Latent Semantic Analysis (LSA) to the job descriptions [38], with each job description represented as a job vector. Latent semantic analysis is a technique used to reduce the number of dimensions in a dataset by uncovering latent or hidden concepts, usually through a matrix factorization technique called singular value decomposition. LSA has been shown to improve the quality of k-means clustering in text documents [62]. The ith coordinate of a job vector is equal to the inverse document frequency (IDF) of the ith word in the set of possible words, provided that this word is mentioned in the given job description at least once (and zero otherwise). We then use LSA to reduce the dimensionality of job vectors from the number of distinct words. Each reduced dimension corresponds to a latent concept in the data. Next, we run k-means clustering on the transformed job vectors, and we report a few top terms (again, ranked by IDF) from each cluster centroid as representatives.

For this analysis, we determine each student to be either junior, intermediate, or senior based on the number of work terms they have completed so far (including the current position). As described in the data overview, students on their first or second work terms are considered junior, those on their third or fourth terms considered intermediate, and those on their fifth or sixth terms considered senior students.

Next, we perform LSA using scikit-learn's Python implementation of LSA (called TruncatedSVD [59]) and k-means clustering (called KMeans [57]) from the same package. This module was chosen because, in addition to being a popular choice for machine learning applications, it conveniently includes implementations of LSA and k-means along with useful text preprocessing features. We tune the number of reduced dimensions for TruncatedSVD using explained variance, identifying the lowest number of dimensions that reach an explained variance of 80% (although it would be ideal to explain all variance in the data, we do not go higher to avoid overfitting). After trying topic counts in the range 30 to 900 (increasing by 10 at each step), we found that 790 was the best number of topics for our dataset.

For KMeans, after finding the best topic count for LSA, we plot the average WCSS (within-cluster sum of squares) for cluster counts from 3 to 20. Using the empirical elbow method to find the optimal number of clusters in the dataset, we locate "elbow" points where the rate of decrease suddenly changes [70]. The resulting plot can be viewed in Figure 2.1.

However, the elbow method is not necessarily optimal for high dimensional data [37] with hundreds or thousands of features. Even after performing LSA, we have 790 dimensions (topics) in the data, so we relying on the elbow method alone may not not result in optimal clusters.Therefore, we further investigate the cluster counts using silhouette analysis [58], a method used to assess the number of clusters visually. By looking at the average silhouette scores and visualizations of cluster counts 7, 9, and 11 (where elbow points appear in the plot), we selected 11 as the best number of clusters with an average silhouette score of 0.014. The silhouette visualization for this cluster count can be found in Figure 2.2.

Finally, to identify popular skills, we report terms that occur at least once in a large percentage of job descriptions. Because skill terms such as "Python" typically appear only once in each job description, we do not count the number of occurrences of each term within a posting. Instead, we report document frequency, the number of documents in which a term occurs at least once. We also identify terms mentioned more in junior jobs (those hiring junior students) than in senior jobs (those hiring senior students).

2.5.3 Bipartite Community Detection

For this analysis, we are interested in what roles are most suited for and which skills are most specifically required for students of specific academic backgrounds, according to the employer. Therefore, as mentioned previously, we only consider academic clusters in this analysis.

Graph 1: Academic Cluster to Skill

In this graph, each node is either an academic advertising cluster or a skill token produced by the job description parser. An edge is added between an academic cluster node and a skill node if at least one job advertising using that cluster also contains that skill term in its required skills section. Because multiple jobs might specify the same skill and academic cluster, the weight of each edge is equal to the number of job descriptions with the same skill and cluster. Therefore, skills and clusters that appear in the same description more

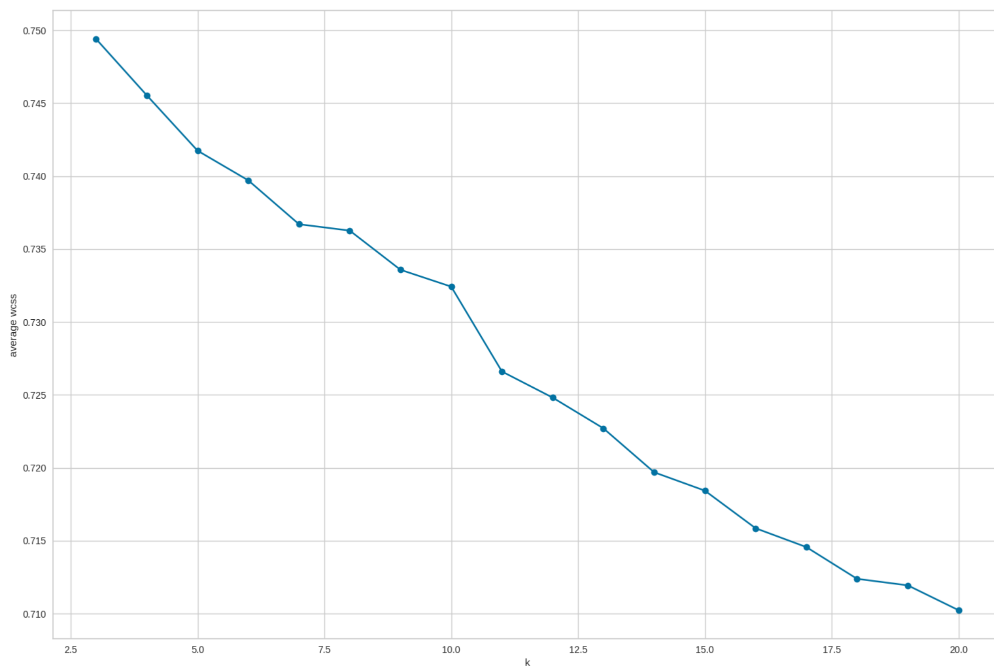


Figure 2.1: Average within-cluster sum of squares (WCSS) vs number of k-means clusters. The x-axis plots k, the number of clusters, while the y-axis shows the corresponding WCSS score (the sum of all distances between points and their cluster centroids) for that number of clusters after running k-means. This value will decrease as the number of clusters increases. The optimal number of clusters can be empirically found by choosing a value for k after which the WCSS score stays constant (or decreases less rapidly), called an elbow point. We identify potential elbow points at cluster counts of 7, 9, and 11.

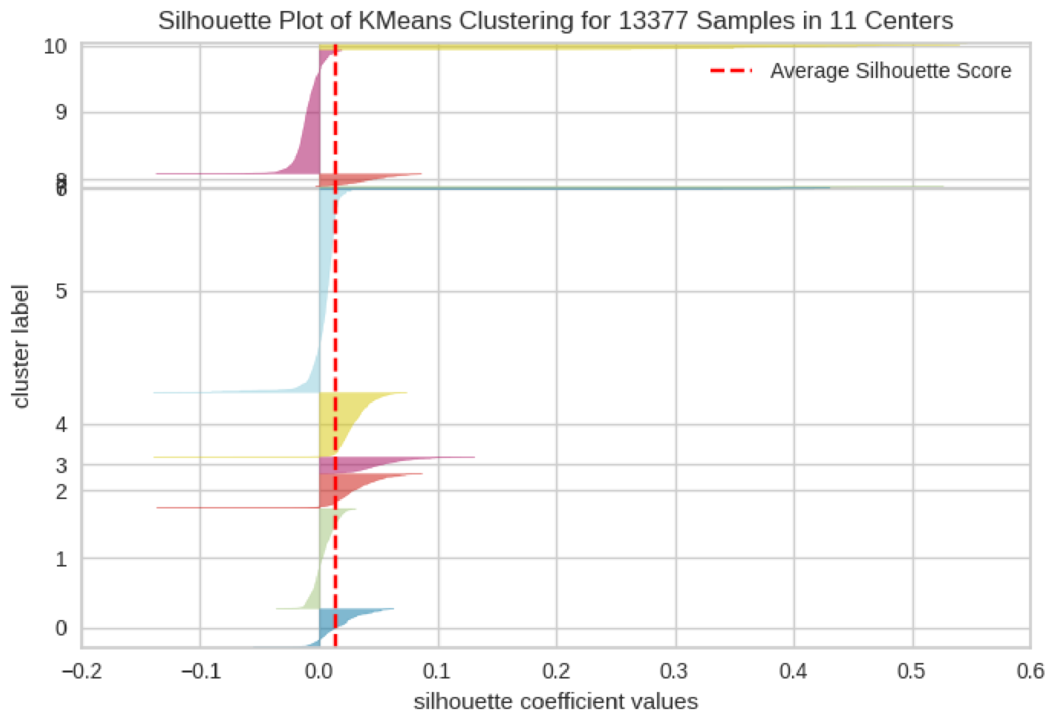


Figure 2.2: A visualization of the silhouette coefficients for each of 11 k-means clusters. Values range from -1 (indicating that the skill may have been assigned to the wrong cluster) to +1 (indicating that the given skill is strongly associated with its cluster and far from other clusters).

often will be more strongly linked. The result is an undirected, weighted, bipartite graph mapping academic advertising clusters to required skills.

Graph 2: Academic Cluster to Job Title

In this graph, each node is either a job title or an academic cluster. An edge is added between an academic cluster node and a job title node if at least one job with that title specifies the given academic cluster. As described above, each job title was aggregated into a single node by concatenating all co-op job posting clusters for all jobs with that title. Because multiple jobs may have the same job title, each edge is also weighted. Each time an academic cluster is found in the same job description as a given title, the edge weight is increased by one. Therefore, the resulting edge weight is equal to the number of job descriptions containing both the given job title and the given academic cluster. The result is an undirected, weighted, bipartite graph mapping academic advertising clusters to job titles.

A small example skill graph (Figure 2.3) is constructed using only two job postings (Table 2.2), both real examples taken from the dataset, but slightly modified for privacy reasons. This simple graph demonstrates not only how job descriptions are parsed into skill tokens, but also how skills and co-op job posting clusters are connected in the bipartite graph.

Community Detection

For both graphs, we perform community detection using the Leiden algorithm with the Constant Potts Model quality function (CPM) using the Leidenalg Python package [68]. We first tune resolution parameters specific to bipartite community detection as specified in the Leidenalg documentation and then use Gephi [5] to visualize the resulting communities. Gephi is an open-source network analysis and visualization platform that can be used to explore many types of graphs. For each community, we compute the degree and weighted degree of each node within their community. The degree of a node is the number of other nodes in the same community connected by a direct edge, while the weighted degree is the sum of the weights of these edges. Together, these statistics represent how strongly each node is linked to other nodes in its community. We then report the academic clusters and top 15 skills (for graph 1) or top 15 job titles (for graph 2) sorted by weighted degree (from high to low) associated with those clusters.

An observant reader might wonder why we do not also construct a bipartite graph between job titles and required skills. First, this graph would not provide information

Table 2.2: Job postings used to generate the example graph (Figure 2.3), including the title, original required skills section, tokens produced after parsing, and co-op job posting clusters.

Job Title	Original Required Skills	Parsed Tokens	Co-op Job Posting Clusters
Full Stack Software Engineering Intern	Must be taking a Computer Science or related technical degree. Experience programming in at least one of these languages: Java, C#, Python, C++. Eager to learn and passionate.	comput scienc technic languag java c# python c++ eager learn passion	Theme Computing: Software; ENG Electrical and Computer Engineering; ENG Software Engineering; MATH Computer Science
Research Assistant - Geochemistry	Applicants need introductory chemistry and should have experience using Microsoft Excel and CorelDraw. Nice to have experience with SigmaPlot, Python and Microsoft Access as well.	applic introductori chemistri msexcel coreldraw sigmaplot python msaccess	Theme Natural Resource Management; Theme Scientific Experimental Design and Laboratory Assistance; SCI Chemical Sciences; SCI Earth, Environmental and Geological Sciences

to answer our research questions, because it does not determine which skills are most important for students with specific academic backgrounds. Furthermore, students in each academic program can already examine the two graphs described above to understand the skills and job titles for which employers expect them to be well-suited. Second, the lack of structure makes insights from this graph more difficult to justify. Each academic advertising cluster is clearly defined and well-structured. However, it is more difficult to group job titles because different employers may decide to use different titles for the same job, leading to noise. Additionally, these insights could not be targeted to any particular academic program grouping, meaning that they are not useful in determining employer expectations for specific groups of students.

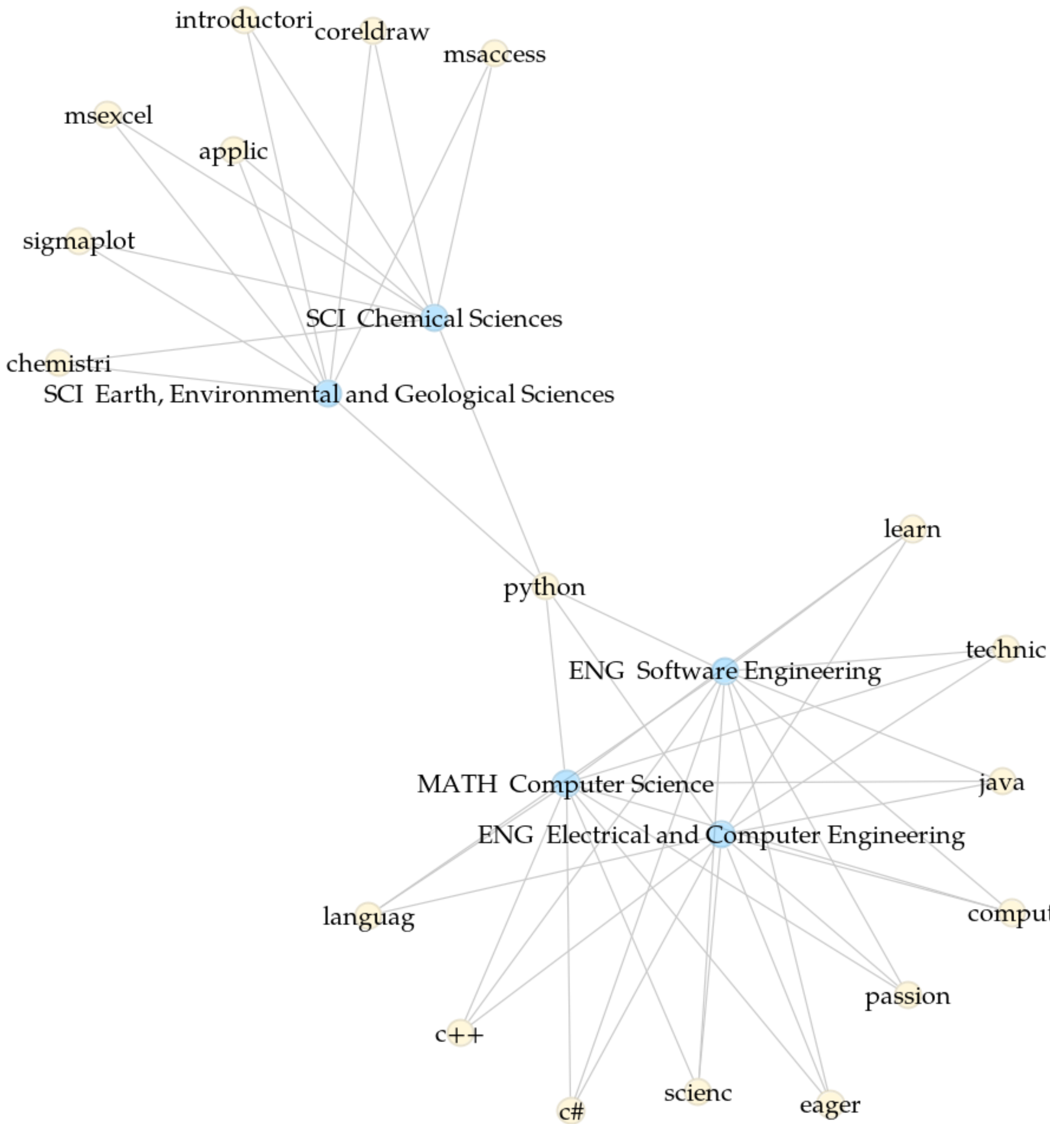


Figure 2.3: A small example bipartite skill graph, generated using the job postings in Table 2.2 All nodes are labelled and colored either blue (academic clusters) or yellow (skill tokens). Only the token *python* appears in both required skills sections, so it is connected to all academic clusters.

2.6 Results

2.6.1 K-Means Clustering and Frequency Analysis

The top 25 most frequent skills, abilities, and attributes for all students (regardless of academic program) are listed in Table 2.3. When comparing skills required more frequently for junior (Table 2.4) and senior (Table 2.5) students we find that attention to detail (*attentiontodetail*), social media (*media*) and communication (*communic*) are most important for junior students. For senior students, science (*scienc*), the programming language Python (*python*), and data (*data*) are most relevant. In all three tables, the Original Words column shows phrases from the original text that produced these tokens. The original text is lowercased and punctuation is removed to avoid displaying near-duplicate strings like "communication" and "Communication.".

Finally, Table 2.6 shows the eleven skill clusters generated using K-means. For each cluster, we assign a manual label (Label) and provide the terms closest to the cluster centroid (Tokens in cluster centroid) and the percentage of all (%All), junior (%Jr), intermediate (%Intr), and senior (%Sr) jobs assigned to that cluster. Note that the type of job is determined by the work experience of the student who filled it; for example, a job is considered to be a "junior job" if it was filled by a junior student. Here is a detailed list of the eleven clusters and representative phrases from the original textthat produced each token (in brackets):

- In the Research and Design cluster, we find the terms *communic* ("communication skills"), *knowledg* ("knowledge", "knowledgeable"), *team* ("team", "team player"), *excel* ("excellent", "excel"), *mechan* ("mechanical", "mechanics"), *design* ("design", "designation"), *learn* ("learn", "learning"), *project* ("projects", "projections"), *softwar* ("software"), *research* ("research", "researching").

- In the Communication cluster, we find the terms *excel* ("excellent", "excel"), *communic* ("communication skills"), *team* ("team", "team player"), *msoffic* ("microsoft office", "ms office"), *written* ("written"), *attentiontodetail* ("attention to detail"), *timemanag* ("time management", "manage time"), *interperson* ("interpersonal", "interpersonal relations"), *profici* ("proficiency", "proficient"), *verbal* ("verbal", "verbally").

- In the Programming and Software Development cluster, we find the terms *develop* ("development", "developing"), *java* ("java"), *test* ("testing", "test", "tests", "tested"), *softwar* ("software", "softwares"), *scienc* ("science", "sciences"), *comput* ("computer", "computing"), *languag* ("languages", "language"), *code* ("code", "coding"), *oop* ("object oriented", "oop"), *python* ("python").

- In the Web Development cluster, we find the terms *javascript* ("javascript"), *css* ("css"), *react* ("react", "reacts"), *html* ("html"), *web* ("web"), *framework* ("frameworks", "framework"), *develop* ("development", "developing"), *databas* ("database", "databases"), *api* ("apis", "api"), *sql* ("sql").

- In the Data Science cluster, we find the terms *statist* ("statistics", "statistical"), *scienc* ("science", "sciences"), *mathemat* ("mathematics", "mathematical"), *comput* ("computer", "computing"), *sql* ("sql"), *data* ("data"), *python* ("python"), *analyt* ("analytical", "analytics"), *actuari* ("actuarial", "actuaries"), *vba* ("vba").

- In the Scripting cluster, we find the terms *linux* ("linux"), *python* ("python"), *system* ("systems", "system"), *script* ("scripting", "scripts"), *softwar* ("software", "softwares"), *languag* ("languages", "language"), *embed* ("embedded", "embed"), *knowledg* ("knowledge", "knowledgeable"), *comput* ("computer", "computing"), *develop* ("development", "developing").

- In the Finance cluster, we find the terms *account* ("accounting", "accountability"), *financ* ("finance", "financing"), *focus* ("focus", "focused"), *busi* ("business", "busy"), *cpa* ("cpa"), *analyt* ("analytical", "analytics"), *excel* ("excellent", "excel"), *lead* ("leadership", "leading"), *communic* ("communication skills"), *team* ("team", "team player").

- In the Graphic Design cluster, we find the terms *adob* ("adobe"), *media* ("media"), *social* ("social", "socializing"), *photoshop* ("photoshop", "intermediate photoshop"), *illustr* ("illustrator", "illustration"), *market* ("marketing", "markets"), *creativ* ("creative", "creativity"), *indesign* ("indesign"), *design* ("design", "designation"), *graphic* ("graphic", "graphics").

- In the Vehicle Access cluster, we find the terms *driver* ("driver", "drivers"), *valid* ("valid", "validation"), *vehicl* ("vehicle", "vehicles"), *access* ("access", "accessibility"), *msoffic* ("microsoft office", "ms office"), *excel* ("excellent", "excel"), *civil* ("civil"), *licenc* ("licence"), *communic* ("communication skills"), *knowledg* ("knowledge", "knowledgeable").

- In the Remote Work cluster, we find the terms *genuin* ("genuine", "genuinely"), *variat* ("variations", "variation"), *zoom* ("zoom"), *teleconferenc* ("teleconferencing"), *uncertainti* ("uncertainty"), *juggl* ("juggle", "juggling"), *webex* ("webex"), *workload* ("workload", "workloads"), *deal* ("deal", "dealing"), *volatil* ("volatility").

- In the Inclusivity and Integrity cluster, we find the terms *articul* ("articulate", "articulating"), *trust* ("trust", "trusted"), *root* ("root", "rooted"), *authent* ("authentication", "authentic"), *genuin* ("genuine", "genuinely"), *inclus* ("inclusive", "inclusion"), *peopl* ("people", "peoples"), *cpa* ("cpa"), *idea* ("ideas", "idea"), *resili* ("resilience", "resilient").

Table 2.3: Top 25 most frequent attributes in the required skills section of filled job descriptions in 2021.

Token	Original Words	Freq. in 2021
communic	“communication skills”, “communication”, “verbal communication skills”, “communicate”	51.8%
knowledg	“knowledge”, “knowledgeable”, “knowledgable”	43.7%
excel	“excellent”, “excel”, “excellence”, “excels”, “excelled”	43.5%
team	“team”, “team player”, “teams”, “teamwork”, “a”	41.1%
develop	“development”, “developing”, “develop”, “developed”, “developer”	31.7%
softwar	“software”, “softwares”	28.9%
comput	“computer”, “computing”, “computers”, “computational”, “compute”	28.7%
written	“written”	24.2%
design	“design”, “designation”, “designing”, “designs”, “designed”	22.9%
analyt	“analytical”, “analytics”, “analytic”, “analytically”	22.4%
scienc	“science”, “sciences”	21.9%
problemsolv	“problem solving”, “problem solver”, “problem solve”, “problem solvers”, “and problem”	21.2%
manag	“management”, “manage”, “managing”, “manager”, “managers”	20.6%
data	“data”	20.4%
applic	“applications”, “application”, “applicants”, “applicable”, “applicant”	18.6%
msoffic	“microsoft office”, “ms office”, “msoffice”, “current ms office”, “ms office365”	18.3%
python	“python”	18.2%
profici	“proficiency”, “proficient”, “proficiently”, “proficiencies”	17.9%
interperson	“interpersonal”, “interpersonal relations”, “excellent interpersonal”, “interpersonal and relationship”, “interpersonal written verbal”	17.0%
learn	“learn”, “learning”, “learned”, “learns”, “learnings”	17.0%
busi	“business”, “busy”, “businesses”	16.6%
tool	“tools”, “tooling”, “tool”	16.5%
system	“systems”, “system”, “systemic”	16.4%
project	“projects”, “project”, “projections”, “projection”, “projected”	15.9%
demonstr	“demonstrated”, “demonstrate”, “demonstrates”, “demonstrating”, “demonstrable”	15.7%

Table 2.4: Frequency differences for attributes of junior jobs in 2021 required skills.

Token	Original Words	Jr.	Sr.	Δ
attentiontodetail	“attention to detail”, “attention to details”, “high attention to detail”, “amazing attention to detail”	17.7%	12.4%	5.3%
media	“media”	7.6%	2.3%	5.3%
communic	“communication skills”, “communication”, “verbal communication skills”, “communicate”	54.0%	48.9%	5.2%
customerservic	“customer service”, “communication customer”, “customer services”	7.5%	2.6%	4.9%
timemanag	“time management”, “manage time”, “time manager”, “time manage”, “manages time”	14.7%	9.8%	4.8%
msoffic	“microsoft office”, “ms office”, “msoffice”, “current ms office”, “ms office365”	20.3%	15.5%	4.8%
account	“accounting”, “accountability”, “accountable”, “account”, “accounts”	8.6%	4.1%	4.5%
web	“web”	14.4%	10.8%	3.6%
supervis	“supervision”, “supervised”, “supervise”, “supervising”	7.8%	4.5%	3.4%
interperson	“interpersonal”, “interpersonal relations”, “excellent interpersonal”, “interpersonal and relationship”, “interpersonal written verbal”	17.4%	14.2%	3.3%
writtencomm	“written communication skills”	12.7%	9.5%	3.3%
social	“social”, “socializing”, “socials”	6.7%	3.4%	3.2%
task	“tasks”, “task”, “tasked”	10.9%	7.9%	3.0%
adapt	“adapt”, “adaptability”, “adaptable”, “adapts”, “adapting”	8.8%	5.9%	2.9%
content	“content”, “contents”, “contentful”	4.3%	1.4%	2.9%
priorit	“prioritize”, “prioritization”, “prioritizing”, “prioritizes”, “priorit”	7.7%	4.9%	2.8%
edit	“editing”, “edit”, “edition”, “edited”, “edits”	4.0%	1.2%	2.8%
multitask	“multi task”, “multitask”, “multi tasking”, “multitasking”, “multi tasker”	8.0%	5.3%	2.7%
oral	“oral”, “orally”, “orale”	9.0%	6.5%	2.5%
photoshop	“photoshop”, “intermediate photoshop”, “sketchup photoshop”	4.1%	1.7%	2.4%

Table 2.5: Frequency differences for attributes of senior jobs in 2021 required skills.

Token	Original Words	Sr.	Jr.	Δ
scienc	"science", "sciences"	27.8%	17.5%	10.3%
python	"python"	24.3%	14.1%	10.3%
data	"data"	24.7%	16.5%	8.2%
c++	"c++"	13.4%	6.9%	6.5%
system	"systems", "system", "systemic"	20.9%	14.4%	6.4%
algorithm	"algorithms", "algorithm", "algorithmic"	8.1%	2.2%	5.9%
design	"design", "designation", "designing", "designs", "designed"	26.5%	20.6%	5.9%
build	"building", "build", "builds", "buildings"	13.6%	7.8%	5.8%
structur	"structures", "structural", "structure", "structured", "structuring"	9.1%	3.4%	5.7%
comput	"computer", "computing", "computers", "computational", "compute"	32.3%	26.6%	5.7%
develop	"development", "developing", "develop", "developed", "developer"	34.2%	28.7%	5.5%
mechan	"mechanical", "mechanics", "mechanisms", "mechanically", "mechanic"	8.9%	3.5%	5.4%
bachelor	"bachelor", "bachelors", "bachelor's"	10.2%	4.8%	5.4%
statist	"statistics", "statistical", "statistic"	10.0%	5.1%	4.9%
java	"java"	12.8%	8.0%	4.8%
languag	"languages", "language"	17.6%	13.0%	4.6%
model	"modeling", "modelling", "models", "model"	9.8%	5.4%	4.4%
project	"projects", "project", "projections", "projection", "projected"	18.6%	14.3%	4.4%
passion	"passion", "passionate", "passions", "passionately", "passioned"	14.4%	10.1%	4.2%
process	"process", "processes", "processinag", "processed", "processe"	16.1%	11.9%	4.2%

Table 2.6: Required skill clusters of 2021 jobs, including a manual label and the portion of all, junior, intermediate, and senior jobs assigned to each cluster.

Label	Tokens in cluster centroid	%All	%Jr	%Intr	%Sr
Research & Design	communic, knowledg, team, excel, mechan, design, learn, project, softwar, research	30.7%	30.6%	29.0%	33.4%
Communication	excel, communic, team, msoffic, written, attentiontodetail, timemanag, interperson, profici, verbal	17.9%	21.5%	17.6%	13.6%
Programming and Software Development	develop, java, test, softwar, scienc, comput, languag, code, oop, python	11.5%	8.9%	11.8%	14.7%
Web Development	javascript, css, react, html, web, framework, develop, databas, api, sql	8.3%	8.9%	8.0%	7.7%
Data Science	statist, scienc, mathemat, comput, sql, data, python, analyt, actuari, vba	8.3%	5.7%	8.8%	11.1%
Scripting	linux, python, system, script, softwar, languag, embed, knowledg, comput, develop	7.2%	6.0%	7.2%	8.9%
Finance	account, financ, focus, busi, cpa, analyt, excel, lead, communic, team	6.0%	5.9%	8.3%	2.8%
Graphic Design	adob, media, social, photoshop, illustr, market, creativ, indesign, design, graphic	5.6%	8.6%	4.4%	3.5%
Vehicle Access	driver, valid, vehicl, access, msoffic, excel, civil, licenc, communic, knowledg	2.9%	2.5%	3.2%	3.0%
Remote Work	genuin, variat, zoom, teleconferenc, uncertainti, juggl, webex, workload, deal, volatil	1.3%	1.4%	1.4%	1.1%
Inclusivity and Integrity	articul, trust, root, authent, genuin, inclus, peopl, cpa, idea, resili	0.2%	0.1%	0.4%	0.1%

2.6.2 Bipartite Community Detection

This analysis produced two graphs:

1. **Skill Graph**: The graph between academic clusters and skills (referred to as the "skill graph" in this section) consists of 6359 nodes and 76901 edges.
2. **Job Title Graph**: The graph between academic clusters and job titles (referred to as the "job title graph" in this section) has 5881 nodes and 26993 edges.

Although the number of nodes is similar for each graph, the job title graph is considerably less dense with approximately 65% fewer edges.

After performing community detection on each graph, we found 10 communities for the skill graph and 12 communities for the job title graph. In the following sections, we describe all communities using tables. For the job title graph, we also include images of each community's structure, created using Gephi by applying the Force Atlas 2 layout algorithm, which aims to repel weakly connected nodes while keeping strongly connected ones together in the visualization [30]. Because there are considerably more edges in the skill graph, visualizations using Gephi did not reveal interesting insights about the structure of skill graph communities, so no images of this graph are included.

Graph 1: Academic Cluster to Skill

In addition to the 10 communities described below, 83 skills were assigned to their own single-node community.

- 67 of these skills were only listed in jobs that never advertised to any academic cluster (that is, they used thematic clusters only). As a result, these skills did not form any edges to academic clusters and could not become part of any larger community.
- The remaining 16 nodes corresponded to the following skill terms: *arcgi* (ArcGIS), *urban*, *spatial*, *configur* (configure), *climat* (climate), *tea*, *failur* (failure), *landscap* (landscape), *logist* (logistics), *fair*, *rate*, *knack*, *interfac* (interface), *studio*, *esitm* (a typo for "estimate"), and *ethnic*. These required skills were evenly linked to a wide variety of academic clusters, so they did not get assigned to any particular cluster. It is likely that this is the result of an employer targeting all 45 academic clusters regardless of whether they are appropriate, as occurred in 10 unique jobs in the dataset (0.07%), and that these terms did not appear in many documents.

- **Community 0: Arts and Humanities** contains 949 nodes (14.92%) and 6584 edges (8.56%).
  - This community includes eight of the ten program groupings from the faculty of Arts, along with Recreation and Leisure studies from the faculty of Health. Students in these fields are associated with communication skills, creativity, and time management.

Table 2.7: The academic advertising clusters and top 15 skills from Community 0 in the skill graph, sorted by weighted degree within the community.

Parsed Token	**Original (Advertising Cluster or Skill)**	**Degree in Community**	**Weighted Degree in Community**
N/A	ARTS Social Sciences	824	19236
N/A	ARTS English Language and Literature	799	15724
N/A	AHS Recreation and Leisure Studies	669	12819
N/A	ARTS Humanities	791	12509
N/A	ARTS Political Science	722	12035
N/A	ARTS Sociology and Legal Studies	733	11700
N/A	ARTS/SCI Psychology	719	11111
N/A	ARTS Languages and Cultures	672	8149
N/A	ARTS Fine and Performing Arts	655	7322
communic	"communication skills", "communication", "verbal communication skills", "communicate"	9	2920
excel	"excellent", "excel", "excellence", "excels", "excelled"	9	2484
team	"team", "team player", "teams", "teamwork"	9	2208
knowledg	"knowledge", "knowledgeable", "knowledgable"	9	1542
written	"written"	9	1512
manag	"management", "manage", "managing", "manager", "managers"	9	1257
creativ	"creative", "creativity", "creatively"	9	1210
timemanag	"time management", "manage time", "time manager", "time manage", "manages time"	9	1198
msoffic	"microsoft office", "ms office", "msoffice", "current ms office", "ms office365"	9	1156
attentiontodetail	"attention to detail", "attention to details", "high attention to detail", "amazing attention to detail"	9	1121

develop	"development", "developing", "develop", "developed", "developer"	9	1085
demonstr	"demonstrated", "demonstrate", "demonstrates", "demonstrating", "demonstrable"	9	1016
problemsolv	"problem solving", "problem solver", "problem solve", "problem solvers", "and problem"	9	991
write	"writing", "write", "writes", "writings"	9	880
learn	"learn", "learning", "learned", "learns", "learnings"	9	877

- **Community 1: Computer Science, Software, and Hardware Engineering** consists of 1982 nodes (31.17%) and 6682 edges (8.69%).
  - The skills most specific to this community (Table 2.8) are highly technical, such as firmware, J2EE (used for web-based applications), SASS (Syntactically Awesome Style Sheets), DynamoDB (a NoSQL database), and Apache Maven (used for project management).

Table 2.8: The academic advertising clusters and top 15 skills from Community 1 in the skill graph, sorted by weighted degree within the community.

Parsed Token	Original (Advertising Cluster or Skill)	Degree in Community	Weighted Degree in Community
N/A	ENG Software Engineering	1623	6483
N/A	ENG Electrical and Computer Engineering	1585	6171
N/A	MATH Computer Science	1510	5987
N/A	ENG Systems Design and Biomedical Engineering	998	3209
N/A	ENG Mechanical and Mechatronics Engineering	966	3040
firmwar	"firmware"	5	219
j2ee	"j2ee"	5	212
sass	"sass"	5	186
dynamodb	"dynamodb"	5	185
maven	"maven"	5	185
uml	"uml"	4	151
bluetooth	"bluetooth"	5	143
i2c	"i2c"	5	134

microprocess	"microprocessors", "microprocessor", "of microprocessor", "hardware microprocessor", "hardware microprocessors"	5	133
repositori	"repository", "repositories"	5	131
backbon	"backbone"	5	127
ember	"ember"	5	118
es6	"es6"	5	113
semiconductor	"semiconductor", "semiconductors"	5	105
laravel	"laravel"	5	104

- **Community 2: Business** consists of 591 nodes (9.29%) and 2328 edges (3.03%).
  - This community (Table 2.9) demonstrates that there are also tools specific to business positions. Employers expect these students to be familiar with Scrum (the project management framework), QuickBooks (an accounting software package), and specific math and statistical concepts like matrices and bias.

Table 2.9: The academic advertising clusters and top 15 skills from Community 2 in the skill graph, sorted by weighted degree within the community.

Parsed Token	**Original (Advertising Cluster or Skill)**	**Degree in Community**	**Weighted Degree in Community**
N/A	ARTS Business	537	1506
N/A	MATH Business	481	1239
N/A	SCI Business	382	905
N/A	ENV Business, Enterprise and Development	360	769
N/A	ARTS Economics	294	679
N/A	ARTS Global Business and Digital Arts	274	555
diplomat	"diplomatic", "diplomatically"	6	75
scrum	"scrum"	6	64
quickbook	"quickbooks", "quickbook"	5	49
matrix	"matrix", "matrixed"	6	47
feet	"feet"	6	47
bias	"bias", "biases"	6	43
central	"centrally", "central", "centralized", "centralize", "centralization"	6	42

net	"net", "nets"	6	41
pitch	"pitch", "pitching"	5	40
browser	"browser", "browsers"	6	39
waterloowork	"waterlooworks"	6	39
wirefram	"wireframes", "wireframing", "wireframe"	6	37
irregular	"irregular"	5	37
gimp	"gimp"	5	37
probe	"probing", "probe", "probes"	6	36

- **Community 3: Environment and Geological Sciences** consists of 711 nodes (11.18%) and 2658 edges (3.46%).
  - In addition to using environmental terminology more frequently (e.g., water, geology, conservation, weather) when advertising to students from this community (Table 2.10), employers also expect them to be familiar with AutoCAD, a web app used to create 2D and 3D drawings.

Table 2.10: The academic advertising clusters and top 15 skills from Community 3 in the skill graph, sorted by weighted degree within the community.

Parsed Token	**Original (Advertising Cluster or Skill)**	**Degree in Community**	**Weighted Degree in Community**
N/A	ENG Civil, Environmental and Geological Engineering	568	2839
N/A	ENV Environment, Resources and Sustainability	510	1497
N/A	ENV Geography and Environmental Management	479	1495
N/A	SCI Earth, Environmental and Geological Sciences	500	1470
N/A	ENV Planning	313	968
N/A	ENV Geomatics	288	880
civil	"civil"	6	454
autocad	"autocad"	6	452
water	"water", "watered"	6	275
geolog	"geology", "geological"	6	118
licenc	"licence"	6	117

chemic	“chemical”, “chemicals”	6	108
conserv	“conservation”, “conservative”	6	95
wast	“waste”, “wasted”	6	94
gps	“gps”	6	88
qgis	“qgis”	6	86
weather	“weather”	6	86
ecolog	“ecology”, “ecological”	6	81
earth	“earth”	6	79
hydrogeolog	“hydrogeology”, “hydrogeological”	5	74
geograph	“geographic”, “geographically”, “geographical”, “geographers”	6	73

- **Community 4: Math, Statistics, and Optimization** consists of 600 nodes (9.44%) and 2189 edges (2.85%).
  - Employers expect math students in this community (Table 2.11) to be familiar with the Society of Actuaries (SOA) and programming languages such as Typescript, HTML5, Racket, and Ruby.

Table 2.11: The academic advertising clusters and top 15 skills from Community 4 in the skill graph, sorted by weighted degree within the community.

Parsed Token	**Original (Advertising Cluster or Skill)**	**Degree in Community**	**Weighted Degree in Community**
N/A	MATH Computing and Financial Management	452	1869
N/A	MATH Statistics and Actuarial Science	481	1780
N/A	MATH Applied Mathematics	474	1432
N/A	MATH Pure Mathematics	419	1325
N/A	MATH Combinatorics and Optimization	363	954
soa	“soa”, “service oriented architecture”, “service oriented architectures”	5	149
forecast	“forecasting”, “forecasts”	5	139
typescript	“typescript”	5	106
html5	“html5”	5	101

deriv	“derive”, “derivatives”, “deriving”, “derivative”	5	91
racket	“racket”	5	86
rubi	“ruby”, “rubi”	5	81
cpa	“cpa”	5	73
backend	“backend”	5	72
postgr	“postgres”	5	71
frontend	“frontend”	5	67
ror	“ruby on rails”, “rails”, “with ruby on rails”, “ror”	5	66
hp	“hp”	5	65
kera	“keras”	5	65
load	“load”, “loading”, “loads”, “loaded”	5	63

- **Community 5: Architecture, Pharmacy, and Teaching** consists of 453 nodes (7.12%) and 1288 edges (1.67%).
  - These students have (according to employer expectations) a requirement for machine learning and AI skills in common: algorithms, TensorFlow, PyTorch, along with other tech skills like Rhinoceros 3D, Javascript, and concepts like object-oriented programming (OOP), and application programming interfaces (API) (Table 2.12).

Table 2.12: The academic advertising clusters and top 15 skills from Community 5 in the skill graph, sorted by weighted degree within the community.

Parsed Token	Original (Advertising Cluster or Skill)	Degree in Community	Weighted Degree in Community
N/A	ENG Architectural Engineering	356	1769
N/A	ENG Architecture	371	1693
N/A	SCI Pharmacy	296	1211
N/A	MATH Teaching	265	1031
revit	“revit”	2	145
algorithm	“algorithms”, “algorithm”, “algorithmic”	4	128
patient	“patient”, “patients”, “patiently”	4	127
linux	“linux”	4	117

framework	"frameworks", "framework"	4	116
electr	"electrical", "electric", "electricity"	4	96
tensorflow	"tensorflow"	4	88
distribut	"distributed", "distribution", "distributes", "distributions"	4	83
cand	"cand", "candid"	4	83
control	"control", "controls", "controlled", "controllers", "controller"	4	81
pytorch	"pytorch"	4	80
rhino	"rhino"	2	68
javascript	"javascript"	4	67
unix	"unix"	4	67
oop	"object oriented", "oop", "of object oriented", "object orientated", "and object oriented"	4	65

- **Community 6: General Sciences** consists of 453 nodes (7.12%) and 1192 edges (1.55%).
  - Students in biology, chemistry, physics, and nanotechnology (Table 2.13) are expected to be specifically familiar with techniques and tools such as soldering, assays (a technique to investigate the composition of a metal or ore), microscopes, sensors, and Arduinos.

Table 2.13: The academic advertising clusters and top 15 skills from Community 6 in the skill graph, sorted by weighted degree within the community.

Parsed Token	**Original (Advertising Cluster or Skill)**	**Degree in Community**	**Weighted Degree in Community**
N/A	ENG Nanotechnology Engineering	214	991
N/A	SCI Biological Sciences	278	911
N/A	ENG Chemical Engineering	256	823
N/A	SCI Chemical Sciences	248	800
N/A	SCI Physics	196	641
nanotechnolog	"nanotechnology"	5	157
optic	"optical", "optics"	5	118

molecular	"molecular"	5	86
microbiolog	"microbiology", "microbiological"	5	79
solder	"soldering", "solder"	5	64
biomed	"biomedical"	5	61
assay	"assays", "assay"	5	61
cell	"cell", "cells"	4	61
microscopi	"microscopy"	5	58
bioinformat	"bioinformatics", "bioinformatic"	4	58
arduino	"arduino", "arduinos"	5	55
wet	"wet"	5	50
pharmaceut	"pharmaceutical", "pharmaceuticals"	5	47
sensor	"sensors", "sensor"	5	47
trial	"trial", "trials"	5	46

- **Community 7: Finance and Accounting** consists of 245 nodes (3.85%) and edges (0.59%).
  - In addition to tools such as Visual Basic for Application (VBA) and concepts such as auditing and risk management, finance and accounting students are uniquely expected to be lifelong learners and focus on extracurriculars (Table 2.14).

Table 2.14: The academic advertising clusters and top 15 skills from Community 7 in the skill graph, sorted by weighted degree within the community.

Parsed Token	**Original (Advertising Cluster or Skill)**	**Degree in Community**	**Weighted Degree in Community**
N/A	ARTS/MATH Finance	229	1515
N/A	ARTS/MATH/SCI Chartered Professional Accounting	222	1050
vba	"vba"	2	181
pep	"pep"	2	71
continuallearn	"continuously learn", "continuous learning", "continual learning", "continuously learning", "continuous learner"	2	71
acumen	"acumen"	2	67

actuari	"actuarial", "actuaries", "actuary"	2	67
riskmanag	"risk management", "to have knowledge of risk management"	2	66
correspond	"correspondence", "corresponding", "correspond", "correspondences"	2	66
root	"root", "rooted", "rooting"	2	55
advic	"advice"	2	54
audit	"audit", "audits", "auditing", "audition"	2	46
extracurricular	"extracurricular", "extracurriculars"	2	44
valuat	"valuation", "valuations"	2	41
commenc	"commencement", "commencing", "commences", "commenced"	2	40
legal	"legally", "legal", "legals"	2	39
statement	"statements", "statement"	2	39

- **Community 8: Health** consists of 236 nodes (3.71%) and 403 edges (0.52%).
  - Public health and kinesiology students (Table 2.15 are expected by employers to understand anatomy and kinesiology. They must also be dependable and be prepared for multi-disciplinary work.

Table 2.15: The academic advertising clusters and top 15 skills from Community 8 in the skill graph, sorted by weighted degree within the community.

Parsed Token	**Original (Advertising Cluster or Skill)**	**Degree in Community**	**Weighted Degree in Community**
N/A	AHS Kinesiology	208	1212
N/A	AHS Public Health and Health Systems	195	852
kinesiolog	"kinesiology"	2	126
crimin	"criminal"	2	70
prevent	"prevention", "prevent", "prevents", "preventative", "preventing"	2	68
crisi	"crisis"	2	64
exercis	"exercise", "exercises", "exercising", "exercised"	2	53
anatomi	"anatomy"	2	40

safeti	"safety"	2	39
rehabilit	"rehabilitation"	2	39
multidisciplin	"multi disciplinary", "multidisciplinary", "multi disciplined", "multidiscipline", "multiple disciplines"	2	36
dependablil	"dependable", "dependability"	2	27
treatment	"treatment", "treatments"	2	26
safe	"safe", "safely"	2	25
mutual	"mutual", "mutually"	2	24
cognit	"cognitive"	2	24
signific	"significant", "significance"	2	22

- **Community 9: Management Sciences** consists of 56 nodes (0.88%) and 55 edges (0.07%).
  - Management Sciences students (Table 2.16) have the most unique skill requirements and are grouped by themselves. The list of skills in this community is highly specific, including Oracle Primavera P6, Robotic Process Automation (RPA), and standpipe plumbing.

Table 2.16: The academic advertising clusters and top 15 skills from Community 9 in the skill graph, sorted by weighted degree within the community.

Parsed Token	**Original (Advertising Cluster or Skill)**	**Degree in Community**	**Weighted Degree in Community**
N/A	ENG Management Sciences	55	101
p6	"p6"	1	7
rpa	"rpa"	1	4
obc	"obc"	1	3
plumb	"plumbing"	1	3
sprinkler	"sprinkler"	1	3
standpip	"standpipe"	1	3
stochast	"stochastic", "stochastics"	1	3
atlassian	"atlassian"	1	3
tese	"tesing"	1	3

catheth	“cathether”	1	3
ppap	“ppaps”, “ppap”	1	3
compliment	“compliment”	1	3
insul	“insulation”	1	3
relationshp	“relationshps”	1	3
apqp	“apqp”	1	3

Graph 2: Academic Cluster to Job Title

In addition to the 12 communities described below, 153 job titles were assigned to their own single-node community.

- 151 of these nodes corresponded to jobs that never advertised to any academic cluster (that is, they used thematic clusters only). As a result, they did not have any edges to academic clusters and could not form part of any larger community.
- The remaining 2 nodes corresponded to the job titles “Product Analyst” and “Product & Project Management Fellow”. These job titles were evenly linked to 20 and 21 academic clusters (respectively), so they did not get assigned to any particular cluster.

Note that identifying information, such as specific course numbers or employer names, is replaced with [REDACTED] in the results below for privacy reasons.

In addition to presenting a table with the top job titles or skills for each community, we also visualize the graph structure using Gephi. In each graph below, academic clusters are light grey and are labelled with identifying numbers (see image captions), while the darker grey nodes are job titles. These graph images help to demonstrate the relative size and density of each community. Each community (including the image from Gephi and table summarizing the top 15 roles) can be explored in detail below:

- **Community 0: Computer Science and Software Engineering** consists of 1286 nodes (21.87%) and 3160 edges (11.71%).
  - Employers expect these students (Figure 2.4 and Table 2.17)to fill jobs involving software, mobile, firmware, full stack, backend, and automation development.

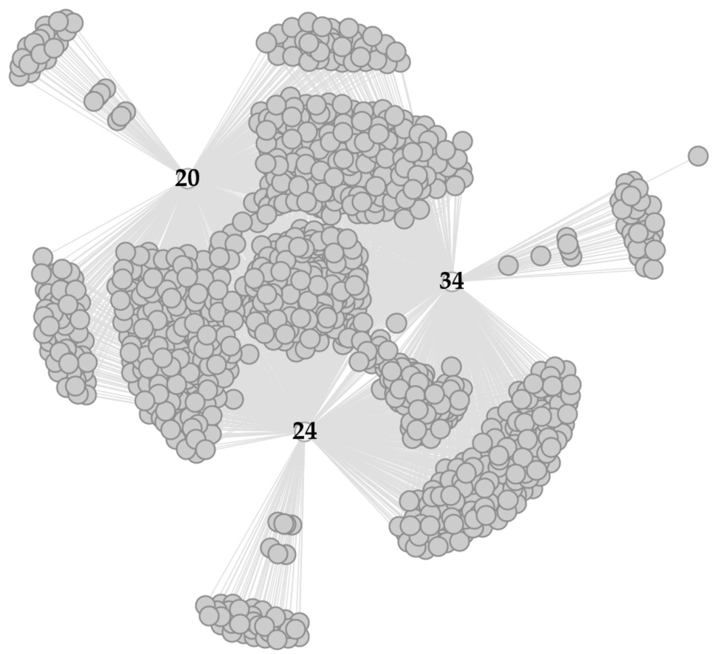


Figure 2.4: Visualization of the structure of Community 0 from the job title graph. The labelled nodes correspond to the following academic clusters: 20 (ENG Electrical and Computer Engineering), 24 (ENG Software Engineering), 34 (MATH Computer Science).

Table 2.17: The academic advertising clusters and top 15 job titles from Community 0 in the job title graph.

Node ID	**Degree in Community**	**Weighted Degree in Community**
ENG Software Engineering	1135	1966
MATH Computer Science	1123	1947
ENG Electrical and Computer Engineering	902	1547
Software Developer Intern	3	51
Software Developer, Engineering	3	48
Mobile Developer (Android)	3	30
Firmware Development	3	30

Full Stack Software Developer	3	28
Software Engineering, Marketplace & Logistics	3	27
Software Developer Co-op	3	26
Software Engineering - Analytics	3	21
Backend Developer Intern	3	21
Automation Developer	3	21
Android Engineering	3	21
Backend Software Engineering	3	21
Full Stack Developer Intern	3	20
Junior Software Engineering	3	18
Backend 'Java 8/Kotlin' Engineering	3	18

- **Community 1: Math, Statistics, Optimization, and Teaching** consists of 566 nodes (9.62%) and 1795 edges (6.65%).
  - In general, these students (Figure 2.5 and Table 2.18) are suited for roles focusing on math and finance. However, employers also expect to hire these students for teaching assistant positions, data science, and software development.

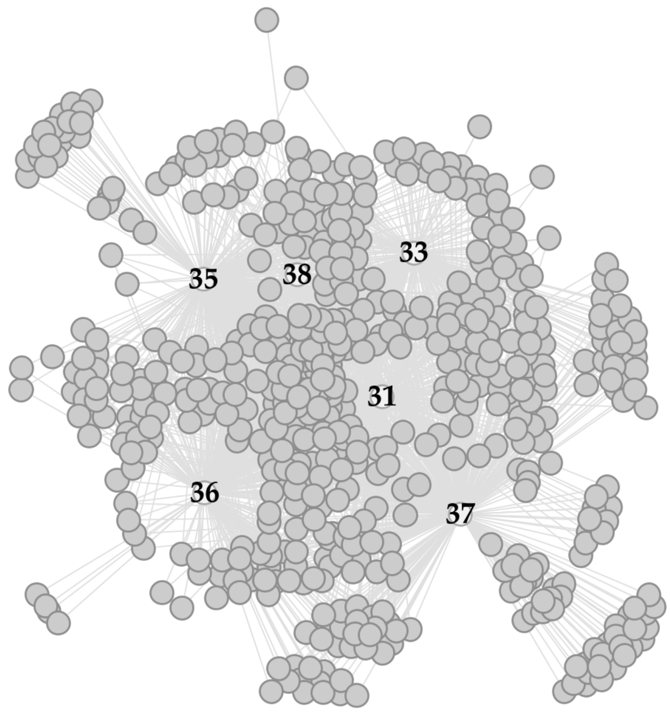


Figure 2.5: Visualization of the structure of Community 1 from the job title graph. The labelled nodes correspond to the following academic clusters: 31 (MATH Applied Mathematics), 33 (MATH Combinatorics and Optimization), 35 (MATH Computing and Financial Management), 36 (MATH Pure Mathematics), 37 (MATH Statistics and Actuarial Science), 38 (MATH Teaching).

Table 2.18: The academic advertising clusters and top 15 job titles from Community 1 in the job title graph.

Node ID	**Degree in Community**	**Weighted Degree in Community**
MATH Statistics and Actuarial Science	442	803
MATH Computing and Financial Management	374	661
MATH Applied Mathematics	353	644
MATH Pure Mathematics	304	513
MATH Combinatorics and Optimization	264	460

MATH Teaching	58	113
Actuarial Assistant	4	43
Financial Analyst	6	42
CS [REDACTED] Instructional Support Assistant	6	42
Data Engineering	5	39
Data Science	5	37
Agile Software Engineering	6	36
Risk Management	4	35
Actuarial Analyst	3	31
Full Stack Developer	5	29
Software Engineering Intern	5	27
Business Insights & Analytics Co-op	5	25
CS [REDACTED] ONLINE Instructional Support Assistant	6	24
CS [REDACTED] Instructional Support Assistant	6	24
QA Analyst	5	24
Online Learning Assistant - Math	6	24

- **Community 2: Arts and Humanities** consists of 466 nodes (7.92%) and 1730 edges (6.41%).
  - Students in the arts (Figure 2.6 and Table 2.19) are sought after as teaching and research assistants, marketing associates, designers, and technical writers. They are associated with a wide variety of roles (especially those related to customer service, user experience, or social media) but generally employers do not advertise technical roles as strongly to these students.

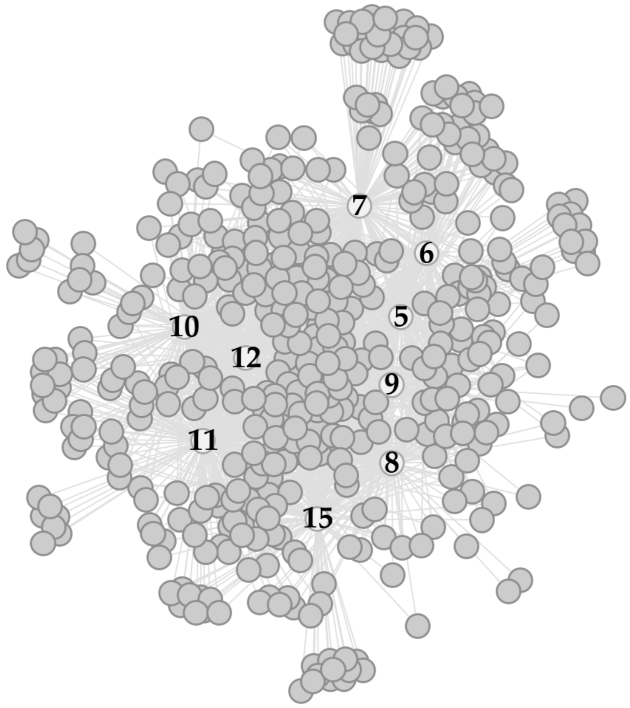


Figure 2.6: Visualization of the structure of Community 2 from the job title graph. The labelled nodes correspond to the following academic clusters: 5 (ARTS English Language and Literature), 6 (ARTS Fine and Performing Arts), 7 (ARTS Global Business and Digital Arts), 8 (ARTS Humanities), 9 (ARTS Languages and Cultures), 10 (ARTS Political Science), 11 (ARTS Social Sciences), 12 (ARTS Sociology and Legal Studies), 15 (ARTS/SCI Psychology).

Table 2.19: The academic advertising clusters and top 15 job titles from Community 2 in the job title graph.

Node ID	**Degree in Community**	**Weighted Degree in Community**
ARTS Social Sciences	281	514
ARTS English Language and Literature	240	472
ARTS Global Business and Digital Arts	220	443
ARTS Humanities	205	384

ARTS Sociology and Legal Studies	177	355
ARTS Political Science	180	340
ARTS/SCI Psychology	165	318
ARTS Languages and Cultures	148	278
ARTS Fine and Performing Arts	114	224
Teaching Assistant	8	254
Research Assistant	9	71
Developer	9	36
Group Home Support Care Worker	6	36
Course and Technical Support Assistant - Arts	9	36
Product Designer	9	35
Bilingual (French & English) Customer Service Representative	9	32
Sales & Marketing Fellow	6	32
Technical Writer	4	31
Special Projects Assistant	9	30
Junior Officer	7	28
Marketing Coordinator	5	28
Junior Policy Analyst	7	27
Online Learning Assistant - Arts	9	27
Marketing & Customer Service Associate	9	26

- **Community 3: Business** consists of 498 nodes (8.47%) and 1317 edges (4.88%).
  - Business students (Figure 2.7 and Table 2.20) are expected to fill roles associated with operations, customer success, sales, marketing, and human resources.

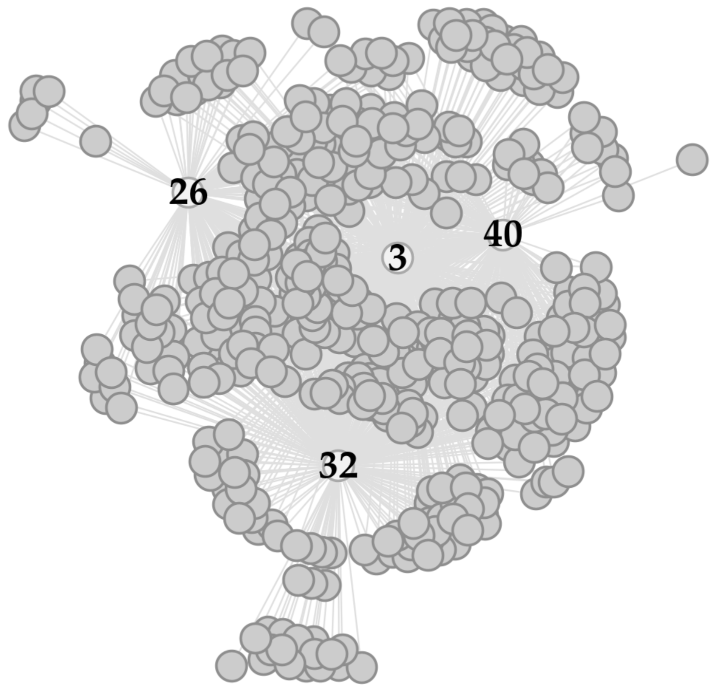


Figure 2.7: Visualization of the structure of Community 3 from the job title graph. The labelled nodes correspond to the following academic clusters: 3 (ARTS Business), 26 (ENV Business, Enterprise and Development), 32 (MATH Business), 40 (SCI Business).

Table 2.20: The academic advertising clusters and top 15 job titles from Community 3 in the job title graph.

Node ID	**Degree in Community**	**Weighted Degree in Community**
ARTS Business	439	710
MATH Business	388	610
SCI Business	278	444
ENV Business, Enterprise and Development	212	337
Supply Chain Associate	3	27
Business Systems Analyst	4	23
Business Operations	4	21

Customer Success Specialist	4	21
Operations Technician	4	21
Business Development Assistant	4	20
Business Analytics	3	19
Inside Sales Representative	3	16
Business Development and Marketing	4	16
English Language Institute Programs Assistant	4	16
Business Systems Analyst, Co-op	4	15
On-Site Support Technician Co-op	3	14
Pre-Sales International Markets Coordinator	4	13
Digital Communications Assistant	4	13
Operations Agent	3	12

- **Community 4: Mechanical, Biomedical, and Nanotechnology Engineering** consists of 664 nodes (11.29%) and 1243 edges (4.60%).
  - Although employers expect these engineering students (Figure 2.8 and Table 2.21) to fill engineering roles related to their programs, they also might fill roles involving software development, quality assurance, and design. In particular, when looking past the top 15 roles, titles such as "iOS Developer", "Game Programmer", and "Junior Developer" demonstrate that coding skills are important for all engineers (not only those in programs explicitly related to software or hardware).

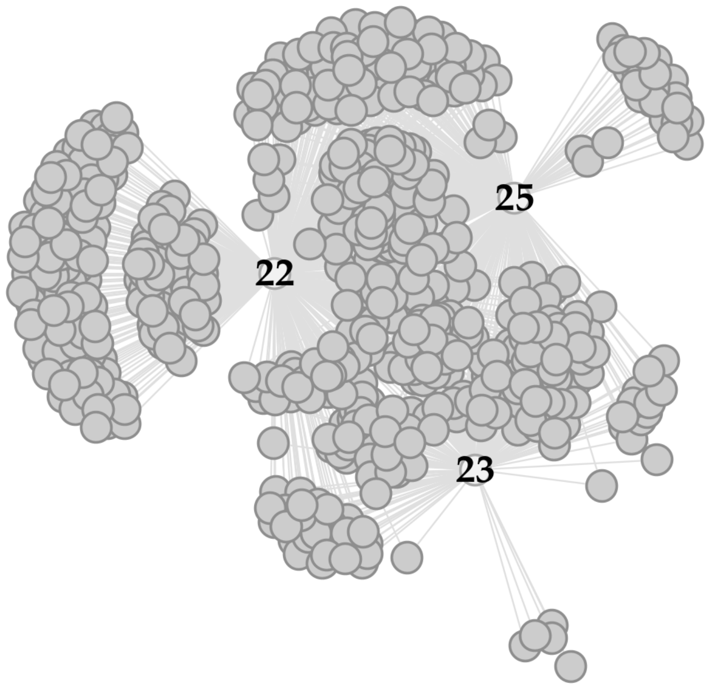


Figure 2.8: Visualization of the structure of Community 4 from the job title graph. The labelled nodes correspond to the following academic clusters: 22 (ENG Mechanical and Mechatronics Engineering), 23 (ENG Nanotechnology Engineering), 25 (ENG Systems Design and Biomedical Engineering).

Table 2.21: The academic advertising clusters and top 15 job titles from Community 4 in the job title graph.

Node ID	**Degree in Community**	**Weighted Degree in Community**
ENG Mechanical and Mechatronics Engineering	604	1095
ENG Systems Design and Biomedical Engineering	425	780
ENG Nanotechnology Engineering	214	353
Hardware Engineering	3	27
Mechatronics Engineering	3	22
Electrical Engineering	3	21

Full Stack Web Developer	3	20
Mechanical Designer	2	19
Hardware Developer	3	18
Quality Project Coordinator	3	18
Product Engineering	3	17
Test Automation Engineering	3	16
Fullstack Developer	3	16
Embedded Software Developer	2	15
Design and Manufacturing Support Technician	3	15
Quality Assurance Engineering - [REDACTED]	2	14
Software Development	3	13
Product Development	2	13

- **Community 5: Environment and Geological Sciences** consists of 409 nodes (6.95%) and edges (4.49%).
  - This is another case where employers appear to expect tech skills from students not explicitly enrolled in computer science or software engineering. For example, these students (Figure 2.9 and Table 2.22) are targeted for compiler software engineering, AI development, and big data development roles in addition to geospatial intelligence and geomatics.

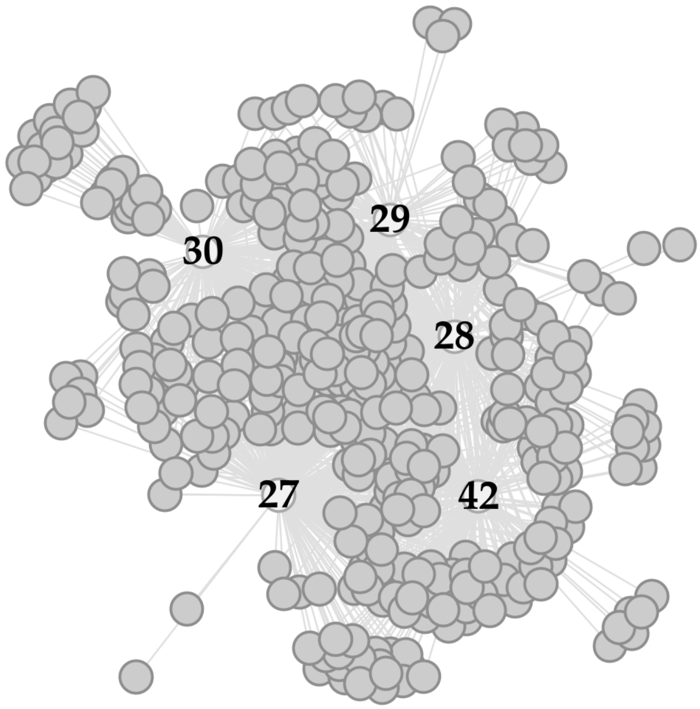


Figure 2.9: Visualization of the structure of Community 5 from the job title graph. The labelled nodes correspond to the following academic clusters: 27 (ENV Environment, Resources and Sustainability), 28 (ENV Geography and Environmental Management), 29 (ENV Geomatics), 30 (ENV Planning), 42 (SCI Earth, Environmental and Geological Sciences).

Table 2.22: The academic advertising clusters and top 15 job titles from Community 5 in the job title graph.

Node ID	**Degree in Community**	**Weighted Degree in Community**
ENV Geography and Environmental Management	302	548
ENV Environment, Resources and Sustainability	298	494
ENV Planning	216	437
SCI Earth, Environmental and Geological Sciences	230	401

ENV Geomatics	167	313
Student Planner	4	41
GIS Technician	5	32
Compiler Software Engineering	5	30
Research Analyst	5	28
Junior Environmental Enforcement Analyst	5	27
GIS Assistant	4	24
Mechanical Engineering	5	20
Geomatics Technician	5	20
GIS Analyst	4	20
AI Developer	5	20
Big Data Platform Developer	5	20
Distribution & Geospatial Intelligence Analyst	5	16
Junior GIS Assistant	4	16
Course and Technical Support Assistant - Engineering	4	16
Software Engineering	5	15

- **Community 6: Finance, Accounting, and Economics** consists of 455 nodes (7.74%) and 873 edges (3.23%).
  - These students (Figure 2.10 and Table 2.23) are most targeted for business, finance, and accounting roles. There is a notable absence of IT-related roles in this community, perhaps indicating that finance does not rely as heavily on tech skills.

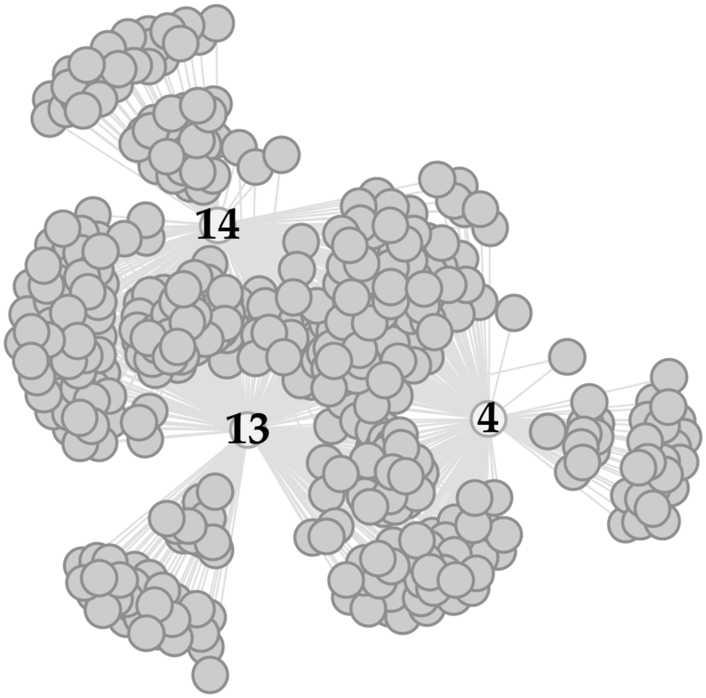


Figure 2.10: Visualization of the structure of Community 6 from the job title graph. The labelled nodes correspond to the following academic clusters: 4 (ARTS Economics), 13 (ARTS/MATH Finance), 14 (ARTS/MATH/SCI Chartered Professional Accounting).

Table 2.23: The academic advertising clusters and top 15 job titles from Community 6 in the job title graph.

Node ID	**Degree in Community**	**Weighted Degree in Community**
ARTS/MATH Finance	355	698
ARTS/MATH/SCI Chartered Professional Accounting	292	657
ARTS Economics	226	379
Business Finance Co-op	3	72
Staff Accountant	3	64
Assurance & Accounting Student	2	44

Business Analyst	3	35
Accounting Associate	3	27
Assurance (CPA)	2	21
Junior Accountant	2	20
Finance, Intern (CPA Stream)	2	20
Property Accountant	3	18
Finance	3	17
Accounting Intern	2	16
Accounting Analyst	3	16
Tax Analyst	3	15
Associate	3	15
Business Finance	3	14

- **Community 7: Architecture and Environmental Engineering** consists of 448 nodes (7.61%) and 711 edges (2.63%).
  - These students (Figure 2.11 and Table 2.24) are generally expected to fill architectural, project management, construction, and online learning assistant roles. Although IT-related roles do not appear in the top 15, there are roles such as "Data Software Engineering" and "Software Developer" within the top 30.

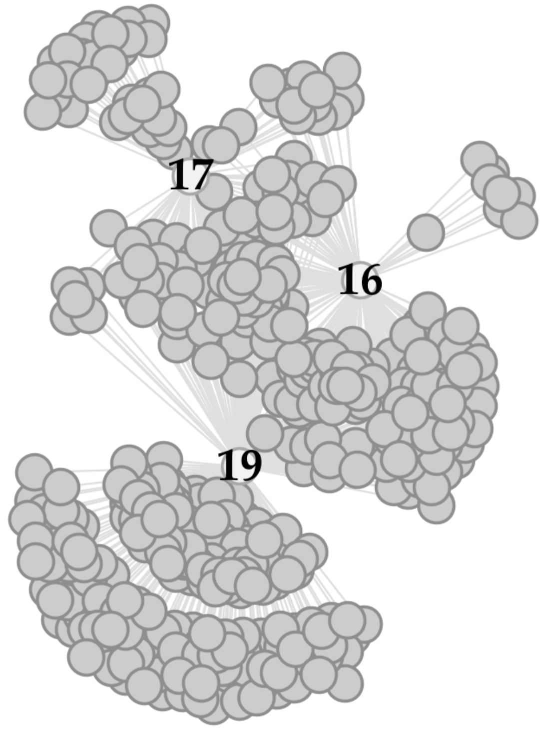


Figure 2.11: Visualization of the structure of Community 7 from the job title graph. The labelled nodes correspond to the following academic clusters: 16 (ENG Architectural Engineering), 17 (ENG Architecture), 19 (ENG Civil, Environmental and Geological Engineering).

Table 2.24: The academic advertising clusters and top 15 job titles from Community 7 in the job title graph.

Node ID	**Degree in Community**	**Weighted Degree in Community**
ENG Civil, Environmental and Geological Engineering	386	722
ENG Architectural Engineering	196	344
ENG Architecture	129	294
Architectural Assistant	3	66
Field Engineering/Project Coordination	2	30

Project Coordinator	3	24
Engineering Ideas Clinic Research Assistant	3	24
Architectural Intern	3	16
Senior Online Learning Assistant - Faculty of Engineering	3	15
Junior Project Coordinator	3	12
Civil Engineering Assistant	1	12
Online Learning Assistant - Engineering	3	12
Civil Engineering	2	12
Project Coordinator and Estimator	3	12
Assistant Project Manager	3	11
Planning Project Coordinator	2	10
Engineering Sales Assistant	1	10
Construction Inspection Assistant	2	10

- **Community 8: Biology, Chemistry, and Pharmacy** consists of 442 nodes (7.52%) and 740 edges (2.74%).
  - Students in these fields (Figure 2.12 and Table 2.25) are targeted for varied roles including "Production Worker", "Math Tutor", and "Lab Assistant". They are often considered a good fit for research and development or educational positions.

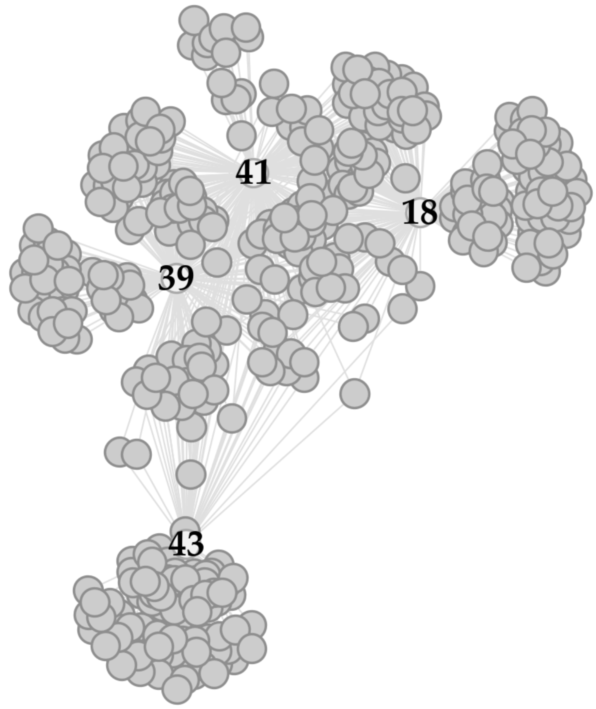


Figure 2.12: Visualization of the structure of Community 8 from the job title graph. The labelled nodes correspond to the following academic clusters: 18 (ENG Chemical Engineering), 39 (SCI Biological Sciences), 41 (SCI Chemical Sciences), 43 (SCI Pharmacy).

Table 2.25: The academic advertising clusters and top 15 job titles from Community 8 in the job title graph.

Node ID	**Degree in Community**	**Weighted Degree in Community**
SCI Chemical Sciences	211	391
ENG Chemical Engineering	193	361
SCI Biological Sciences	190	348
SCI Pharmacy	146	322
Pharmacy Student	2	73
Clinical Research Assistant	3	29

Production Worker	3	21
Math Tutor	4	20
Lab Assistant	3	19
Laboratory Technician	3	17
Micro Engineering Dynamics Automation Lab	2	16
Research Assistant - Chemical Science	3	15
Food Technologist	3	15
Regulatory Affairs Associate	3	13
Product Development Assistant	3	13
Investment Banking Analyst	4	12
R&D Assistant	3	12
Sensors Research Samurai	3	12
Technical Support to R&D	3	12

- **Community 9: Health, Recreation, and Kinesiology** consists of 241 nodes (4.10%) and 393 edges (1.46%).
  - Unsurprisingly, most of the roles that target these students (Figure 2.13 and Table 2.26) are related to health care, animal care, and community oriented roles. Many of the positions also involve receptionist or customer service work that centers around interfacing with other people.

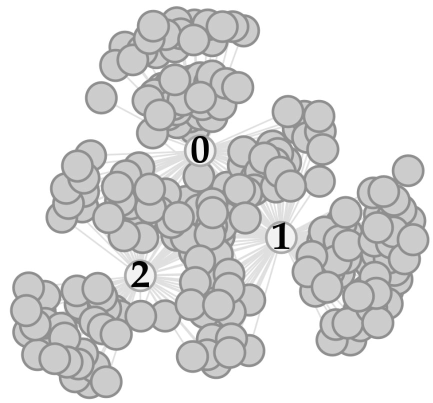


Figure 2.13: Visualization of the structure of Community 9 from the job title graph. The labelled nodes correspond to the following academic clusters: 0 (AHS Kinesiology), 1 (AHS Public Health and Health Systems), 2 (AHS Recreation and Leisure Studies).

Table 2.26: The academic advertising clusters and top 15 job titles from Community 9 in the job title graph.

Node ID	**Degree in Community**	**Weighted Degree in Community**
AHS Kinesiology	140	352
AHS Public Health and Health Systems	141	293
AHS Recreation and Leisure Studies	112	231
Pedorthic Assistant	2	32
Administrative Assistant	3	31
Animal Care & Research Assistant	3	30
[REDACTED] Community Assistant - Western or Atlantic Canada	3	27
Rehabilitation Assistant	3	22
Quality Assurance Specialist	3	18
Medical Receptionist	3	18
Optometric Assistant	3	18
[REDACTED] - Vocational/Residential Assistant	3	15
Customer Service Associate	2	14

Service Desk Specialist	3	13
Health Care Administrator	2	10
Assistant Physical Training Instructor	3	9
Health Planning and Performance Analyst	3	9
Move Your Mind Coordinator	3	9

- **Community 10: Management Sciences** consists of 148 nodes (2.52%) and 147 edges (0.54%).
  - This community (Figure 2.14 and Table 2.27) includes only one academic cluster (ENG - Management Sciences) and includes roles related to logistics, reliability, and maintenance . However, there are also numerous IT-related roles, including software development and web application development. Note that in this and the following community, there is only one academic advertising cluster. This leads the visualization to have a different look compared to the previous ones. Because there are edges only between advertising clusters and skills, the Force Atlas algorithm used to visualize these communities attempts to disperse the skills as far from each other as possible (as they are less strongly connected to each other). This results in the skills being visualized in a circular pattern evenly around the advertising cluster.

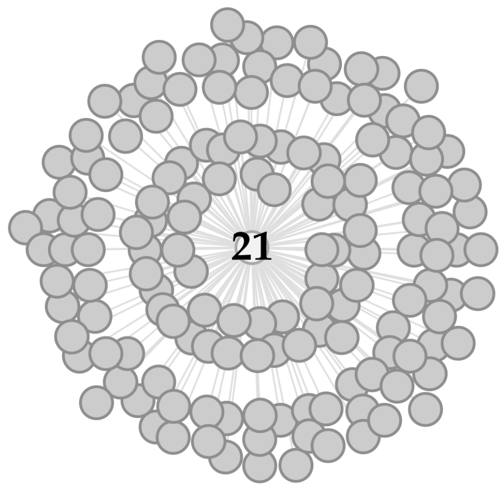


Figure 2.14: Visualization of the structure of Community 10 from the job title graph. The labelled nodes correspond to the following academic clusters: 21 (ENG Management Sciences).

Table 2.27: The academic advertising clusters and top 15 job titles from Community 10 in the job title graph.

Node ID	**Degree in Community**	**Weighted Degree in Community**
ENG Management Sciences	147	213
Logistics Apprentice	1	4
Reliability Engineering	1	4
Business Intelligence Coordinator	1	3
Digital & Payments	1	3
Engineering Maintenance Student	1	3
Junior Designer - Mechanical	1	3
Junior Manufacturing Engineering	1	3
Maintenance Scheduling and Planning Support	1	3
Online Software Support Associate	1	3
Product Management	1	3
Project Administrator - Pumps	1	3

Research Assistant - Management Sciences	1	3
Software Developer (Cloud)	1	3
Technical Implementation & Support Specialist	1	3
Web Application Developer	1	3

- **Community 11: Physics** consists of 105 nodes (1.79%) and 104 edges (0.39%).
  - Jobs targeting physics students form their own community (Figure 2.15 and Table 2.28), where roles often involve medical physics, 2D image sensing, and other physics related concepts. However, there are also some other trends: research and development, engineering, and IT (including software engineering and machine learning).

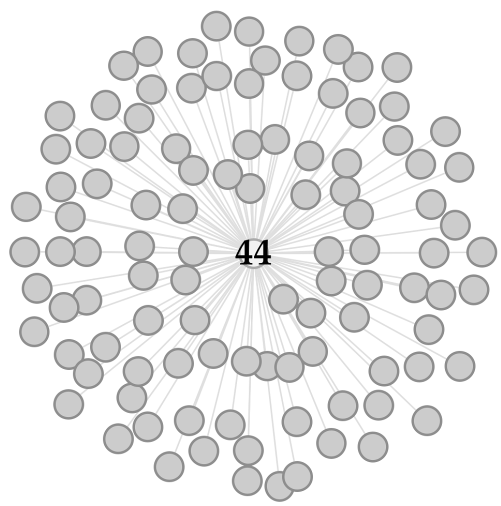


Figure 2.15: Visualization of the structure of Community 11 from the job title graph. The labelled nodes correspond to the following academic clusters: 44 (SCI Physics).

Table 2.28: The academic advertising clusters and top 15 job titles from Community 11 in the job title graph.

Node ID	Degree in Community	Weighted Degree in Community
SCI Physics	104	160
Nuclear Innovation Engineering	1	8
[REDACTED] Global Markets Quantitative Analyst	1	6
R&D Associate	1	4
R&D Co-op Student	1	4
Scientific Programmer for Satellite Monitoring of Air Pollution	1	4
Software Engineering Co-op	1	4
[REDACTED] Global Markets Developer	1	3
Defence Analytics Assistant	1	3
Medical Physics Assistant	1	3
Numerical Environmental Modelling	1	3
Optical Algorithms Engineering	1	3
Undergraduate Research Assistant on [REDACTED] Experiments	1	3
2D Image Sensor Test	1	2
Agricultural Environmental Sustainability Analyst	1	2
Engineering & Computer Science Trainee in Advanced Telecommunications Research & Development	1	2

2.7 Discussion

2.7.1 Types of Available Co-op Jobs

When analyzing the top 25 most frequent terms in the required skills section (Table 2.3), we find several interesting tokens. First, communication is the most frequently mentioned term, appearing in 51.8% of filled job descriptions' required skills sections. This may indicate that employers value communication skills strongly in a remote work context, as in

2021, many jobs were performed remotely due to the COVID-19 pandemic. Second, many frequent skills are related to information technology (IT), including "develop" (31.7%), "software" (28.9%), "data" (20.4%), and "python" (18.2%).

There are also notable differences between the skills required for junior and senior students (Tables 2.4 and 2.5). In particular, technology-related skills are more common for senior students (science, Python, data, C++, system, and algorithm), indicating that employers expect these students to have concrete knowledge and skills, especially in tech. This aligns with the findings of Chopra & Golab [13]. However, notably, our analysis is not specific to IT-related jobs, so it is interesting that senior students are still expected to possess significantly more tech skills. For junior students, people skills are crucial. In particular, customer service, social media, and communication are all required more for junior than for senior jobs.

When analysing the k-means clusters in Table 2.6, we notice several interesting trends. First, the Communication cluster is assigned to 21.5% of Junior jobs but only 13.6% of senior jobs, while many tech-related clusters (e.g., Programming and Software Development) make up a larger portion of senior jobs. This is aligned with the findings from Table 2.5, where senior students are more associated with technical skills.

As the majority of jobs in 2021 were still remote, it is interesting to observe that one cluster is dedicated to skills related to Remote Work, including Zoom, teleconferencing, and the ability to multitask and handle uncertainty. Only two of the clusters focus on soft skills (Communication, Inclusivity and Integrity), while the rest are largely focused on hard skills. Although many of the remaining clusters are tech-related, Graphic Design, Finance, and Vehicle Access are notable exceptions. It is clear that jobs in these areas require a specific skill set, or, in the case of Vehicle Access, perhaps these tokens typically appear together when a job requires driving skills. It is also interesting to note the clusters where there are more intermediate jobs than either junior or senior: Finance, Vehicle Access, and Inclusivity and Integrity. This may indicate that junior students are less qualified for these jobs, while senior students may be overqualified or may be more interested in other types of jobs.

2.7.2 Employer Expectations of Job Roles and Required Skills

2.7.3 Academic Cluster and Required Skill Communities

The Arts and Humanities community (Table 2.7) is surprisingly similar to the Research & Design cluster found using k-means in the previous analysis. Perhaps this indicates that employers expect Arts students to possess excellent communication, creativity, and people skills over more concrete technical skills. This is in contrast to the Computer Science, Software, and Hardware community (Table 2.8). Students in these traditionally IT-related fields are expected by employers to possess significant technical knowledge. Many other student groups also require highly specific technical knowledge, including the Management Sciences (Table 2.16) and General Sciences groupings. However, it is also common for students to be expected to possess both technical skills and soft skills. For example, students in Health (public health and kinesiology) are expected to be both knowledgeable in their field and be dependable and open to multi-disciplinary work. When examining the business community (Table 2.9), the source of the term diplomat was unclear. However, through manual inspection, we found that these students are expected to be diplomatic team players who carefully consider the interests of stakeholders.

Data science, artificial intelligence, and machine learning skills appear to be in demand for students in a diverse range of fields: Computer Science, Software, and Hardware Engineering (Table 2.8, as well as Architecture, Pharmacy, and Teaching (Table 2.12. After manually inspecting Community 5: Architecture, Pharmacy, and Teaching, we found that these students are linked primarily by the AI-related skills required for their jobs. Jobs targeting these students typically involve teaching assistant, research assistant, and software development positions with reliance on technical skills, programming languages such as Python and Javascript, and machine learning tools.

Academic Cluster and Job Title Communities

In general, when analysing these communities, it is clear that some program groupings are expected by employers to be more interdisciplinary than others. For example, while employers have a clear and narrow set of expectations for computer science and software engineering students, the roles targeting science, arts, and business students are considerably more varied. Another notable point is the prevalence of IT-related roles targeted to students not in traditionally IT-related programs. For example, students in biology, chemistry, and pharmacy (Table 2.25) are linked to “Sensors Research Samurai”, in environment and geological sciences (Table 2.22) are linked to “AI Developer”, and students in business (Table 2.20 to “Digital Communications Assistant”. Students in math and statistics (Table 2.18) are expected to fill roles in data science more strongly than any other group, showing that employers expect these skills from this group in particular. However, students in finance form one notable exception. They do not appear to be targeted for software/hardware development, AI, machine learning, or data science related roles. This could indicate that the finance sector has been slow to adopt modern technology. Alternatively, employers might simply expect finance students to have a specific financial skill set, where tech skills are secondary.

When examining the structure of the graph communities, it is possible to see which academic programs are most strongly linked by the types of roles expected by employers. For example, in Community 0 (Table 2.4), three academic clusters are grouped together. However, we can see more job titles shared between "ENG Software Engineering" and "MATH Computer Science", indicating that employers expect more similarities between these programs than with "ENG Electrical and Computer Engineering". While all three programs are grouped together, this graph structure allows us to see that hardware engineers are expected to fill slightly different roles than software engineers or computer science students. Similarly, there are many job titles linked specifically to "ENG Mechanical and Mechatronics Engineering" in Figure 2.8, differentiating it from nanotechnology and biomedical engineering. In Figure 2.11, we see "ENG Civil, Environmental and Geological Engineering" differentiated in a similar way from architectural engineering programs. "SCI Pharmacy", in Figure 2.12, is also shown to be distinct from biology and chemistry programs. In contrast, employers expect health students to fill very similar roles, as evidenced by the even distribution of job titles in Figure 2.13.

Finally, it was interesting to note that similar groupings of academic clusters emerge for both skill and job title graphs. Despite this fully unsupervised analysis, both graphs cluster Arts students and Math students together, although a few specific programs have moved to other clusters. Although the number of clusters slightly different, there is evidence that employers do believe that these academic programs are related and expect students from these programs to fill similar roles with similar skills.

2.8 Practical Implications

Understanding the expectations of employers for students from different academic backgrounds is important for students, institutions, and researchers. Current students can use this information to understand the roles employers expect them to fill and the skills that are most important for them to acquire. Similarly, institutions can update and develop new courses to ensure that they are teaching their students skills that are most sought-after in their fields and ensure that students are informed about the roles they can expect to fill. Prospective students might use this knowledge to inform their decisions about which programs to pursue, based on the types of jobs they might acquire later in their careers and the skills needed for those positions. Finally, even outside the application of job descriptions, researchers could leverage this graph methodology to co-cluster pairs of interesting attributes in other datasets. For example, a researcher interested in exploring the use of hate speech on the popular social media platform Reddit might co-cluster types of hate speech with the communities that use them (adding an edge between a type of hate speech and a Reddit community each time its members use that speech in their discussions). This graph methodology could be used to determine which communities are most strongly associated with each type of hate speech, by finding groupings of communities and types of hate speech simultaneously. The result would be groups of related communities, joined with the most strongly associated types of hate speech. Ultimately, this methodology is appropriate for any case where the relationship between two variables can be explored through co-clustering.

Chapter 3

Student success in co-operative education: A comparison of remote and in-person workplace performance evaluations

3.1 Introduction

During the COVID-19 pandemic, many positions switched partially or entirely to remote work [35]. This necessary increase in the number of remote positions has normalized remote work, with many current positions still offering remote options. As a result, it is important for employers, institutions, and students to understand what makes students successful in remote co-op positions and how this may differ from success in in-person positions. This motivates the research questions studied in this chapter:

1. Which skills, abilities, and attributes make co-op students successful?
2. How do the most valued skills differ for remote and in-person positions?
3. Where should students improve to be most successful in remote and in-person positions?

In this secondary data analysis, we use over 23,000 remote and over 22,000 in-person co-op performance evaluations collected by the University of Waterloo to uncover key differences between remote and in-person evaluations. With these insights, we recommend skills that will enable students to be more competitive in the co-operative job market.

The next section reviews existing literature on remote and in-person co-operative work positions, followed by an overview of the secondary data used for this study. The results of analyzing two types of feedback written by workplace supervisors at the end of co-operative work terms are then presented. Finally, we discuss practical implications and make recommendations for ways that students, employers, and institutions can be more successful in co-operative education programs.

3.2 Related Work

3.2.1 Employers' Expectations of In-person Co-op Students

Closely related works on understanding employers' expectations of (in-person) co-op students were conducted by Nevison et al. [42] and Coll et al. [16]. Nevison et al. surveyed 376 co-op employers who stated that the most important workplace competencies are relevant work experience and the quantity of work done. Coll et al. surveyed 172 co-op employers who suggested that the most important attributes of new graduates entering the workforce are willingness to learn, teamwork, initiative, and analytical thinking. To the best of our knowledge, this study is the first to explore employers' expectations in the context of remote co-op, based on the results of over 23,000 remote performance evaluations compared with over 22,000 evaluations from in-person positions.

3.2.2 Employers' Perceptions of In-person Co-op Students

There are also studies of co-op employers' perceptions of student competencies (prior to remote working), and whether these perceptions change based on gender and seniority [14, 15, 34]. One study found that in-person co-op students were rated most highly on response to supervision, ability to learn, and interpersonal skills, and the lowest on leadership and creativity [34] Studies also found that female students and students with more work experience were rated higher in almost all evaluation categories [14, 15, 34]. Instead of analyzing co-op employers' perceptions of student competencies, we analyze co-op employers' expectations of remote employees to understand if similar expectations persist before and after the switch to remote co-op.

3.2.3 Remote Co-operative Education

Recent work on remote co-op has focused on making remote work placements beneficial for students [1, 9, 19] and the challenges arising from the transition from in-person to remote co-op [9, 26, 31, 50]. In remote work, students must infer workplace culture and supervisor expectations through short virtual interactions where they previously had the opportunity to immerse themselves in a physical workplace. Studies note that this is an important factor for success in remote co-op positions [19, 50], showing the importance of understanding employer expectations. Thus, this study fills a gap by examining co-op employers' expectations of remote student employees, which may aid in the success of remote co-op programs.

3.2.4 Post-Graduate Employment

In the broader context of post-graduate employment, studies have focused on remote employees' experiences, perceptions, expectations, and well-being [3, 6, 21, 23, 28, 48, 52, 63]. Past studies found that satisfaction and productivity in remote employees increased with strong communication, trust with one's supervisor, and clarity of evaluation and feedback [4, 51]. However, no previous work studying employers' expectations from and satisfaction with remote workers was found by the authors. Consequently, we focus on employer expectations in a remote context, rather than the experiences of employees.

3.3 Data Overview

This analysis was performed using undergraduate co-op performance evaluations collected by the University of Waterloo, a large North American university. The evaluations were completed by workplace supervisors at the end of work terms. In our dataset, most work placements (91.5%) were four months in length, while the rest were eight months (7.8%) or two months (0.7%) duration.

Approximately half of the evaluations correspond to positions from 2019, while the remaining evaluations were from positions in 2021. With high confidence, all or almost all positions in 2019 were believed to be in-person, while positions in 2021 were believed to be remote. This was based on publicly available data on government lockdown restrictions and information from employees working directly in the University of Waterloo's co-op program. Therefore, for the purposes of this analysis, positions in 2019 were treated as in-person and positions in 2021 as remote. Because the sudden switch to remote work in winter 2020 led to many positions being disrupted without warning (partially in-person and partially remote), no data from 2020 was included in the analysis. In total, anonymized student data and employer evaluations for 22,134 in-person and 23,417 remote work placements were incorporated into the analysis. For each placement, the dataset includes:

- **Overall Performance Rating:** on a 7-point scale, with the categories labelled as (in order from 1 to 7) unsatisfactory, marginal, satisfactory, good, very good, excellent, and outstanding.
- **Supervisor's Comments:** optional comments written by the supervisor to explain their choice of evaluation rating for the given student.
- **Recommendation Comments:** optional comments written by the supervisor to discuss areas for improvement and skills the student should develop.

Demographic information was also included for each evaluation. Students were evaluated from eight faculties: Engineering (39.4%), Mathematics (25.4%), Science (9.1%), Arts (13.2%), Environment (6.1%), Health (2.9%), Applied Health Sciences (2.0%), and Interdisciplinary Studies (1.9%). Notably, nearly three-quarters (73.9%) of evaluations corresponded to students in Science, Technology, Engineering, and Mathematics (STEM) related fields. Of these student evaluations, 42.4% corresponded to students identifying as female, 56.6% male, and 1% other. Finally, each evaluation was also labelled with the number of work terms completed by the given student (including the one currently being evaluated). This information can be found in Table 3.1. Our preliminary investigations comparing different demographic groups (including faculty, gender, and number of completed work terms) did not uncover any significant differences, so these were not used as control variables.

In 2020, a new evaluation form was introduced for co-op employers to evaluate their students, following the development of the Future Ready Talent Framework [49], a conceptual model of the most relevant skills for co-op students. Although the sections for written comments did not change, there were changes in wording to other parts of the form. Specifically, different terminology was used for describing students' skills. For example, a skill called "Teamwork" in the 2019 form was referred to as "Collaboration" in the 2021 form. This change is mentioned here because it may have affected the words used by supervisors to describe students' skills, producing superficial differences between the 2019 and 2021 evaluations that are not related to the switch from in-person to remote work. These differences are explored in more detail in the discussion section.

Table 3.1: The number of work terms completed by each student, at the time of evaluation. Because evaluations occur at the end of the work term, students on their first work term have completed one work term and are assigned a work term number of one.

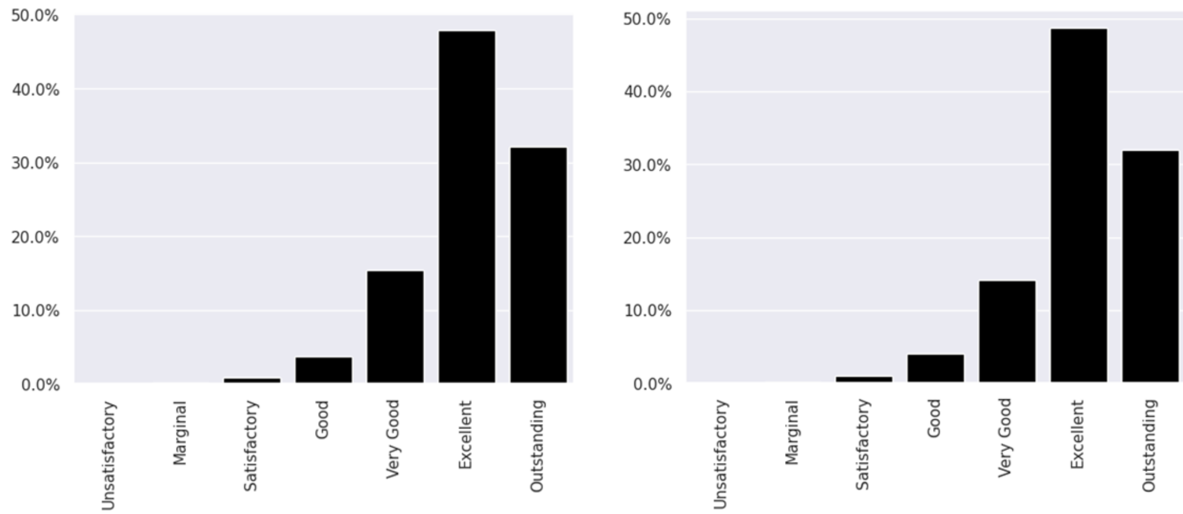
Work Term Number	Percentage of Student Evaluations
1	22.2%
2	20.4%
3	18.5%
4	17.0%
5	12.9%
6	8.1%
7+	0.9%

The overall performance rating distribution for all student placements is shown in Figure 3.1a for 2019 and Figure 3.1b for 2021. In 2021, the majority (80.7%) of all students received a rating of excellent or outstanding. Only nine students (0.04%) in the 2021 data received a rating of unsatisfactory, failing to meet employer expectations. The distribution was similar in the 2019 evaluations, indicating that students are not perceived as being either more or less successful by their employers in remote positions.

3.4 Methods

3.4.1 Logistic Regression to Determine What Makes Students Successful

We investigated supervisor comments using a logistic regression classifier [41] to reveal aspects of successful co-op students. Generally, unstructured text data such as a supervisor's comments requires preprocessing. In this dataset, supervisors often left short comments that did not contain useful information about a student's skills, such as "comments included" or "good job." For this analysis, all comments shorter than 25 characters were removed. This threshold was selected through manual inspection to ensure that most non-useful comments would be removed. This process resulted in 3,841 records being removed from the analysis (out of 23,417 total evaluations) in the remote data and 2,246 (out of 22,134) from the in-person data.



(a) 2019

(b) 2021

Figure 3.1: Distribution of overall performance ratings in student evaluations. Figure a corresponds to the ratings from 2019 evaluations, while Figure b is from 2021.

The Natural Language Toolkit (NLTK) word tokenizer [43] was used to divide each sentence into individual words, which were converted to lowercase and stripped of punctuation. Stop words are common words (such as the, and, or should) which are often deliberately ignored by classifiers because they do not contain important meaning (HaCohen-Kerner et al., 2020). Thus, words from the NLTK English stop word list were removed before applying the NLTK Snowball Stemmer [44] to convert words with similar meanings to a common representation. For example, the Snowball Stemmer converted both ‘innovated’ and ‘innovator’, which are both related to the skill of innovation, to the term innov.

The next step in preprocessing was to convert the text to a vector representation. Each comment produced a vector, a list of n coordinates corresponding to each of the n words in the vocabulary. For a given comment, the ith coordinate was equal to one if the word was present in that comment, and zero otherwise. Each coordinate was then multiplied by the total number of comments and divided by the number of comments in which the corresponding word appeared. For example, if the word ‘design’ appeared in 1000 out of 22000 comments, the coordinate corresponding to ‘design’ would be 22 if ‘design’ appeared in the given comment, and 0 otherwise. This is a measure of how rare the given word is in the set of comments overall. This was done to place more emphasis on rare words that are present in only a few documents (such as a student’s skills) instead of words that appear in almost every document and therefore will not help distinguish outstanding students. Finally, both unigrams (single words) and bigrams (pairs of consecutive words) were included in the analysis as features for logistic regression.

Using these data, a logistic regression classifier was applied to the 2019 and 2021 datasets separately to make a binary prediction: whether or not a student received a rating of outstanding, the highest possible evaluation category. As these were the most successful students according to co-op employers, the classifier was expected to learn what differentiates these outstanding students from others. Next, the classifiers were evaluated using 10-fold cross validation, a standard technique for measuring classification accuracy. In this case, accuracy is defined as the number of correct predictions divided by the total number of predictions. For example, a perfect classifier would achieve an accuracy of 100% by correctly predicting, for all students, whether or not they received a rating of outstanding.

Logistic regression classifiers were used for this analysis because they are easy to implement and interpret. By examining the model weights after training a logistic regression classifier, it is possible to understand which words in the original comments were most strongly associated with outstanding and non-outstanding students. Therefore, 100 terms with the highest model weights and 100 terms with the lowest weights were identified, corresponding to terms that most strongly predicted a label of outstanding and not outstanding, respectively. These terms were selected out of 150,000 total words and pairs of words from each dataset because they best differentiated between the two groups of students. This was done using the eli5 Python module (Korobov & Lopuhin, 2017). To isolate the skills and attributes of the most successful students, only the top terms that were not common English words are reported.

This analysis was applied separately to the data from 2019 and 2021, and the top and bottom terms compared to detect differences between the most important skills for each group of students.

3.4.2 Supervisor Recommendations: Frequency Analysis

Although a logistic regression model can easily identify which skills differentiate recommendations in remote and in-person settings, these skills are not necessarily the most frequent. For example, remote supervisors may recommend improving a skill such as independence more often than in-person supervisors, but this may not be a common recommendation overall. Therefore, to understand which improvements were most recommended by supervisors, a frequency analysis was performed on the supervisor recommendations text field. Preprocessing was performed in the same way as for the logistic regression analysis described above using the NLTK toolkit to convert the text to lowercase, remove punctuation and stop words, and apply a stemmer to each individual word.

When examining the most frequent words individually, it is difficult to extract meaningful skills due to lack of context. For example, if a frequent word is 'ethic', we cannot be certain if it refers to a student's work ethic or to their ability to act ethically towards colleagues, superiors, and customers. However, by examining pairs of words, it is possible to identify skills with more certainty. Therefore, words were counted in overlapping pairs called *bigrams*. For example, in the phrase "problem solving skills", the bigrams "problem solving" and "solving skills" would be counted separately. After preprocessing, similar phrases such as "problem solving" and "problem solve" would be treated as the same phrase *problem solv* (after stemming), so they could be counted together.

This analysis was repeated three times using:

1. Recommendations from in-person evaluations only.
2. Recommendations from remote evaluations only.
3. Recommendations from both in-person and remote evaluations combined, to understand which recommendations were made most frequently overall.

After counting all bigrams, we report the 40 most frequent. This threshold was selected because it includes at least the top ten skill-related words or phrases for each analysis, while being short enough to include in this work.

3.4.3 Supervisor Recommendations: Logistic Regression

We analyze the supervisor recommendations text field using a logistic regression classifier to identify differences between the recommendations given by 2021 supervisors and those given in 2019. The classifier was trained using the recommendation comments from 2019 and 2021 to predict one of the following two labels:

- In-person, if the recommendation comment was from 2019, or
- Remote, if the recommendation comment was from 2021.

The goal of this process was to identify the features that most strongly distinguish remote recommendations from those given to in-person students. For this, eli5 was used to display the top 100 terms that most strongly predicted a label of in-person and the bottom 100 terms that most strongly predicted a label of remote. These terms were compared to identify differences in the recommendations made for remote and in-person students. All preprocessing steps were identical to those used in the above logistic regression classifiers used to identify what makes students successful.

3.5 Results

3.5.1 Logistic Regression to Determine What Makes Students Successful

In the 2021 data, out of 23417 total student evaluations, 19204 performance evaluations were kept after removing short supervisor comments. Of these students 37.5% were given a performance evaluation rating of outstanding by their employer, while the remaining 62.4% were given a lower overall rating, from excellent to unsatisfactory. For the purposes of classification, these remaining students were considered to be *not outstanding*. The logistic regression classifier achieved an average prediction accuracy of 79.6% using 10-fold cross validation. A majority-class baseline model that simply predicts *not outstanding* for every student would only achieve an accuracy of 62.4%. The logistic regression model improves upon this baseline, indicating that it has learned important differences between outstanding and non-outstanding students. These results were similar for the 2019 data, where 19,888 evaluations were kept after removing short supervisor comments, with 35.1% corresponding to outstanding students. Although a baseline majority class classifier would produce an accuracy of 64.9%, the logistic regression classifier trained on the 2019 data achieved an accuracy of 80.3%, again showing an ability to learn important patterns from the data.

The model weights of the logistic regression classifier were examined to understand which terms are the strongest predictors of an outstanding student. After discarding terms that are not related to a student's skills, abilities, or knowledge (for example *went*, *expect*, and *student*), only the top 100 most and least predictive terms are kept for analysis.

Tables 3.2 and 3.3 (2021) and Tables 3.4 and 3.5 (2019) summarize the relevant terms that most strongly predicted an overall performance rating of outstanding along with their weight. Tables 3.2 and 3.4 list terms that are more likely to be present for (and therefore predict) outstanding students, while Tables 3.3 and 3.5 list terms that instead predict non-outstanding students. Each term's weight is either positive or negative. If a term has a positive weight, it indicates that a student with that term in their evaluation is more likely to receive a rating of outstanding. On the other hand, a term with a negative weight indicates that a student is likely to receive a non-outstanding rating. The magnitude reflects how strongly the term predicts the given label. More extreme values are better predictors. The original text column shows the words in the original text that produced these terms. This allows the reader to understand how terms in the table correspond to words written by the supervisor. For example, the term *outstand* was produced from the words 'outstanding', 'outstandingly', 'outstand', or 'outstandings' through preprocessing. The words in this column have been lowercased and punctuation removed to prevent multiple entries for words such as 'Outstanding' and 'outstanding', but they are otherwise identical to words appearing directly in the comments. The original words are presented in order from most to least frequent, and in the case where there are more than five original words, only the top five most frequent are presented for brevity. Finally, only meaningful skill-related terms were kept, so the rank column is included to show the position of the included terms in the ranking. If a value from 1 to 100 is missing from this column, it indicates that that term was not meaningful and was removed. Some examples of removed terms include *went*, *student*, *company*, *amaz* (amazing), *incred* (incredible), *intern*, *year*, and *level*. These terms were removed because they do not help to explain what skills, abilities, and attributes differentiate outstanding students. For transparency, the full tables can be viewed in Tables A.1, A.2, A.3, and A.4 in Appendix A to examine the terms that were removed.

Overall, many of the strongest predictors were common to students in both 2019 and 2021. Both remote and in-person outstanding students were characterized as good leaders, designers, and innovators, based on the presence of terms *lead*, *design*, and *innov* in both Table 3.2 and Table 3.4 However, notably, the term *independ* was a predictor of a remote outstanding student, while the term *work independ* predicted a non-outstanding student in in-person settings. This suggests that independence may be a skill that is more valued in remote settings. Further analysis of these results will be presented in the discussion section.

Table 3.2: A selection of the top 100 terms predicting a outstanding student in 2021 evaluations of mostly remote positions.

Rank	Term	Weight	Original Text

1	outstand	16.846	“outstanding”, “outstandingly”, “outstand”, “outstandings”
3	expect	4.510	“expectations”, “expected”, “expectation”, “expect”, “expecting”
9	exceed	3.349	“exceeded”, “exceed”, “exceeding”, “exceeds”, “exceedingly”
14	lead	3.127	“leadership”, “lead”, “leading”, “leads”, “leaded”
19	fulltim	2.874	“fulltime”, “fulltimer”, “fulltimers”, “fulltimes”
23	product	2.721	“product”, “production”, “productive”, “products”, “productivity”
25	deliv	2.677	“deliver”, “delivered”, “delivering”, “delivers”, “delived”
27	problem	2.670	“problem”, “problems”
29	hire	2.619	“hire”, “hired”, “hiring”, “hires”, “hirings”
32	engin	2.450	“engineering”, “engineer”, “engineers”, “engine”, “engines”
33	design	2.413	“design”, “designs”, “designing”, “designed”, “designer”
35	critic	2.381	“critical”, “criticism”, “critically”, “criticisms”, “criticize”
36	implement	2.321	“implementation”, “implement”, “implemented”, “implementing”, “implementations”
39	featur	2.305	“features”, “feature”, “featured”, “featuring”, “featureful”
41	impact	2.289	“impact”, “impactful”, “impacted”, “impacts”, “impacting”
43	servic	2.224	“service”, “services”, “servicing”, “serviceability”, “serviceable”
45	custom	2.175	“customer”, “customers”, “custom”, “customizing”, “customized”
50	constant	2.056	“constantly”, “constant”, “constants”
51	initi	2.046	“initiative”, “initial”, “initiatives”, “initially”, “initiate”
52	innov	2.034	“innovative”, “innovation”, “innovate”, “innovations”, “innovating”
53	solut	2.030	“solutions”, “solution”, “solutioning”, “solutation”, “solutioned”
58	dedic	1.993	“dedication”, “dedicated”, “dedicate”, “dedicating”, “dedicates”
63	quick	1.898	“quickly”, “quick”, “quickness”
70	new	1.794	“new”, “newness”
73	organ	1.771	“organization”, “organized”, “organizing”, “organize”, “organizations”

74	scope	1.770	"scope", "scoping", "scoped", "scopes"
77	code	1.740	"code", "coding", "codes", "coded", "codeing"
86	consist	1.664	"consistently", "consistent", "consistency", "consistantly", "consisted"
87	cultur	1.638	"culture", "cultural", "cultures", "culturally", "cultureal"
91	qualiti	1.564	"quality", "qualities"
92	independ	1.554	"independently", "independent", "independence", "independantly", "independant"
96	model	1.524	"model", "models", "modeling", "modelling", "modelled"
97	care	1.503	"care", "carefully", "careful", "cares", "caring"
98	complex	1.490	"complex", "complexity", "complexities", "complexed", "complexes"
99	fast	1.485	"fast", "fastly", "fasted"

Table 3.3: A selection of the top 100 terms predicting a non-outstanding student in 2021 evaluations of mostly remote positions.

Rank	Term	Weight	Original Text
1	overal	-4.612	"overall"
8	remot	-1.980	"remote", "remotely"
16	met expect	-1.695	"met expectations", "met all expectations", "met our expectations", "met the expectations", "met my expectations"
22	struggl	-1.381	"struggled", "struggle", "struggling", "struggles"
25	ask question	-1.333	"ask questions", "asking questions", "ask more questions", "asked questions", "asks questions"
34	friend	-1.12	"friendly", "friends", "friend"
38	confid	-1.072	"confidence", "confident", "confidently", "confidant", "confidance"
39	assign time	-1.069	"assignments on time", "assigned to him in a timely", "assignments in a timely", "assigned to her in a timely", "assignment on time"
41	time manner	-1.037	"timely manner", "time manner", "timely manners", "time in the manner", "timing manner"
42	work home	-1.037	"working from home", "work from home", "worked from home", "working at home", "works from home"

44	comfort	-1.024	"comfortable", "comfort", "comfortably", "comfortability", "comforted"
49	learn	-1.008	"learn", "learning", "learned", "learns", "learnings"
52	virtual	-0.992	"virtual", "virtually", "virtualization"
59	remot work	-0.944	"remote work", "remote working", "remotely working", "remotely for this work", "remote from work"
64	hard worker	-0.922	"hard worker", "hard of a worker"
71	improv	-0.879	"improve", "improvement", "improved", "improving", "improvements"
77	improv term	-0.819	"improved over the term", "improvement over the term", "improvement in terms", "improved as the term", "improvement this term"
82	ask help	-0.806	"ask for help", "asking for help", "asked for help", "asks for help", "asked to help"
94	engag	-0.755	"engaged", "engagement", "engage", "engaging", "engagements"
95	follow instruct	-0.748	"follow instructions", "followed instructions", "follows instructions", "following instructions", "follow the instructions"
97	perform task	-0.735	"perform tasks", "performing tasks", "perform the tasks", "performed all tasks", "performed tasks"
100	manag workload	-0.723	"manage her workload", "manage his workload", "managed her workload", "manage workload", "manage workloads"

Table 3.4: A selection of the top 100 terms predicting a outstanding student in 2019 evaluations of mostly in person positions.

Rank	Term	Weight	Original Text
1	outstand	17.581	"outstanding", "outstandingly", "outstand", "outstandings"
3	expect	4.489	"expectations", "expected", "expectation", "expect", "expecting"
4	exceed	4.123	"exceeded", "exceed", "exceeding", "exceeds", "exceedingly"
6	fulltim	3.958	"fulltime", "fulltimer", "fulltimers", "fulltimes"
14	hire	3.287	"hire", "hired", "hiring", "hires", "hirings"
15	lead	3.236	"leadership", "lead", "leading", "leads", "leaded"

25	engin	2.643	"engineering", "engineer", "engineers", "engine", "engines"
31	solut	2.450	"solutions", "solution", "solutioning", "solutation", "solutioned"
33	idea	2.368	"ideas", "idea"
35	graduat	2.316	"graduate", "graduation", "graduates", "graduated", "graduating"
38	complex	2.265	"complex", "complexity", "complexities", "complexed", "complexes"
40	design	2.200	"design", "designs", "designing", "designed", "designer"
43	critic	2.167	"critical", "criticism", "critically", "criticisms", "criticize"
49	respons	2.130	"responsibilities", "responsibility", "responsible", "responsive", "response"
50	impact	2.121	"impact", "impactful", "impacted", "impacts", "impacting"
54	deliv	2.049	"deliver", "delivered", "delivering", "delivers", "delived"
55	innov	2.039	"innovative", "innovation", "innovate", "innovations", "innovating"
56	consist	2.026	"consistently", "consistent", "consistency", "consistantly", "consisted"
62	insight	1.975	"insights", "insight", "insightful", "insightfulness"
63	develop	1.929	"development", "develop", "developing", "developed", "developer"
65	implement	1.926	"implementation", "implement", "implemented", "implementing", "implementations"
66	product	1.925	"product", "production", "productive", "products", "productivity"
78	contribut	1.838	"contributions", "contributed", "contribute", "contribution", "contributing"
80	featur	1.835	"features", "feature", "featured", "featuring", "featureful"
84	research	1.805	"research", "researching", "researched", "researcher", "researchers"
89	dedic	1.767	"dedication", "dedicated", "dedicate", "dedicating", "dedicates"
96	challeng	1.719	"challenges", "challenging", "challenge", "challenged", "challenger"
97	code	1.711	"code", "coding", "codes", "coded", "codeing"

Table 3.5: A selection of the top 100 terms predicting a non-outstanding student in 2019 evaluations of mostly in person positions.

Rank	Term	Weight	Original Text
1	good	-4.311	"good", "goods", "goodly", "goodness"
2	overal	-2.428	"overall"
10	task	-1.860	"tasks", "task", "tasked", "tasking", "taskings"
14	ask question	-1.642	"ask questions", "asking questions", "ask more questions", "asked questions", "asks questions"
24	quiet	-1.377	"quiet", "quietly", "quietness"
33	sometim	-1.210	"sometimes", "sometime"
37	struggl	-1.163	"struggled", "struggle", "struggling", "struggles"
39	friend	-1.138	"friendly", "friends", "friend"
40	slow	-1.136	"slow", "slowing", "slowed", "slows", "slowness"
53	instruct	-0.959	"instructions", "instruction", "instructional", "instructed", "instructing"
55	improv	-0.950	"improve", "improvement", "improved", "improving", "improvements"
65	understand task	-0.906	"understand the task", "understand the tasks", "understanding of the task", "understanding of the tasks", "understand tasks"
66	engag	-0.901	"engaged", "engagement", "engage", "engaging", "engagements"
76	confid	-0.843	"confidence", "confident", "confidently", "confidant", "confidance"
79	clarif	-0.834	"clarification", "clarifications"
80	time manner	-0.832	"timely manner", "time manner", "timely manners", "time in the manner", "timing manner"
83	demonstr willing	-0.806	"demonstrated a willingness", "demonstrated willingness", "demonstrated his willingness", "demonstrates a willingness", "demonstrated the willingness"
87	work independ	-0.798	"work independently", "worked independently", "working independently", "works independently", "worked very independently"
88	divers	-0.798	"diverse", "diversity", "diversed", "diversely", "diversities"
93	assist	-0.774	"assist", "assistance", "assisted", "assistant", "assisting"
98	focus	-0.761	"focus", "focused", "focusing", "focuses", "focuse"

3.5.2 Supervisor Recommendations: Frequency Analysis

There were 15728 non-blank supervisor recommendation comments in 2019 and 14166 comments in 2021. Comments that were left blank (i.e., not filled in by the supervisor) were discarded.

The analysis of frequent skills can be found in Table 3.6. This table shows the fifty most frequent bigrams for each of the in-person, remote, and all recommendation comments, along with the count, indicating the number of times the bigrams appeared across all recommendation comments in the given dataset. From the bigram column of these tables, the following top ten skills were extracted, in order. Explanations of the original skills that produced these terms, which were deduced through manual inspection of supervisor recommendation comments containing these terms, are provided in brackets:

- **In-Person:** *ask question* (asking questions), *communic skill* (communication skills), *problem solv* (problem solving), *attent detail* (attention to detail), *technic skill* (technical skills), *take initi* (take initiative), *work ethic* (work ethic), *oral communic* (oral communication), *public speak* (public speaking), *softwar develop* (software development), and *written communic* (written communication).
- **Remote:** *ask question* (asking questions), *communic skill* (communication skills), *technic skill* (technical skills), *critic think* (critical thinking), *seek opportun* (seek opportunities), *take initi* (take initiative), *profession develop* (professional development), *softwar develop* (software development), *work ethic* (work ethic), and *attent detail* (attention to detail).
- **Both (in-person and remote combined):** *ask question* (asking questions), *communic skill* (communication skills), *technic skill* (technical skills), *problem solv* (problem solving), *take initi* (take initiative), *attent detail* (attention to detail), *softwar develop* (software development), *work ethic* (work ethic), *profession develop* (professional development), *critic think* (critical thinking).

Table 3.6: The most frequent bigrams (pairs of words) in recommendation comments written by the supervisor for in-person, remote, and all placements combined.

In-Person (2019)		Remote (2021)		All (2019 and 2021)	
Bigram	Count	Bigram	Count	Bigram	Count

would recommend	971	ask question	836	would recommend	1693
continu develop	803	continu develop	730	ask question	1639
ask question	803	work term	724	continu develop	1533
communic skill	789	would recommend	722	work term	1505
work term	781	continu work	499	communic skill	1279
continu work	763	communic skill	490	continu work	1262
would encourag	589	would encourag	470	would encourag	1059
would like	482	technic skill	459	would like	876
encourag continu	451	coop term	409	coop term	835
problem solv	441	continu learn	407	technic skill	817
coop term	426	critic think	397	encourag continu	797
good work	392	would like	394	continu learn	764
attent detail	384	encourag continu	346	good work	697
technic skill	358	learn new	307	problem solv	636
continu learn	357	good work	305	keep good	608
team member	347	area develop	279	take initi	595
make sure	344	keep learn	275	team member	583
take initi	341	develop skill	275	attent detail	583
keep good	334	keep good	274	learn new	577
continu improv	317	seek opportun	266	develop skill	565
great work	292	great work	258	make sure	561
develop skill	290	take initi	254	great work	550
work ethic	288	profession develop	252	recommend continu	522
take time	282	recommend continu	251	softwar develop	507
would benefit	279	softwar develop	245	continu improv	506
oral communic	279	continu build	244	would benefit	499
recommend continu	271	team member	236	work ethic	498
public speak	271	gain experi	233	profession develop	495
would also	270	career path	229	seek opportun	492
like see	270	futur career	224	critic think	486
learn new	270	would benefit	220	continu build	483
area improv	265	make sure	217	gain experi	482
softwar develop	262	work ethic	210	keep learn	475
written communic	260	continu seek	209	would also	473
futur work	255	futur work	208	futur work	463

present skill	251	would also	203	take time	451
gain experi	249	attent detail	199	area improv	450
profession develop	243	problem solv	195	like see	431
continu build	239	continu improv	189	also encourag	406
seek opportun	226	continu grow	185	continu seek	405

3.5.3 Supervisor Recommendations: Logistic Regression

The results of the logistic regression analysis can be found in Table 3.7 (terms that predict remote students) and Table A.6 (terms that predict in-person students). For brevity, only a selection of the top terms are included. However, removed terms can be examined in Appendix A Tables A.5 and A.6, where the tables are listed in full. As in Tables 3.2, 3.3, 3.4, and 3.5, the weight is either negative for in-person and positive for remote, while the magnitude of the weight shows how strongly it predicts whether or not a student is remote. The rank and original text columns are also as described in the logistic regression results to predict whether or not a student is outstanding.

When making recommendations, in-person students were encouraged to improve their communication (especially verbal communication and presentation skills, based on the presence of terms such as *present*, *public speak*, *oral*, and *assert*), while remote students were encouraged to be more innovative (*innov*), curious (*curios*), and ask questions (*question*).

Table 3.7: Selected terms from the top 100 terms predicting a remote student (2021) from the supervisors' recommendation comments.

Rank	Term	Weight	Original Text
2	collabor	4.979	"collaboration", "collaborative", "collaborate", "collaborated", "collaborating"
3	innov	4.663	"innovative", "innovation", "innovate", "innovations", "innovating"
6	critic think	3.829	"critical thinking", "critically think", "critically thinking", "critical think", "critical thinkings"
8	mindset	3.391	"mindset", "mindsets"
9	critic	3.293	"critical", "criticism", "critically", "criticisms", "criticize"
12	selfmanag	2.664	"selfmanagement", "selfmanage", "selfmanaged", "selfmanaging", "selfmanageable"

14	technolog	2.432	“technologies”, “technology”, “technological”, “technologically”, “technolog”
16	learn	2.353	“learn”, “learning”, “learned”, “learns”, “learnings”
17	curios	2.311	“curiosity”, “curiosities”
21	selfassess	2.114	“selfassessment”, “selfassess”, “selfassessments”, “self-assessed”, “selfassessing”
22	technic skill	2.112	“technical skills”, “technical skill”, “technically skilled”, “technical and other skills”, “technical and skill”
27	disciplin	1.888	“discipline”, “disciplines”, “disciplined”
31	innov mindset	1.849	“innovation mindset”, “innovative mindset”, “innovative mindsets”, “innovate mindset”, “innovating mindset”
36	communic	1.647	“communication”, “communicate”, “communicating”, “communications”, “communicated”
41	agil	1.557	“agile”, “agility”
42	opportun	1.525	“opportunities”, “opportunity”, “opportune”
47	good communic	1.376	“good communication”, “good communicator”, “good at communicating”, “good at communication”, “good communications”
48	think critic	1.371	“think critically”, “thinking critically”, “thinks critically”, “think more critically”, “think about critical”
51	self manag	1.332	“self management”, “self manage”, “self managed”, “self managing”, “self manages”
52	opportun learn	1.322	“opportunities to learn”, “opportunity to learn”, “opportunities for learning”, “opportunities and learning”, “opportunity for learning”
53	make suggest	1.313	“make suggestions”, “making suggestions”, “make more suggestions”, “make a suggestion”, “make some suggestions”
58	question	1.230	“questions”, “question”, “questioning”, “questioned”, “questionable”
68	collabor communic	1.166	“collaboration and communication”, “collaborate and communicate”, “collaboration communication”, “collaborated and communicated”, “collaborating and communicating”
76	literaci	1.137	“literacy”
78	academ	1.130	“academic”, “academics”, “academically”, “academe”, “academical”
81	collabor skill	1.125	“collaboration skills”, “collaborative skills”, “collaboration skill”, “collaborative skill”, “collaborating skills”

85	confid speak	1.116	"confidence to speak", "confidence when speaking", "confidence in speaking", "confident to speak", "confident when speaking"
86	honest	1.113	"honest", "honestly"
92	technolog skill	1.094	"technological skills", "technology skills", "technologies and skills", "technology skill", "technological skill"
93	encourag think	1.088	"encouraged to think", "encourage him to think", "encourage you to think", "encourage her to think", "encourage him to do so i think"

Table 3.8: Selected terms from the top 100 terms predicting an in-person student (2019) from the supervisors' recommendation comments.

Rank	Term	Weight	Original Text
1	written	-2.795	"written"
2	interperson	-2.671	"interpersonal", "interpersonally", "interperson"
4	public speak	-2.246	"public speaking"
5	present	-2.118	"presentation", "presented", "present", "presentations", "presenting"
6	entrepreneuri	-2.064	"entrepreneurial", "entrepreneurialism", "entrepreneurially"
7	interperson communic	-1.931	"interpersonal communication", "interpersonal and communication", "interpersonal communications", "interpersonal and communications"
8	depend	-1.926	"dependable", "dependability", "depend", "depending", "dependencies"
9	proactiv	-1.868	"proactive", "proactively", "proactiveness", "proactivity", "proactivelys"
10	oral	-1.830	"oral", "orally", "orall"
11	assert	-1.827	"assertive", "assertiveness", "assert", "asserting", "assertively"
14	quiet	-1.749	"quiet", "quietly", "quietness"
15	problem solv	-1.725	"problem solving", "problem solve", "problems to solve", "problem solved", "problem to solve"
16	teamwork	-1.723	"teamwork", "teamworker", "teamworking", "teamworks"
17	public	-1.715	"public", "publication", "publications", "publicly", "publically"

18	oral communic	-1.701	"oral communication", "oral communications", "oral communicator", "oral and communication", "oral the communication"
20	speak	-1.661	"speaking", "speak", "speaks", "speakings"
21	punctual	-1.650	"punctual", "punctuality", "punctually", "punctuallity"
23	interact	-1.562	"interactions", "interact", "interaction", "interacted", "interacting"
24	audienc	-1.555	"audience", "audiences"
25	enthusiasm	-1.532	"enthusiasm", "enthusiasms"
27	independ	-1.504	"independently", "independent", "independence", "independantly", "independant"
29	attent	-1.476	"attention", "attentive", "attentively", "attentiveness", "attentions"
30	peopl	-1.381	"people", "peoples"
33	solv skill	-1.337	"solving skills", "solving skill", "solveing skills", "solving these are skills"
35	document	-1.314	"documentation", "documents", "document", "documented", "documenting"
40	matur	-1.277	"maturity", "mature", "matures", "matured", "maturely"
41	languag	-1.269	"language", "languages"
42	problemsolv	-1.268	"problemsolving", "problemsolve", "problemsolver", "problemsolved"
47	lead	-1.221	"leadership", "lead", "leading", "leads", "leaded"
50	creativ	-1.189	"creative", "creativity", "creatively", "creativeness", "creativities"
52	improv written	-1.186	"improve his written", "improving written", "improve written", "improving his written", "improve her written"
66	listen	-1.130	"listening", "listen", "listens", "listener", "listened"
68	intuit	-1.129	"intuition", "intuitive", "intuit", "intuitively", "intuitions"
69	activ listen	-1.126	"active listening", "active listener", "actively listening", "actively listens", "actively listened"
70	consist	-1.122	"consistently", "consistent", "consistency", "consistantly", "consisted"
71	verbal	-1.116	"verbal", "verbally", "verbalize", "verbalized", "verbalizing"
75	communic skill	-1.093	"communication skills", "communications skills", "communication skill", "communicative skills", "communications skill"
79	social	-1.072	"social", "socially", "socialize", "socializing", "socials"

86	opinion	-1.055	"opinions", "opinion", "opinionated"
97	present skill	-1.017	"presentation skills", "presentation skill", "presentations skills", "presenting skills", "presentational skills"

3.6 Discussion

3.6.1 A Comparison of Remote and In-Person Evaluations

The analysis of supervisors' comments highlights skills that distinguish outstanding students in both remote and in-person settings. Outstanding students in both contexts are described as designers, innovators, and leaders. This result differs from prior work by Coll et al. [16], who found that team leadership was one of the least important competencies, as perceived by employers. This indicates that the most successful students do not only produce good work; they also have the soft skills required to innovate, design novel solutions, and demonstrate leadership.

For remote students, independence (through the presence of the term *independ*) was found to be a strong predictor of an outstanding student. However, the term *independ* was found to strongly predict an outstanding student in remote settings while *work independ* predicted a non-outstanding student in in-person settings. To better understand this result, a more in-depth manual analysis was performed to analyze comments that contained these keywords. Through this analysis of comments mentioning independence, it was found that:

- Independence was usually mentioned in a positive way to praise a student's ability to produce work or solve problems independently. However, too much independence could be a negative trait. For example, some employers stated that their student was too independent and did not collaborate well with others.

- Independence was usually only one of many skills mentioned by the employer. Although independence was important, it was not the sole distinguishing factor for an outstanding student.

- Students at all performance evaluation levels were praised for their independence, but students with higher performance ratings (excellent and above) were more likely to receive a positive comment than those with lower ratings (good and below).

- In remote settings, many employers discussed the student's ability to succeed despite working remotely, showing that independence was an asset for this remote work. Also, independence was associated with initiative: being proactive, seeking new tasks, solving their own problems, and developing useful novel features independently.

- In in-person settings, employers seemed more likely to mention a student's flexibility - for example, that the student was both independent and a good collaborator.

This result differs from the work of Nevison et al. [42], who found that independence only had medium importance. Overall, independence and related skills such as self-management appear to be more important in remote work where students cannot be directly supervised most of the time.

Interestingly, the term *cultur* points to remote outstanding students commonly having a cultural awareness or fitting well into workplace culture. Students were likely expected to acquire this through remote interactions, such as virtual lunches, perhaps requiring extra effort that was praised by employers. Culture was highly rated in the study by Nevison et al. [42], so this finding aligns with prior work.

For in-person students, the term *idea* was a strong predictor of an outstanding student. However, when examining the context in which this word is used, it was found that it is used in a similar way to *innov* (e.g., a student bringing fresh and creative ideas). Although it may support innovation being more important in in-person settings, it is possible this is the result of wording changes in the evaluation form: the word 'ideas' appears seven times in the detailed skill descriptions in the 2019 form, but only once in the 2021 form. A second term of importance is *research*. It was found to be related not only to a student's ability to work as a research assistant, but also a student's ability to seek solutions to problems and learn independently in general.

3.6.2 Supervisor Recommendations for Remote and In-Person Positions

Supervisors of both in-person and remote students commonly recommended that students be more inquisitive, with asking questions being the most frequent recommendation. Furthermore, good communication was recommended more frequently than technical and software development skills regardless of context. Other important skills include taking initiative, attention to detail, and work ethic, although notably work ethic is more frequently recommended for remote students rather than in-person students. This may be the result of students feeling less motivated to do work from home, so a good work ethic is more valuable to employers.

When making recommendations, supervisors of in-person students were more likely to recommend presentation or communication skills, as can be observed from the top terms *written*, *interperson* (e.g., interpersonal communication), *public speak* (e.g., public speaking), and *present*. This may be a result of in-person students having more need to interact directly with peers and supervisors. However, it should be noted that students in remote settings were encouraged to improve their communication as well, showing that good communication is important for all co-op students. This aligns with prior work [4, 51] which found that strong communication was extremely important for remote employees.

Remote students were encouraged to be more innovative, perhaps showing that producing creative solutions is more important in a remote environment. Self-management and technological skills were also important in a remote context, which likely reflects the remote student's need to use technology to communicate with other team members as well as to complete work.

It is also important to consider the change in evaluation form when analyzing these logistic regression results. Terms such as *interperson* and *teamwork* were predictors of in-person recommendations, but these can be explained by changes in wording in the evaluation forms. The 2019 version of the form included skills called "interpersonal communication" and "teamwork". In the 2021 form, the skills were changed to remove the word interpersonal when referencing communication skills and to use "collaboration" instead of "teamwork".

Finally, it should be noted that a switch from in-person to remote work may not be the only explanation for these findings. Work has changed as a result of the COVID-19 pandemic beyond an increase in remote work. For example, employers' priorities have shifted, and they may now hire students for different kinds of work. These changes may also be reflected in these data and could also explain the discrepancies between the 2019 and 2021 evaluations.

3.7 Practical Implications

By analyzing the evaluation criteria and supervisors' comments, we discovered the most important skills for success in remote and in-person co-op positions. Furthermore, it was possible to learn which skills supervisors wanted students to improve. This information may be beneficial to students who wish to improve their job prospects, employers who aim to recruit the best talent, and institutions who may shift their educational policies accordingly. By understanding which skills are most valued in remote and in-person work, students can focus on acquiring skills that match the type of position they aim to fill. Employers may change their recruitment strategies to attract students with the most valuable skill sets. Institutions may offer additional workshops to improve highly valued soft skills such as communication. Strategies such as these can help co-op students to be more successful in the work-term positions.

3.7.1 Recommendations

From our findings, we make the following recommendations for students in co-operative education programs.

Students who plan to work remotely should focus on improving their work ethic and ability to work independently. Having a strategy for staying motivated and focusing on mastering technological skills including communication technologies will help students succeed in remote work.

Students who aim for in-person positions should ensure that they have excellent communication skills. Practicing presenting, speaking and writing clearly, and asking questions will be important for these students to ensure they interact effectively with their colleagues and supervisors.

Finally, students who fall short of receiving the highest evaluation ratings should focus on asking questions to ensure they fully understand their assigned tasks. Learning how to effectively manage workloads is also important to ensure work timeframes are met. If in a remote context, these students should aim to be more motivated and independent. In an in-person setting, they should instead focus on interacting positively and collaborating with colleagues and supervisors to achieve a rating of outstanding.

Chapter 4

Conclusion

4.1 Job Description Analysis

In this investigation, we explored the types of jobs available to current co-op students and the expectations of employers for students with specific academic backgrounds.

Using k-means clustering and frequency-based analyses, we identified groups of required skills representing the types of available jobs. We found that communication was the most frequently required skill regardless of a student's specific academic program. Technical skills were more strongly associated with senior than junior students, who were more associated with people-oriented skills such as customer service. In the eleven k-means clusters, we identified groups of skills associated with research and design, communication, technical skills (programming, software and web development, data science, and scripting), finance, graphic design, access to a vehicle, inclusivity and integrity, and remote work. These groupings indicate the types of jobs available to co-op students.

Next, by constructing two bipartite graphs, we investigated the relationships between academic advertising clusters, used by employers to target students with specific backgrounds, and job description attributes: required skills or job roles (represented by job titles). After co-clustering the advertising clusters and job attributes, we summarized employer expectations for students with specific academic backgrounds. We found significant differences between employer expectations for students with different academic backgrounds, with specific technical skills required for many programs. While technical and data science skills and job roles were prevalent for many program groupings, they were notably absent from finance and arts groupings. Overall, this demonstrates that employers have clear expectations about the types of skills and roles that are most suitable for each type of student.

4.2 Performance Evaluation Analysis

Our comparison of the employer evaluations of remote and in-person undergraduate co-op students uncovered key factors that contribute to student success. By analyzing written comments and recommendations from supervisors using logistic regression, we determined which skills correlate most strongly with the overall performance rating. Through analysis of supervisor comments, we found that the most successful students were characterized by employers as being good leaders and innovators, with remote students also being praised for their independence. Finally, supervisors recommended that remote students become more innovative and learn technological skills while in-person students should improve their oral communication and presentation abilities. Ultimately, these important skills are valued by employers and should therefore be acquired by students in their early careers.

4.3 Comparisons

Although our two analyses used different datasets (job descriptions in Chapter 2 and performance evaluations in Chapter 3), these findings together paint a complete picture of the skills and roles employers expect from students in remote contexts. The job description analysis focused on discovering specific (often technical) skills required for each type of student, while the performance evaluation analysis determined the skills that make co-op students outstanding when compared to their peers. In particular, we found communication is the most commonly required skill in job descriptions. This aligns with our finding from the supervisor recommendation comments, that improved communication was recommended more frequently than technical and software development skills in both remote and in-person settings. Although very specific technical skills were expected from students in different academic programs, our findings show that students who wish to be seen as outstanding must also focus on improving their soft skills, such as communication and work ethic.

4.4 Insights for Data Scientists

Several insights emerged from this work that might inform future data scientists carrying out a study in an entirely different application area. First, working with small, domain-specific datasets is challenging for several reasons. Often, sophisticated deep learning models fail on this type of data because they cannot capture domain-specfic information and may overfit on the limited data. Smaller datasets may lack labels and structure that would make analyses easier, because creating such labels requires significant time and effort. However, as we demonstrated in the job description analysis in Chapter 2, it is still possible to generate interesting findings from such data. Simple models not involving deep learning, such as logistic regression, k-means, frequency analyses, and community detection can be surprisingly effective on this type of data. Furthermore, the data can often contain unique features, such as the co-op job posting clusters in Chapter 2's job description dataset, which can spark ideas for novel analyses. As a result, we recommend that future data scientists explore their data carefully, searching for unique features and creating custom preprocessing approaches that facilitate the extraction of key domain-specific information.

4.5 Future Work

There are several avenues available for future work.

4.5.1 Job Description Analysis

Based on our job description analysis, future research could use the bipartite graph co-clustering methodology to investigate relationships between other variables of interest. Our analysis focused on employer expectations and the factors that make co-op students successful. Future work might explore the factors that make employers and jobs attractive to students, so that employers could better advertise themselves to attract top talent. For example, a bipartite graph between prospective employee groups and jobs could be constructed to determine which jobs are most attractive to students, based on information about the jobs they view and/or apply to. Such an analysis could reveal what factors make jobs attractive to students in different disciplines.

4.5.2 Performance Evaluation Analysis

The findings from the performance evaluations also have important implications for future research. We identified important skills for successful remote and in-person co-op students but did not determine the reasons why employers value these skills most highly. These findings could drive future studies involving direct interviews with employers to understand why highly valued skills may have changed with the switch to remote work. One method that would simplify the analysis might provide employers with a list of skills, ask them to rate (on a scale from one to five) how important they believe that skill is for student success, and to provide a text explanation for their rating. This could be linked to a particular student's performance to make the process less abstract, since it might be difficult for employers to accurately judge each skill in isolation. By having separate text comments for each skill along with a numeric rating, it would be easy to cluster the terms to find the reasons why a particular skill might be more valued.

It would also be interesting to explore why oral communication and presentation skills were more valued for in-person positions. As mentioned in the discussion, there may also have been additional factors (other than the switch to remote work) that contributed to the differences found through this investigation. Future research, perhaps using interviews or surveys, might assess how employers' priorities have shifted and whether this aligns with the skills found in this analysis.

Finally, although these findings were specific to co-operative education in an undergraduate program, they may suggest similar trends for remote work in general. For example, future work might distinguish undergraduate and graduate co-op students or remote employees in general, who are employed for longer periods and might be more affected by remote work. These findings could therefore be applicable to a wide range of researchers who are interested in employer satisfaction in modern work environments.

References

[1] Erik R Alanson, Erin M Alanson, Brittany Arthur, Aaron Burdette, Christopher Cooper, and Michael Sharp. Re-envisioning work-integrated learning during a pandemic: Cincinnati's experiential explorations program. *International Journal of Work-Integrated Learning*, 21(5):505–519, 2020.

[2] Earl Anderson, Nancy Johnston, Larry Iles, Norah Mcrae, Nancy Reed, and Julie Walchli. Co-operative education and student recruitment, engagement and success: Early findings from a multi-institutional study in british columbia. *Journal of Cooperative Education and Internships*, 46(1):58–76, 2012.

[3] Stijn Baert, Louis Lippens, Eline Moens, Johannes Weytjens, and Philippe Sterkens. The covid-19 crisis and telework: A research survey on experiences, expectations and hopes. 2020.

[4] Ellen Baker, Gayle C Avery, and John Crawford. Home alone: The role of technology in telecommuting. *Information Resources Management Journal (IRMJ)*, 19(4):1–22, 2006.

[5] Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy. Gephi: an open source software for exploring and manipulating networks. In *Proceedings of the international AAAI conference on web and social media*, volume 3, pages 361–362, 2009.

[6] Grigore Belostecinic, Radu Ioan Mogoș, Maria Loredana Popescu, Sorin Burlacu, Carmen Valentina Rădulescu, Dumitru Alexandru Bodislav, Florina Bran, and Mihaela Diana Oancea-Negescu. Teleworking—an economic and social impact during covid-19 pandemic: A data mining analysis. *International Journal of Environmental Research and Public Health*, 19(1):298, 2021.

[7] Gilles Bisson and Fawad Hussain. Chi-sim: A new similarity measure for the co-clustering task. In *2008 Seventh International Conference on Machine Learning and Applications*, pages 211–217. IEEE, 2008.

[8] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008.

[9] Tracey Bowen. Work-integrated learning placements and remote working: Experiential learning online. *International Journal of Work-Integrated Learning*, 21(4):377–386, 2020.

[10] Nora Brüning and Patricia Mangeol. What skills do employers seek in graduates?: Using online job posting data to support policy and practice in higher education. 2020.

[11] Xiaoyan Cai, Wenjie Li, and Renxian Zhang. Combining co-clustering with noise detection for theme-based summarization. *ACM Transactions on Speech and Language Processing (TSLP)*, 10(4):1–27, 2014.

[12] Jay Choi, Brian Foster-Pegg, Joel Hensel, and Oliver Schaer. Using graph algorithms for skills gap analysis. In *2021 Systems and Information Engineering Design Symposium (SIEDS)*, pages 1–6. IEEE, 2021.

[13] Shivangi Chopra and Lukasz Golab. Job description mining to understand work-integrated learning. *International Educational Data Mining Society*, 2018.

[14] Shivangi Chopra, Abeer Khan, Melicaalsadat Mirsafian, and Lukasz Golab. Gender differences in work-integrated learning assessments. In *EDM*, 2019.

[15] Shivangi Chopra, Abeer Khan, Melicaalsadat Mirsafian, and Lukasz Golab. Gender differences in work-integrated learning experiences of stem students: From applications to evaluations. *International Journal of Work-Integrated Learning*, 21(3):253–274, 2020.

[16] Richard K Coll, Karsten Zegwaard, and Dave Hodges. Science and technology stakeholders' ranking of graduate competencies part 1: Employers perspective. *International Journal of Work-Integrated Learning*, 3(2):19, 2002.

[17] Diego Collarana, Mikhail Galkin, Christoph Lange, Simon Scerri, Sören Auer, and Maria-Esther Vidal. Synthesizing knowledge graphs from web sources with the

minteˆ++ framework. In *The Semantic Web–ISWC 2018: 17th International Semantic Web Conference, Monterey, CA, USA, October 8–12, 2018, Proceedings, Part II 17*, pages 359–375. Springer, 2018.

[18] Maurits de Groot, Jelle Schutte, and David Graus. Job posting-enriched knowledge graph for skills-based matching. *arXiv preprint arXiv:2109.02554*, 2021.

[19] Bonnie Amelia Dean and Matthew Campbell. Reshaping work-integrated learning in a post-covid-19 world of work. *International Journal of Work-Integrated Learning*, 21(4):355–364, 2020.

[20] Inderjit S Dhillon. Co-clustering documents and words using bipartite spectral graph partitioning. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 269–274, 2001.

[21] Marta Fana, Santo Milasi, Joanna Napierala, Enrique Fernández-Macías, and Ignacio González Vázquez. Telework, work organisation and job quality during the covid-19 crisis: a qualitative study. Technical report, JRC Working Papers Series on Labour, Education and Technology, 2020.

[22] Farnoush Farhadi, Elham Hoseini, Sattar Hashemi, and Ali Hamzeh. Teamfinder: A co-clustering based framework for finding an effective team of experts in social networks. In *2012 IEEE 12th International Conference on Data Mining Workshops*, pages 107–114. IEEE, 2012.

[23] Alan Felstead and Golo Henseke. Assessing the growth of remote working and its consequences for effort, well-being and work-life balance. *New Technology, Work and Employment*, 32(3):195–212, 2017.

[24] Santo Fortunato. Community detection in graphs. *Physics reports*, 486(3-5):75–174, 2010.

[25] Anna Giabelli, Lorenzo Malandri, Fabio Mercorio, Mario Mezzanzanica, and Andrea Seveso. Skills2job: A recommender system that encodes job offer embeddings on graph databases. *Applied Soft Computing*, 101:107049, 2021.

[26] Ainsley S Goldman and Ashley E Sterling. Becoming a part while apart: Building professional identity and membership when working and learning remotely. *International Journal of Work-Integrated Learning*, 21(4):387–399, 2020.

[27] Mahmoud Haddara and Heather Skanes. A reflection on cooperative education: From experience to experiential learning. *International Journal of Work-Integrated Learning*, 8(1):67, 2007.

[28] Lynette Harris. Home-based teleworking and the employment relationship: Managerial challenges and dilemmas. *Personnel review*, 2003.

[29] Information technology — Artificial intelligence — Artificial intelligence concepts and terminology. Standard, International Organization for Standardization, 2021.

[30] Mathieu Jacomy, Tommaso Venturini, Sebastien Heymann, and Mathieu Bastian. Forceatlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software. *PloS one*, 9(6):e98679, 2014.

[31] Debora Jeske and Carol Linehan. Mentoring and skill development in e-internships. *Journal of Work-Applied Management*, 12(2):245–258, 2020.

[32] Shanshan Jia, Xiaoan Liu, Ping Zhao, Chang Liu, Lianying Sun, and Tao Peng. Representation of job-skill in artificial intelligence with knowledge graph analysis. In *2018 IEEE symposium on product compliance engineering-asia (ISPCE-CN)*, pages 1–6. IEEE, 2018.

[33] Yuheng Jiang and Lukasz Golab. On competition for undergraduate co-op placements: A graph mining approach. *International Educational Data Mining Society*, 2016.

[34] Yuheng Helen Jiang, Sally Wai Yin Lee, and Lukasz Golab. Analyzing student and employer satisfaction with cooperative education through multiple data sources. *Asia-Pacific Journal of Cooperative Education*, 16(4):225–240, 2015.

[35] Judie Kay, Norah McRae, and Leoni Russell. Two institutional responses to work-integrated learning in a time of covid-19: Canada and australia. *International Journal of Work-Integrated Learning*, 21(5):491–503, 2020.

[36] Mikhail Korobov and Konstantin Lopuhin. Overview — ELI5 0.11.0 documentation. https://eli5.readthedocs.io/en/latest/overview.html, 2017. Accessed: 2023-3-30.

[37] Sayam Kumar. Silhouette method — better than elbow method to find optimal clusters. https://towardsdatascience.com/silhouette-method-better-than-elbow-method-to-find-optimal-clusters-378d62ff6891. Accessed: 2023-03-30.

[38] Thomas K Landauer and Susan T Dumais. A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological review*, 104(2):211, 1997.

[39] Matthieu Latapy, Clémence Magnien, and Nathalie Del Vecchio. Basic notions for the analysis of large two-mode networks. *Social networks*, 30(1):31–48, 2008.

[40] Wenjing Lyu and Jin Liu. Soft skills, hard skills: What matters most? evidence from job postings. *Applied Energy*, 300:117307, 2021.

[41] Peter McCullagh. *Generalized linear models*. Routledge, 2019.

[42] Colleen Nevison, Lauren Cormier, Judene Pretti, and David Drewery. The influence of values on supervisors' satisfaction with co-op student employees. *International Journal of Work-Integrated Learning*, 19(1):1–11, 2018.

[43] NLTK. Nltk tokenizer package. https://www.nltk.org/api/nltk.tokenize.html. Accessed: 2023-03-30.

[44] NLTK. Snowball stemmer. https://www.nltk.org/api/nltk.stem.snowball.html#module-nltk.stem.snowball. Accessed: 2023-03-30.

[45] University of Waterloo. Waterlooworks has replaced jobmine. https://uwaterloo.ca/hire/waterlooworks-has-replaced-jobmine.

[46] University of Waterloo. Find a job on waterlooworks: Rank/match, Sep 2022.

[47] Mohammad S Parsa and Lukasz Golab. Social media mining to understand the impact of cooperative education on mental health. *International Journal of Work-Integrated Learning*, 22(4):521–537, 2021.

[48] Monica Aureliana Petcu, Maria Iulia Sobolevschi-David, Adrian Anica-Popa, Stefania Cristina Curea, Catalina Motofei, and Ana-Maria Popescu. Multidimensional assessment of job satisfaction in telework conditions. case study: Romania in the covid-19 pandemic. *Sustainability*, 13(16):8965, 2021.

[49] T J Pretti, B Etmanski, and D W Drewery. Development and validation of a future ready talent framework. *International Journal of Work-Integrated Learning*, 22(3):369–383, 2021.

[50] T Judene Pretti, Brittany Etmanski, and Amie Durston. Remote work-integrated learning experiences: Student perceptions. *International Journal of Work-Integrated Learning*, 21(4):401–414, 2020.

[51] Sumita Raghuram, Raghu Garud, Batia Wiesenfeld, and Vipin Gupta. Factors contributing to virtual work adjustment. *Journal of Management*, 27(3):383–405, 2001.

[52] Agota Giedrė Raišienė, Violeta Rapuano, Kristina Varkulevičiūtė, and Katarína Stachová. Working from home—who is happy? a survey of lithuania's employees during the covid-19 quarantine period. *Sustainability*, 12(13):5332, 2020.

[53] Kevin Raison, Noriko Tomuro, Steve Lytinen, and Jose P Zagal. Extraction of user opinions by adjective-context co-clustering for game review texts. In *Advances in Natural Language Processing: 8th International Conference on NLP, JapTAL 2012, Kanazawa, Japan, October 22-24, 2012. Proceedings*, pages 289–299. Springer, 2012.

[54] Manjeet Rege, Ming Dong, and Farshad Fotouhi. Bipartite isoperimetric graph partitioning for data co-clustering. *Data Mining and Knowledge Discovery*, 16:276–312, 2008.

[55] Giulio Rossetti, Letizia Milli, and Rémy Cazabet. Cdlib: a python library to extract, compare and evaluate communities from complex networks. *Applied Network Science*, 4(1):1–26, 2019.

[56] Lea Samek, Mariagrazia Squicciarini, and Emile Cammeraat. The human capital behind ai: Jobs and skills demand from online job postings. 2021.

[57] Scikit-Learn. Kmeans. https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html. Accessed: 2023-03-30.

[58] Scikit-Learn. Selecting the number of clusters with silhouette analysis on kmeans clustering. https://scikit-learn.org/stable/auto_examples/cluster/plot_kmeans_silhouette_analysis.html#. Accessed: 2023-03-30.

[59] Scikit-Learn. Truncatedsvd. https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.TruncatedSVD.html. Accessed: 2023-03-30.

[60] Walid Shalaby, BahaaEddin AlAila, Mohammed Korayem, Layla Pournajaf, Khalifeh AlJadda, Shannon Quinn, and Wlodek Zadrozny. Help me find a job: A graph-based approach for job recommendation at scale. In *2017 IEEE international conference on big data (big data)*, pages 1544–1553. IEEE, 2017.

[61] MS Sodhi and Byung-Gak Son. Content analysis of or job advertisements to infer required skills. *Journal of the Operational Research Society*, 61(9):1315–1327, 2010.

[62] Wei Song and Soon Cheol Park. A novel document clustering model based on latent semantic analysis. In *Third International Conference on Semantics, Knowledge and Grid (SKG 2007)*, pages 539–542. IEEE, 2007.

[63] Gregory K Stephens and Bernadette Szajna. Perceptions and expectations: Why people choose a telecommuting work style. *International Journal of Electronic Commerce*, 3(1):70–85, 1998.

[64] Raphael Tackx, Fabien Tarissan, and Jean-Loup Guillaume. Comsim: a bipartite community detection algorithm using cycle and node's similarity. In *Complex Networks & Their Applications VI: Proceedings of Complex Networks 2017 (The Sixth International Conference on Complex Networks and Their Applications)*, pages 278–289. Springer, 2018.

[65] Hibiki Taguchi, Tsuyoshi Murata, and Xin Liu. Bimlpa: community detection in bipartite networks by multi-label propagation. In *Proceedings of NetSci-X 2020: Sixth International Winter School and Conference on Network Science 6*, pages 17–31. Springer, 2020.

[66] Glenn R Thiel and Nell T Hartley. Cooperative education: A natural synergy between business and academia. *SAM Advanced Management Journal*, 62(3):19, 1997.

[67] Andrew Toulis and Lukasz Golab. Graph mining to characterize competition for employment. In *Proceedings of the 2nd International Workshop on Network Data Analytics*, pages 1–7, 2017.

[68] V A Traag. Introduction. https://leidenalg.readthedocs.io/en/stable/intro.html, 2016. Accessed: 2023-03-30.

[69] Vincent A Traag, Ludo Waltman, and Nees Jan Van Eck. From louvain to leiden: guaranteeing well-connected communities. *Scientific reports*, 9(1):5233, 2019.

[70] Edy Umargono, Jatmiko Endro Suseno, and SK Vincensius Gunawan. K-means clustering optimization using the elbow method and early centroid determination based on mean and median formula. In *The 2nd International Seminar on Science and Technology (ISSTEC 2019)*, pages 121–129. Atlantis Press, 2020.

[71] Chirayu Upadhyay, Hasan Abu-Rasheed, Christian Weber, and Madjid Fathi. Explainable job-posting recommendations using knowledge graphs and named entity recognition. In *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 3291–3296. IEEE, 2021.

[72] Guandong Xu, Yu Zong, Peter Dolog, and Yanchun Zhang. Co-clustering analysis of weblogs using bipartite spectral projection approach. In *Knowledge-Based and Intelligent Information and Engineering Systems: 14th International Conference, KES 2010, Cardiff, UK, September 8-10, 2010, Proceedings, Part III 14*, pages 398–407. Springer, 2010.

[73] Katharina Anna Zweig and Michael Kaufmann. A systematic approach to the one-mode projection of bipartite graphs. *Social Network Analysis and Mining*, 1:187–218, 2011.

Appendix A

Logistic Regression Tables Including All Terms

A.1 Top 100 Terms Predicting Outstanding and Non-Outstanding Students

Full tables including all of the top 100 terms predicting outstanding and non-outstanding students after logistic regression analysis on supervisors' evaluation comments. Tables A.1, A.2, A.3, and A.4 correspond to Tables 3.2, 3.3, 3.4, and 3.5 in the main text (respectively), where less interesting terms were removed for brevity.

Table A.1: The top 100 terms predicting an outstanding student in 2021 evaluations of mostly remote positions, corresponding to Table 3.2 in the main text.

Rank	Term	Weight	Original Text
1	outstand	+16.846	"outstanding", "outstandingly", "outstand", "outstandings"
2	went	+5.509	"went"
3	expect	+4.510	"expectations", "expected", "expectation", "expect", "expecting"
4	student	+4.030	"student", "students", "studentitis"
5	compani	+4.025	"company", "companies", "companied", "companys"
6	amaz	+3.911	"amazing", "amazed", "amazingly", "amaze", "amazes"

7	outstand perform	+3.606	"outstanding performance", "outstanding performer", "outstanding in her performance", "outstanding in his performance", "outstanding for her performance"
8	incred	+3.450	"incredible", "incredibly"
9	exceed	+3.349	"exceeded", "exceed", "exceeding", "exceeds", "exceedingly"
10	intern	+3.313	"intern", "internal", "interns", "international", "internally"
11	extrem	+3.306	"extremely", "extreme", "extremly", "extremally", "extremes"
12	year	+3.269	"year", "years", "yearly"
13	outstand job	+3.204	"outstanding job", "outstanding in this job", "outstand job", "outstand jobs", "outstanding in her job"
14	lead	+3.127	"leadership", "lead", "leading", "leads", "leaded"
15	level	+3.013	"level", "levels", "levelled", "leveling", "levelling"
16	outstand rate	+2.949	"outstanding rating", "outstanding ratings", "outstanding rate", "outstand rating"
17	high	+2.946	"high", "highly", "highs"
18	gone	+2.922	"gone"
19	fulltim	+2.874	"fulltime", "fulltimer", "fulltimers", "fulltimes"
20	perform outstand	+2.866	"performance was outstanding", "performance is outstanding", "performance has been outstanding", "performed outstanding", "performance as outstanding"
21	employe	+2.745	"employee", "employees", "employeed"
22	rate	+2.726	"rating", "rate", "rated", "ratings", "rates"
23	product	+2.721	"product", "production", "productive", "products", "productivity"
24	entir	+2.691	"entire", "entirely", "entires", "entirity", "entirly"
25	deliv	+2.677	"deliver", "delivered", "delivering", "delivers", "delived"
26	everi	+2.673	"every"
27	problem	+2.670	"problem", "problems"
28	impress	+2.658	"impressed", "impressive", "impression", "impress", "impressively"
29	hire	+2.619	"hire", "hired", "hiring", "hires", "hirings"
30	took	+2.614	"took", "tooks"
31	coop student	+2.572	"coop student", "coop students", "coops students", "coops student", "coop a student"

32	engin	+2.450	“engineering”, “engineer”, “engineers”, “engine”, “engines”
33	design	+2.413	“design”, “designs”, “designing”, “designed”, “designer”
34	mani	+2.387	“many”
35	critic	+2.381	“critical”, “criticism”, “critically”, “criticisms”, “criticize”
36	implement	+2.321	“implementation”, “implement”, “implemented”, “implementing”, “implementations”
37	huge	+2.315	“huge”, “hugely”
38	goe	+2.307	“goes”
39	featur	+2.305	“features”, “feature”, “featured”, “featuring”, “featureful”
40	deserv	+2.302	“deserves”, “deserving”, “deserved”, “deserve”, “deservers”
41	impact	+2.289	“impact”, “impactful”, “impacted”, “impacts”, “impacting”
42	outstand work	+2.284	“outstanding work”, “outstanding to work”, “outstanding in his work”, “outstanding during his work”, “outstanding working”
43	servic	+2.224	“service”, “services”, “servicing”, “serviceability”, “serviceable”
44	way	+2.223	“way”, “ways”
45	custom	+2.175	“customer”, “customers”, “custom”, “customizing”, “customized”
46	provid	+2.123	“provided”, “provide”, “providing”, “provides”, “provider”
47	everyth	+2.104	“everything”
48	abil	+2.084	“ability”, “abilities”, “abillity”, “abillities”
49	exceed expect	+2.073	“exceeded expectations”, “exceeded our expectations”, “exceeded my expectations”, “exceeded all expectations”, “exceed expectations”
50	constant	+2.056	“constantly”, “constant”, “constants”
51	initi	+2.046	“initiative”, “initial”, “initiatives”, “initially”, “initiate”
52	innov	+2.034	“innovative”, “innovation”, “innovate”, “innovations”, “innovating”
53	solut	+2.030	“solutions”, “solution”, “solutioning”, “solutation”, “solutioned”
54	coop	+2.028	“coop”, “coops”
55	signific	+2.009	“significant”, “significantly”, “significance”
56	multipl	+2.000	“multiple”, “multiples”, “multipled”, “multiplicative”

57	perfect	+1.998	"perfect", "perfectly", "perfection", "perfecting", "perfected"
58	dedic	+1.993	"dedication", "dedicated", "dedicate", "dedicating", "dedicates"
59	senior	+1.960	"senior", "seniors", "seniority", "seniorities"
60	comment	+1.942	"comments", "comment", "commented", "commenting", "commentable"
61	signific exceed	+1.937	"significantly exceeded", "significantly exceeding", "significantly exceed", "significantly exceeds"
62	tremend	+1.913	"tremendous", "tremendously", "tremendeous"
63	quick	+1.898	"quickly", "quick", "quickness"
64	expect coop	+1.887	"expectations for a coop", "expectations of a coop", "expected of a coop", "expected from a coop", "expect from a coop"
65	best coop	+1.883	"best coop", "best coops", "best of any of the coop"
66	best	+1.869	"best", "bested", "bests"
67	rare	+1.866	"rare", "rarely", "rares", "raring"
68	far	+1.852	"far"
69	outstand coop	+1.823	"outstanding coop", "outstanding coops", "outstanding during her coop", "outstanding for her coop", "outstanding in his coop"
70	new	+1.794	"new", "newness"
71	trust	+1.785	"trust", "trusted", "trusting", "trusts"
72	absolut	+1.780	"absolute", "absolutely", "absolutly", "absoluted"
73	organ	+1.771	"organization", "organized", "organizing", "organize", "organizations"
74	scope	+1.770	"scope", "scoping", "scoped", "scopes"
75	surpass	+1.751	"surpassed", "surpassing", "surpass", "surpasses"
76	alway	+1.745	"always", "alway", "alwayes"
77	code	+1.740	"code", "coding", "codes", "coded", "codeing"
78	alreadi	+1.731	"already"
79	fantast	+1.716	"fantastic", "fantastically", "fantastics"
80	finish	+1.715	"finish", "finished", "finishing", "finishes", "finisheds"
81	say	+1.713	"say", "saying", "says"
82	love	+1.701	"love", "loved", "loves", "lovely", "loving"
83	amaz job	+1.687	"amazing job", "amazing jobs"
84	truli	+1.683	"truly"

85	highest	+1.683	"highest"
86	consist	+1.664	"consistently", "consistent", "consistency", "consistantly", "consisted"
87	cultur	+1.638	"culture", "cultural", "cultures", "culturally", "cultureal"
88	project	+1.614	"project", "projects", "projecting", "projection", "projections"
89	day	+1.599	"day", "days"
90	offer	+1.590	"offer", "offered", "offering", "offers", "offerings"
91	qualiti	+1.564	"quality", "qualities"
92	independ	+1.554	"independently", "independent", "independence", "independantly", "independant"
93	taken	+1.547	"taken"
94	went expect	+1.537	"went above expectations", "went above our expectations", "went above what was expected", "went above my expectations", "went above all expectations"
95	major	+1.532	"major", "majority", "majoring", "majors"
96	model	+1.524	"model", "models", "modeling", "modelling", "modelled"
97	care	+1.503	"care", "carefully", "careful", "cares", "caring"
98	complex	+1.490	"complex", "complexity", "complexities", "complexed", "complexes"
99	fast	+1.485	"fast", "fastly", "fasted"
100	return	+1.480	"return", "returning", "returns", "returned", "returner"

Table A.2: The top 100 terms predicting a non-outstanding student in 2021 evaluations of mostly remote positions, corresponding to Table 3.3 in the main text.

Rank	Term	Weight	Original Text
1	overal	-4.162	"overall"
2	good	-4.126	"good", "goods", "goodly", "goodness"
3	term	-2.615	"term", "terms", "termly", "termed"
4	excel job	-2.520	"excellent job", "excelent job", "excellent jobs", "excelled at her job", "excellant job"
5	met	-2.317	"met"
6	work	-2.304	"work", "working", "worked", "works", "workings"
7	pleasur	-2.185	"pleasure", "pleasurable", "pleasuring", "pleasureable", "pleasureful"

8	remot	-1.980	"remote", "remotely"
9	thank	-1.940	"thank", "thanks", "thankful", "thanking", "thankfully"
10	work term	-1.934	"work term", "work terms", "work this term", "work with this term", "working with you this term"
11	howev	-1.892	"however"
12	encourag	-1.799	"encourage", "encouraged", "encouraging", "encouragement", "encourages"
13	perform excel	-1.793	"performance was excellent", "performance is excellent", "performed excellently", "performance has been excellent", "performed excellent"
14	<BIAS>	-1.763	
15	good job	-1.758	"good job", "good a job", "good jobs", "good at her job", "good at his job"
16	met expect	-1.695	"met expectations", "met all expectations", "met our expectations", "met the expectations", "met my expectations"
17	xd br	-1.694	"xd br", "xd same as above br", "xd and br", "xd so br", "xd these are br"
18	appreci	-1.666	"appreciated", "appreciate", "appreciation", "appreciative", "appreciates"
19	pleas	-1.642	"pleased", "please", "pleasing", "pleases"
20	progress	-1.596	"progress", "progressed", "progresses", "progression", "progressing"
21	unfortun	-1.551	"unfortunately", "unfortunate"
22	struggl	-1.381	"struggled", "struggle", "struggling", "struggles"
23	excel perform	-1.354	"excellent performance", "excellent performer", "excelent performance", "excellent she performed", "excel he perform"
24	delight	-1.342	"delighted", "delight", "delightful", "delights", "delightfully"
25	ask question	-1.333	"ask questions", "asking questions", "ask more questions", "asked questions", "asks questions"
26	great addit	-1.298	"great addition", "great additional", "great a addition", "great additions", "great but it s what she did in addition"
27	exceed perform	-1.296	"exceeded all performance", "exceeded performance", "exceeded some performance", "exceeded the performance", "exceeded our performance"
28	excel coop	-1.274	"excellent coop", "excel at this coop", "excelled in the coop", "excelled in this coop", "excellent during his coop"

29	quit	-1.256	"quite", "quit", "quitely", "quits", "quitting"
30	workterm	-1.241	"workterm", "workterms"
31	enjoy	-1.235	"enjoyed", "enjoy", "enjoys", "enjoyable", "enjoying"
32	best futur	-1.207	"best in her future", "best in his future", "best in your future", "best in the future", "best in future"
33	met exceed	-1.201	"met and exceeded", "met or exceeded", "met all and exceeded", "met but exceeded", "met exceeded"
34	friend	-1.120	"friendly", "friends", "friend"
35	great coop	-1.113	"great coop", "great during her coop", "greatly while on her coop", "great at the coop", "great but as a coop"
36	past month	-1.087	"past months", "past few months", "past month", "past few month", "past a few months"
37	task	-1.085	"tasks", "task", "tasked", "tasking", "taskings"
38	confid	-1.072	"confidence", "confident", "confidently", "confidant", "confidance"
39	assign time	-1.069	"assignments on time", "assigned to him in a timely", "assignments in a timely", "assigned to her in a timely", "assignment on time"
40	demonstr good	-1.043	"demonstrated good", "demonstrated a good", "demonstrated very good", "demonstrates good", "demonstrate good"
41	time manner	-1.037	"timely manner", "time manner", "timely manners", "time in the manner", "timing manner"
42	work home	-1.037	"working from home", "work from home", "worked from home", "working at home", "works from home"
43	overal pleas	-1.031	"overall i am very pleased", "overall we are very pleased", "overall we were very pleased", "overall very pleased", "overall i was very pleased"
44	comfort	-1.024	"comfortable", "comfort", "comfortably", "comfortability", "comforted"
45	email	-1.019	"email", "emails", "emailing", "emailed"
46	expos	-1.018	"exposed", "expose", "exposing", "exposes", "exposer"
47	workload	-1.011	"workload", "workloads"
48	meet expect	-1.008	"meet expectations", "meeting expectations", "meet the expectations", "meets expectations", "meet our expectations"
49	learn	-1.008	"learn", "learning", "learned", "learns", "learnings"
50	lack	-0.997	"lack", "lacking", "lacks", "lacked"
51	gain	-0.995	"gain", "gained", "gaining", "gains", "gainned"

52	virtual	-0.992	“virtual”, “virtually”, “virtualization”
53	posit attitud	-0.987	“positive attitude”, “positive attitudes”, “positive can do attitude”, “positive and can do attitude”, “positive attitud”
54	second half	-0.982	“second half”
55	excel addit	-0.981	“excellent addition”, “excellent in addition”, “excellent additionally”, “excellent additions”, “excellent the additional”
56	offic	-0.973	“office”, “offices”, “officer”, “officers”, “offics”
57	demonstr strong	-0.970	“demonstrated strong”, “demonstrated a strong”, “demonstrates strong”, “demonstrated very strong”, “demonstrated a very strong”
58	half	-0.954	“half”, “halfs”
59	remot work	-0.944	“remote work”, “remote working”, “remotely working”, “remotely for this work”, “remote from work”
60	great	-0.937	“great”, “greatly”, “greatness”, “greatful”, “greate”
61	andrew	-0.936	“andrew”, “andrews”
62	complet task	-0.926	“complete tasks”, “completing tasks”, “completed tasks”, “complete the tasks”, “completed all tasks”
63	bit	-0.923	“bit”, “bits”
64	hard worker	-0.922	“hard worker”, “hard of a worker”
65	varieti	-0.914	“variety”, “varieties”
66	kevin	-0.911	“kevin”
67	desir	-0.901	“desire”, “desired”, “desires”, “desirable”, “desiring”
68	condit	-0.899	“conditions”, “condition”, “conditioning”, “conditional”
69	midterm	-0.897	“midterm”, “midterms”
70	work deliv	-0.882	“work delivered”, “work he delivered”, “work and delivered”, “work she delivered”, “work and delivering”
71	improv	-0.879	“improve”, “improvement”, “improved”, “improving”, “improvements”
72	john	-0.860	“john”, “johns”
73	valu member	-0.850	“valued member”, “value for members”, “valued by all members”
74	work abl	-0.842	“work and was able”, “work she was able”, “work he was able”, “work and is able”, “work she is able”
75	coop term	-0.841	“coop term”, “coop terms”, “coop this term”, “coops this term”, “coops term”
76	nice	-0.838	“nice”, “nicely”

77	improv term	-0.819	“improved over the term”, “improvement over the term”, “improvement in terms”, “improved as the term”, “improvement this term”
78	good perform	-0.819	“good performance”, “good performer”, “good at performing”, “good he performed”, “good her performance”
79	given work	-0.813	“given work”, “given the work”, “given and worked”, “given her work”, “given his work”
80	jason	-0.811	“jason”, “jasons”
81	term abl	-0.811	“term he was able”, “term she was able”, “term and was able”, “term he is able”, “term we were able”
82	ask help	-0.806	“ask for help”, “asking for help”, “asked for help”, “asks for help”, “asked to help”
83	display great	-0.802	“displayed great”, “displayed a great”, “displays a great”, “displaying great”, “display great”
84	thank hard	-0.792	“thank you for all your hard”, “thank you for your hard”, “thanks for all your hard”, “thank you for all of your hard”, “thanks for all of your hard”
85	forward futur	-0.792	“forward to future”, “forward to his future”, “forward for future”, “forward in future”, “forward in your future”
86	grow	-0.788	“grow”, “growing”, “grows”, “growed”
87	cours term	-0.783	“course of the term”, “course of his term”, “course of her term”, “course of this term”, “course this term”
88	abl make	-0.782	“able to make”, “able make”, “able to makes”, “able to making”, “able to only make”
89	daniel	-0.778	“daniel”, “daniell”, “daniels”, “daniele”
90	lot potenti	-0.776	“lot of potential”, “lots of potential”, “lot of potentials”, “lot potential”, “lots of potentials”
91	excel	-0.765	“excellent”, “excel”, “excelled”, “excellence”, “excels”
92	excel exceed	-0.759	“excellent and exceeded”, “excellent and he has exceeded”, “excellent exceeding”, “excellent and exceeds”, “excel and exceeded”
93	excel term	-0.756	“excellent term”, “excelled this term”, “excellent this term”, “excelled at this term”, “excelled in this term”
94	engag	-0.755	“engaged”, “engagement”, “engage”, “engaging”, “engagements”
95	follow instruct	-0.748	“follow instructions”, “followed instructions”, “follows instructions”, “following instructions”, “follow the instructions”
96	need improv	-0.748	“needs to improve”, “need to improve”, “needs improvement”, “need improvement”, “needed to improve”

97	perform task	-0.735	"perform tasks", "performing tasks", "perform the tasks", "performed all tasks", "performed tasks"
98	great asset	-0.732	"great asset", "great assets", "great and an asset", "great an asset", "great she was an asset"
99	wish best	-0.731	"wish her all the best", "wish him all the best", "wish you all the best", "wish her the best", "wish him the best"
100	manag workload	-0.723	"manage her workload", "manage his workload", "managed her workload", "manage workload", "manage workloads"

Table A.3: The top 100 terms predicting an outstanding student in 2019 evaluations of mostly in person positions, corresponding to Table 3.4 in the main text.

Rank	Term	Weight	Original Text
1	outstand	+17.581	"outstanding", "outstandingly", "outstand", "outstandings"
2	went	+5.133	"went"
3	expect	+4.489	"expectations", "expected", "expectation", "expect", "expecting"
4	exceed	+4.123	"exceeded", "exceed", "exceeding", "exceeds", "exceedingly"
5	extrem	+4.030	"extremely", "extreme", "extremly", "extremally", "extremes"
6	fulltim	+3.958	"fulltime", "fulltimer", "fulltimers", "fulltimes"
7	amaz	+3.622	"amazing", "amazed", "amazingly", "amaze", "amazes"
8	level	+3.592	"level", "levels", "levelled", "leveling", "levelling"
9	comment	+3.448	"comments", "comment", "commented", "commenting", "commentable"
10	outstand perform	+3.434	"outstanding performance", "outstanding performer", "outstanding in her performance", "outstanding in his performance", "outstanding for her performance"
11	exceed expect	+3.425	"exceeded expectations", "exceeded our expectations", "exceeded my expectations", "exceeded all expectations", "exceed expectations"
12	gone	+3.405	"gone"
13	compani	+3.398	"company", "companies", "companied", "companys"
14	hire	+3.287	"hire", "hired", "hiring", "hires", "hirings"
15	lead	+3.236	"leadership", "lead", "leading", "leads", "leaded"

16	outstand rate	+3.233	“outstanding rating”, “outstanding ratings”, “outstanding rate”, “outstand rating”
17	outstand job	+3.213	“outstanding job”, “outstanding in this job”, “outstand job”, “outstand jobs”, “outstanding in her job”
18	high	+3.050	“high”, “highly”, “highs”
19	intern	+3.036	“intern”, “internal”, “interns”, “international”, “internally”
20	everi	+2.942	“every”
21	rate	+2.804	“rating”, “rate”, “rated”, “ratings”, “rates”
22	deserv	+2.781	“deserves”, “deserving”, “deserved”, “deserve”, “deservers”
23	coop student	+2.747	“coop student”, “coop students”, “coops students”, “coops student”, “coop a student”
24	outstand work	+2.719	“outstanding work”, “outstanding to work”, “outstanding in his work”, “outstanding during his work”, “outstanding working”
25	engin	+2.643	“engineering”, “engineer”, “engineers”, “engine”, “engines”
26	perform outstand	+2.534	“performance was outstanding”, “performance is outstanding”, “performance has been outstanding”, “performed outstanding”, “performance as outstanding”
27	year	+2.532	“year”, “years”, “yearly”
28	employe	+2.526	“employee”, “employees”, “employeed”
29	best	+2.490	“best”, “bested”, “bests”
30	truli	+2.485	“truly”
31	solut	+2.450	“solutions”, “solution”, “solutioning”, “solutation”, “solutioned”
32	incred	+2.423	“incredible”, “incredibly”
33	idea	+2.368	“ideas”, “idea”
34	tremend	+2.336	“tremendous”, “tremendously”, “tremendeous”
35	graduat	+2.316	“graduate”, “graduation”, “graduates”, “graduated”, “graduating”
36	impress	+2.304	“impressed”, “impressive”, “impression”, “impress”, “impressively”
37	took	+2.266	“took”, “tooks”
38	complex	+2.265	“complex”, “complexity”, “complexities”, “complexed”, “complexes”
39	student	+2.211	“student”, “students”, “studentitis”

40	design	+2.200	“design”, “designs”, “designing”, “designed”, “designer”
41	fantast	+2.189	“fantastic”, “fantastically”, “fantastics”
42	realli	+2.168	“really”
43	critic	+2.167	“critical”, “criticism”, “critically”, “criticisms”, “criticize”
44	coop	+2.165	“coop”, “coops”
45	everyth	+2.150	“everything”
46	entir	+2.136	“entire”, “entirely”, “entires”, “entirity”, “entirly”
47	far	+2.135	“far”
48	perfect	+2.133	“perfect”, “perfectly”, “perfection”, “perfecting”, “perfected”
49	respons	+2.130	“responsibilities”, “responsibility”, “responsible”, “responsive”, “response”
50	impact	+2.121	“impact”, “impactful”, “impacted”, “impacts”, “impacting”
51	love	+2.119	“love”, “loved”, “loves”, “lovely”, “loving”
52	miss	+2.112	“missed”, “miss”, “missing”, “misses”
53	alreadi	+2.059	“already”
54	deliv	+2.049	“deliver”, “delivered”, “delivering”, “delivers”, “delived”
55	innov	+2.039	“innovative”, “innovation”, “innovate”, “innovations”, “innovating”
56	consist	+2.026	“consistently”, “consistent”, “consistency”, “consistantly”, “consisted”
57	super	+2.026	“super”
58	signific	+2.024	“significant”, “significantly”, “significance”
59	hope	+2.006	“hope”, “hopefully”, “hoping”, “hoped”, “hopes”
60	abil	+2.001	“ability”, “abilities”, “abillity”, “abillities”
61	lucki	+2.000	“lucky”
62	insight	+1.975	“insights”, “insight”, “insightful”, “insightfulness”
63	develop	+1.929	“development”, “develop”, “developing”, “developed”, “developer”
64	way	+1.928	“way”, “ways”
65	implement	+1.926	“implementation”, “implement”, “implemented”, “implementing”, “implementations”
66	product	+1.925	“product”, “production”, “productive”, “products”, “productivity”
67	problem	+1.922	“problem”, “problems”
68	invalu	+1.919	“invaluable”

69	surpass	+1.915	“surpassed”, “surpassing”, “surpass”, “surpasses”
70	execut	+1.908	“execute”, “execution”, “executing”, “executed”, “executive”
71	everi task	+1.901	“every task”, “every tasks”
72	oustand	+1.888	“oustanding”
73	alway	+1.860	“always”, “alway”, “alwayes”
74	demonstr outstand	+1.856	“demonstrated outstanding”, “demonstrated an outstanding”, “demonstrates an outstanding”, “demonstrated his outstanding”, “demonstrating an outstanding”
75	come	+1.855	“come”, “comes”, “coming”, “comings”
76	second	+1.854	“second”, “secondly”, “seconds”, “seconded”, “secondement”
77	talent	+1.850	“talented”, “talent”, “talents”
78	contribut	+1.838	“contributions”, “contributed”, “contribute”, “contribution”, “contributing”
79	offer	+1.836	“offer”, “offered”, “offering”, “offers”, “offerings”
80	featur	+1.835	“features”, “feature”, “featured”, “featuring”, “featureful”
81	result	+1.834	“results”, “result”, “resulted”, “resulting”, “resultant”
82	mani	+1.832	“many”
83	rare	+1.827	“rare”, “rarely”, “rares”, “raring”
84	research	+1.805	“research”, “researching”, “researched”, “researcher”, “researchers”
85	senior	+1.801	“senior”, “seniors”, “seniority”, “seniorities”
86	comment submit	+1.795	“comments submitted”
87	absolut	+1.781	“absolute”, “absolutely”, “absolutly”, “absoluted”
88	project	+1.768	“project”, “projects”, “projecting”, “projection”, “projections”
89	dedic	+1.767	“dedication”, “dedicated”, “dedicate”, “dedicating”, “dedicates”
90	highest	+1.765	“highest”
91	pleas comment	+1.735	“please comment”
92	exemplari	+1.728	“exemplary”
93	step	+1.725	“step”, “steps”, “stepped”, “stepping”
94	went expect	+1.722	“went above expectations”, “went above our expectations”, “went above what was expected”, “went above my expectations”, “went above all expectations”
95	make	+1.722	“make”, “making”, “makes”, “makings”, “maked”

96	challeng	+1.719	“challenges”, “challenging”, “challenge”, “challenged”, “challenger”
97	code	+1.711	“code”, “coding”, “codes”, “coded”, “codeing”
98	short	+1.700	“short”, “shortly”, “shorting”, “shorts”, “shorted”
99	built	+17.581	“built”, “builts”, “builting”
100	surpass expect	+5.133	“surpassed expectations”, “surpassed our expectations”, “surpassed my expectations”, “surpassed the expectations”, “surpassed all expectations”

Table A.4: The top 100 terms predicting a non-outstanding student in 2019 evaluations of mostly in person positions, corresponding to Table 3.5 in the main text.

Rank	Term	Weight	Original Text
1	good	-4.311	“good”, “goods”, “goodly”, “goodness”
2	overal	-2.428	“overall”
3	term	-2.195	“term”, “terms”, “termly”, “termed”
4	work	-2.148	“work”, “working”, “worked”, “works”, “workings”
5	howev	-2.144	“however”
6	met expect	-2.138	“met expectations”, “met all expectations”, “met our expectations”, “met the expectations”, “met my expectations”
7	good job	-2.131	“good job”, “good a job”, “good jobs”, “good at her job”, “good at his job”
8	<BIAS>	-2.006	
9	met	-1.999	“met”
10	task	-1.860	“tasks”, “task”, “tasked”, “tasking”, “taskings”
11	appreci	-1.775	“appreciated”, “appreciate”, “appreciation”, “appreciative”, “appreciates”
12	excel perform	-1.669	“excellent performance”, “excellent performer”, “excelent performance”, “excellent she performed”, “excel he perform”
13	xd br	-1.658	“xd br”, “xd same as above br”, “xd and br”, “xd so br”, “xd these are br”
14	ask question	-1.642	“ask questions”, “asking questions”, “ask more questions”, “asked questions”, “asks questions”
15	general	-1.513	“general”, “generally”, “generalize”, “generalized”, “generalization”

16	lack	-1.506	“lack”, “lacking”, “lacks”, “lacked”
17	meet expect	-1.440	“meet expectations”, “meeting expectations”, “meet the expectations”, “meets expectations”, “meet our expectations”
18	work term	-1.435	“work term”, “work terms”, “work this term”, “work with this term”, “working with you this term”
19	updat	-1.422	“updates”, “update”, “updating”, “updated”, “updater”
20	complet assign	-1.399	“complete assigned”, “complete the assigned”, “completed all assigned”, “completing assigned”, “completed assigned”
21	demonstr good	-1.397	“demonstrated good”, “demonstrated a good”, “demonstrated very good”, “demonstrates good”, “demonstrate good”
22	feedback	-1.389	“feedback”, “feedbacks”
23	excel job	-1.379	“excellent job”, “excelent job”, “excellent jobs”, “excelled at her job”, “excellant job”
24	quiet	-1.377	“quiet”, “quietly”, “quietness”
25	great addit	-1.377	“great addition”, “great additional”, “great a addition”, “great additions”, “great but it s what she did in addition”
26	encourag	-1.372	“encourage”, “encouraged”, “encouraging”, “encouragement”, “encourages”
27	quit	-1.329	“quite”, “quit”, “quitely”, “quits”, “quitting”
28	pleasur	-1.323	“pleasure”, “pleasurable”, “pleasuring”, “pleasureable”, “pleasureful”
29	work good	-1.287	“work was very good”, “work and good”, “work good”, “work was good”, “work with a good”
30	thank	-1.277	“thank”, “thanks”, “thankful”, “thanking”, “thankfully”
31	excel coop	-1.260	“excellent coop”, “excel at this coop”, “excelled in the coop”, “excelled in this coop”, “excellent during his coop”
32	wish best	-1.222	“wish her all the best”, “wish him all the best”, “wish you all the best”, “wish her the best”, “wish him the best”
33	sometim	-1.210	“sometimes”, “sometime”
34	gain	-1.209	“gain”, “gained”, “gaining”, “gains”, “gainned”
35	good perform	-1.181	“good performance”, “good performer”, “good at performing”, “good he performed”, “good her performance”
36	wish	-1.176	“wish”, “wishes”, “wishing”, “wished”
37	struggl	-1.163	“struggled”, “struggle”, “struggling”, “struggles”
38	perform excel	-1.146	“performance was excellent”, “performance is excellent”, “performed excellently”, “performance has been excellent”, “performed excellent”

39	friend	-1.138	“friendly”, “friends”, “friend”
40	slow	-1.136	“slow”, “slowing”, “slowed”, “slows”, “slowness”
41	expos	-1.113	“exposed”, “expose”, “exposing”, “exposes”, “exposer”
42	assign task	-1.095	“assigned tasks”, “assigned task”, “assigned a task”, “assignments and tasks”, “assign tasks”
43	perform good	-1.067	“performance was very good”, “performance was good”, “performance has been very good”, “performance is good”, “performance is very good”
44	bit	-1.063	“bit”, “bits”
45	coop term	-1.060	“coop term”, “coop terms”, “coop this term”, “coops this term”, “coops term”
46	xd overal	-1.058	“xd overall”, “xd his overall”, “xd her overall”, “xd the overall”, “xd an overall”
47	addit team	-1.057	“addition to our team”, “addition to the team”, “addition to any team”, “addition to my team”, “addition to a team”
48	new task	-1.006	“new tasks”, “new task”, “new he was tasked”, “new to a task”
49	pleasant	-0.979	“pleasant”, “pleasantly”
50	assign	-0.973	“assigned”, “assignments”, “assignment”, “assign”, “assigning”
51	good term	-0.966	“good term”, “good in terms”, “good this term”, “good and as the term”, “good during the term”
52	xd area	-0.965	“xd areas”, “xd an area”, “xd the only area”, “xd for areas”, “xd some areas”
53	instruct	-0.959	“instructions”, “instruction”, “instructional”, “instructed”, “instructing”
54	feedback receiv	-0.954	“feedback received”, “feedback he received”, “feedback she received”, “feedback i received”, “feedback she receives”
55	improv	-0.950	“improve”, “improvement”, “improved”, “improving”, “improvements”
56	ticket	-0.946	“tickets”, “ticket”, “ticketing”, “ticketed”
57	comfort	-0.946	“comfortable”, “comfort”, “comfortably”, “comfortability”, “comforted”
58	file	-0.940	“files”, “file”, “filing”, “filed”, “filings”
59	varieti	-0.936	“variety”, “varieties”
60	jack	-0.933	“jack”, “jacks”
61	meet	-0.918	“meetings”, “meet”, “meeting”, “meets”
62	error	-0.917	“errors”, “error”

63	area improv	-0.916	“areas of improvement”, “areas for improvement”, “area of improvement”, “area for improvement”, “areas to improve”
64	receiv	-0.910	“received”, “receiving”, “receive”, “receives”, “receivable”
65	understand task	-0.906	“understand the task”, “understand the tasks”, “understanding of the task”, “understanding of the tasks”, “understand tasks”
66	engag	-0.901	“engaged”, “engagement”, “engage”, “engaging”, “engagements”
67	met exceed	-0.899	“met and exceeded”, “met or exceeded”, “met all and exceeded”, “met but exceeded”, “met exceeded”
68	excel addit	-0.884	“excellent addition”, “excellent in addition”, “excellent additionally”, “excellent additions”, “excellent the additional”
69	various task	-0.879	“various tasks”, “various other tasks”, “various task”
70	semest	-0.871	“semester”, “semesters”
71	br xd	-0.866	“br xd”
72	unfortun	-0.863	“unfortunately”, “unfortunate”
73	fair	-0.861	“fairly”, “fair”, “faire”, “fairness”, “fairs”
74	br	-0.857	“br”
75	construct	-0.854	“constructive”, “construction”, “constructively”, “constructed”, “constructing”
76	confid	-0.843	“confidence”, “confident”, “confidently”, “confidant”, “confidance”
77	valuabl addit	-0.838	“valuable addition”, “valuable additions”
78	eric	-0.834	“eric”
79	clarif	-0.834	“clarification”, “clarifications”
80	time manner	-0.832	“timely manner”, “time manner”, “timely manners”, “time in the manner”, “timing manner”
81	xd	-0.824	“xd”
82	team	-0.824	“team”, “teams”, “teaming”, “teamed”
83	demonstr willing	-0.806	“demonstrated a willingness”, “demonstrated willingness”, “demonstrated his willingness”, “demonstrates a willingness”, “demonstrated the willingness”
84	great job	-0.804	“great job”, “great jobs”, “great a job”, “great at his job”, “great in any job”
85	team summer	-0.800	“team this summer”, “team for the summer”, “team over the summer”, “team during the summer”, “team during his summer”

86	futur career	-0.798	“future career”, “future careers”, “futur career”, “future and career”, “future in her career”
87	work independ	-0.798	“work independently”, “worked independently”, “working independently”, “works independently”, “worked very independently”
88	divers	-0.798	“diverse”, “diversity”, “diversed”, “diversely”, “diversities”
89	observ	-0.790	“observed”, “observe”, “observations”, “observation”, “observing”
90	welcom addit	-0.785	“welcome addition”, “welcomed addition”, “welcome additional”, “welcomed additional”, “welcoming addition”
91	great coop	-0.782	“great coop”, “great during her coop”, “greatly while on her coop”, “great at the coop”, “great but as a coop”
92	team month	-0.775	“team for months”, “team in the months”, “team months”, “team over her month”, “team during her month”
93	assist	-0.774	“assist”, “assistance”, “assisted”, “assistant”, “assisting”
94	strength	-0.773	“strengths”, “strength”
95	especi	-0.771	“especially”
96	daniel	-0.770	“daniel”, “daniell”, “daniels”, “daniele”
97	inform	-0.765	“information”, “informed”, “inform”, “informative”, “informal”
98	focus	-0.761	“focus”, “focused”, “focusing”, “focuses”, “focuse”
99	nice	-0.760	“nice”, “nicely”
100	luck futur	-0.756	“luck in the future”, “luck in your future”, “luck in her future”, “luck in his future”, “luck with your future”

A.2 Top 100 Terms Predicting Remote and In Person Recommendations

Full tables including all of the top 100 terms predicting remote and in-person students after logistic regression analysis on supervisors’ recommendation comments. Tables A.5 and A.6 correspond to Tables 3.7 and 3.8 in the main text (respectively), where less interesting terms were removed for brevity.

Table A.5: The top 100 terms predicting a remote student (2021) from the supervisors' recommendation comments, corresponding to Table 3.7 in the main text.

Rank	Term	Weight	Original Text
1	nt	+8.232	"nt"
2	collabor	+4.979	"collaboration", "collaborative", "collaborate", "collaborated", "collaborating"
3	innov	+4.663	"innovative", "innovation", "innovate", "innovations", "innovating"
4	remot	+4.060	"remote", "remotely"
5	mention	+3.888	"mentioned", "mention", "mentioning", "mentionned", "mentions"
6	critic think	+3.829	"critical thinking", "critically think", "critically thinking", "critical think", "critical thinkings"
7	area develop	+3.464	"areas for development", "areas of development", "area of development", "area for development", "areas to develop"
8	mindset	+3.391	"mindset", "mindsets"
9	critic	+3.293	"critical", "criticism", "critically", "criticisms", "criticize"
10	covid	+3.072	"covid", "covidence"
11	comment	+2.741	"comments", "comment", "commented", "commenting", "commentable"
12	selfmanag	+2.664	"selfmanagement", "selfmanage", "selfmanaged", "selfmanaging", "selfmanageable"
13	virtual	+2.551	"virtual", "virtually", "virtualization"
14	technolog	+2.432	"technologies", "technology", "technological", "technologically", "technolog"
15	nt afraid	+2.385	"nt be afraid", "nt afraid"
16	learn	+2.353	"learn", "learning", "learned", "learns", "learnings"
17	curios	+2.311	"curiosity", "curiosities"
18	workplac	+2.241	"workplace", "workplaces", "workplacements"
19	section	+2.151	"section", "sections"
20	curious	+2.134	"curious", "curiousity", "curiouse", "curiousness", "curiousities"
21	selfassess	+2.114	"selfassessment", "selfassess", "selfassessments", "selfassessed", "selfassessing"
22	technic skill	+2.112	"technical skills", "technical skill", "technically skilled", "technical and other skills", "technical and skill"
23	think	+2.057	"think", "thinking", "thinks", "thinkings"

24	implement	+2.052	"implementation", "implement", "implemented", "implementing", "implementations"
25	remot work	+2.003	"remote work", "remote working", "remotely working", "remotely for this work", "remote from work"
26	hope	+1.924	"hope", "hopefully", "hoping", "hoped", "hopes"
27	disciplin	+1.888	"discipline", "disciplines", "disciplined"
28	reach	+1.877	"reach", "reaching", "reached", "reaches"
29	support	+1.870	"support", "supporting", "supported", "supportive", "supports"
30	comment includ	+1.858	"comments not included", "comments included", "comments include", "comments and including", "comments i ll include"
31	innov mindset	+1.849	"innovation mindset", "innovative mindset", "innovative mindsets", "innovate mindset", "innovating mindset"
32	career	+1.847	"career", "careers"
33	data	+1.835	"data", "datas"
34	career develop	+1.830	"career development", "career develops", "career and development", "career as a developer", "career develop"
35	work remot	+1.655	"working remotely", "work remotely", "worked remotely", "working remote", "working in a remote"
36	communic	+1.647	"communication", "communicate", "communicating", "communications", "communicated"
37	area	+1.640	"areas", "area"
38	develop area	+1.601	"development areas", "development area", "develop in this area", "develop in the areas", "development in this area"
39	pandem	+1.581	"pandemic"
40	lifelong	+1.571	"lifelong"
41	agil	+1.557	"agile", "agility"
42	opportun	+1.525	"opportunities", "opportunity", "opportune"
43	cover	+1.490	"cover", "covered", "covering", "covers"
44	world	+1.481	"world", "worlds", "worldly"
45	want	+1.428	"want", "wants", "wanted", "wanting"
46	previous	+1.377	"previous", "previously"
47	good communic	+1.376	"good communication", "good communicator", "good at communicating", "good at communication", "good communications"
48	think critic	+1.371	"think critically", "thinking critically", "thinks critically", "think more critically", "think about critical"

49	learn opportun	+1.369	"learning opportunities", "learning opportunity", "learn from all opportunities", "learn from the opportunities", "learn from them they are opportunities"
50	skillset	+1.360	"skillset", "skillsets", "skillsetabilities"
51	self manag	+1.332	"self management", "self manage", "self managed", "self managing", "self manages"
52	opportun learn	+1.322	"opportunities to learn", "opportunity to learn", "opportunities for learning", "opportunities and learning", "opportunity for learning"
53	make suggest	+1.313	"make suggestions", "making suggestions", "make more suggestions", "make a suggestion", "make some suggestions"
54	ethan	+1.286	"ethan", "ethans"
55	work area	+1.275	"work area", "work in this area", "work on areas", "working on the areas", "work in areas"
56	onlin	+1.268	"online", "onlined"
57	develop	+1.260	"development", "develop", "developing", "developed", "developer"
58	question	+1.230	"questions", "question", "questioning", "questioned", "questionable"
59	mani	+1.229	"many"
60	skill experi	+1.225	"skills and experience", "skills and experiences", "skills through experience", "skills with experience", "skills with more experience"
61	relev	+1.224	"relevant", "relevance", "relevent", "relevancy", "releve"
62	strength	+1.212	"strengths", "strength"
63	futur internship	+1.211	"future internships", "future internship", "future during her internship", "future over the internship"
64	anoth work	+1.195	"another work", "another for work", "another she worked", "another the work", "another your work"
65	pleasur	+1.189	"pleasure", "pleasurable", "pleasuring", "pleasureable", "pleasureful"
66	love	+1.187	"love", "loved", "loves", "lovely", "loving"
67	hire	+1.175	"hire", "hired", "hiring", "hires", "hirings"
68	collabor communic	+1.166	"collaboration and communication", "collaborate and communicate", "collaboration communication", "collaborated and communicated", "collaborating and communicating"
69	cpa	+1.166	"cpa"
70	train	+1.161	"training", "train", "trained", "trainings", "trains"

71	communic collabor	+1.160	"communication and collaboration", "communication collaboration", "communicate and collaborate", "communicating and collaborating", "communicator and collaborator"
72	inform	+1.155	"information", "informed", "inform", "informative", "informal"
73	journey	+1.152	"journey", "journeys"
74	spring	+1.147	"spring"
75	necessari	+1.137	"necessary"
76	literaci	+1.137	"literacy"
77	pleas	+1.133	"pleased", "please", "pleasing", "pleases"
78	academ	+1.130	"academic", "academics", "academically", "academe", "academical"
79	continu increas	+1.127	"continue to increase", "continue increasing", "continued to increase", "continues to increase", "continually increase"
80	overcom	+1.126	"overcome", "overcoming", "overcomes", "overcomed", "overcomming"
81	collabor skill	+1.125	"collaboration skills", "collaborative skills", "collaboration skill", "collaborative skill", "collaborating skills"
82	state	+1.123	"state", "stated", "states", "stating", "stateful"
83	effect	+1.118	"effectively", "effective", "effectiveness", "effect", "effects"
84	continu posit	+1.117	"continue her positive", "continue the positive", "continue to be a positive", "continually had a positive", "continue his positive"
85	confid speak	+1.116	"confidence to speak", "confidence when speaking", "confidence in speaking", "confident to speak", "confident when speaking"
86	honest	+1.113	"honest", "honestly"
87	planner	+1.113	"planner", "planners"
88	work home	+1.107	"working from home", "work from home", "worked from home", "working at home", "works from home"
89	note	+1.102	"notes", "noted", "note", "noting"
90	glad	+1.099	"glad", "gladly", "gladfully"
91	skill develop	+1.095	"skill development", "skills development", "skills developed", "skills to develop", "skill to develop"
92	technolog skill	+1.094	"technological skills", "technology skills", "technologies and skills", "technology skill", "technological skill"

93	encourag think	+1.088	"encouraged to think", "encourage him to think", "encourage you to think", "encourage her to think", "encourage him to do so i think"
94	believ	+1.086	"believe", "believes", "believed", "believing", "believer"
95	develop quot	+1.080	"development quot", "developer quot", "develop a quot", "develop her quot", "develop quot"
96	develop section	+1.080	"development section", "develop our section", "develop section", "developement section", "development sections"
97	improv busi	+1.078	"improve business", "improving the business", "improve the business", "improve her business", "improve both the business"
98	demonstr strong	+1.077	"demonstrated strong", "demonstrated a strong", "demonstrates strong", "demonstrated very strong", "demonstrated a very strong"
99	strong	+1.073	"strong", "strongly", "stronge"
100	lifelong learn	+1.071	"lifelong learning"

Table A.6: The top 100 terms predicting an in-person student (2019) from the supervisors' recommendation comments, corresponding to Table 3.8 in the main text.

Rank	Term	Weight	Original Text
1	written	-2.795	"written"
2	interperson	-2.671	"interpersonal", "interpersonally", "interperson"
3	quot	-2.592	"quot", "quotes", "quote", "quoting"
4	public speak	-2.246	"public speaking"
5	present	-2.118	"presentation", "presented", "present", "presentations", "presenting"
6	entrepreneuri	-2.064	"entrepreneurial", "entrepreneurialism", "entrepreneurially"
7	interperson communic	-1.931	"interpersonal communication", "interpersonal and communication", "interpersonal communications", "interpersonal and communications"
8	depend	-1.926	"dependable", "dependability", "depend", "depending", "dependencies"
9	proactiv	-1.868	"proactive", "proactively", "proactiveness", "proactivity", "proactivelys"
10	oral	-1.830	"oral", "orally", "orall"

11	assert	-1.827	“assertive”, “assertiveness”, “assert”, “asserting”, “assertively”
12	sometim	-1.803	“sometimes”, “sometime”
13	ensur	-1.788	“ensure”, “ensuring”, “ensured”, “ensures”
14	quiet	-1.749	“quiet”, “quietly”, “quietness”
15	problem solv	-1.725	“problem solving”, “problem solve”, “problems to solve”, “problem solved”, “problem to solve”
16	teamwork	-1.723	“teamwork”, “teamworker”, “teamworking”, “teamworks”
17	public	-1.715	“public”, “publication”, “publications”, “publicly”, “publically”
18	oral communic	-1.701	“oral communication”, “oral communications”, “oral communicator”, “oral and communication”, “oral the communication”
19	issu	-1.683	“issues”, “issue”, “issuing”, “issued”
20	speak	-1.661	“speaking”, “speak”, “speaks”, “speakings”
21	punctual	-1.650	“punctual”, “punctuality”, “punctually”, “punctuallity”
22	request	-1.615	“requests”, “requested”, “request”, “requesting”, “requester”
23	interact	-1.562	“interactions”, “interact”, “interaction”, “interacted”, “interacting”
24	audienc	-1.555	“audience”, “audiences”
25	enthusiasm	-1.532	“enthusiasm”, “enthusiasms”
26	work abil	-1.504	“work and ability”, “working ability”, “work ability”, “work and his ability”, “work and her ability”
27	independ	-1.504	“independently”, “independent”, “independence”, “independantly”, “independant”
28	order	-1.494	“order”, “orders”, “ordering”, “orderly”, “ordered”
29	attent	-1.476	“attention”, “attentive”, “attentively”, “attentiveness”, “attentions”
30	peopl	-1.381	“people”, “peoples”
31	resolv	-1.364	“resolve”, “resolved”, “resolving”, “resolves”, “resolver”
32	littl	-1.338	“little”
33	solv skill	-1.337	“solving skills”, “solving skill”, “solveing skills”, “solving these are skills”
34	qualiti	-1.324	“quality”, “qualities”
35	document	-1.314	“documentation”, “documents”, “document”, “documented”, “documenting”

36	approach	-1.295	“approach”, “approaches”, “approached”, “approachable”, “approaching”
37	discuss person	-1.295	“discussed in person”, “discuss personal”, “discuss with other persons”, “discussed personally”, “discussed these in person”
38	offic	-1.287	“office”, “offices”, “officer”, “officers”, “offics”
39	career softwar	-1.278	“career in software”, “career as a software”, “career in the software”, “career as software”
40	matur	-1.277	“maturity”, “mature”, “matures”, “matured”, “maturely”
41	languag	-1.269	“language”, “languages”
42	problemsolv	-1.268	“problemsolving”, “problemsolve”, “problemsolver”, “problemsolved”
43	construct critic	-1.243	“constructive criticism”, “constructive criticisms”, “construction criticism”
44	br xd	-1.239	“br xd”
45	given	-1.239	“given”
46	manner	-1.226	“manner”, “mannered”, “manners”, “mannerism”, “mannerly”
47	lead	-1.221	“leadership”, “lead”, “leading”, “leads”, “leaded”
48	br	-1.214	“br”
49	enhanc	-1.190	“enhance”, “enhanced”, “enhancing”, “enhancements”, “enhancement”
50	creativ	-1.189	“creative”, “creativity”, “creatively”, “creativeness”, “creativities”
51	someth	-1.187	“something”
52	improv written	-1.186	“improve his written”, “improving written”, “improve written”, “improving his written”, “improve her written”
53	watch	-1.184	“watch”, “watching”, “watched”, “watches”, “watchful”
54	easili	-1.170	“easily”
55	need work	-1.168	“needs to work”, “need to work”, “needed to work”, “needs more work”, “needs work”
56	challeng face	-1.167	“challenges faced”, “challenges he faced”, “challenges she faced”, “challenges we faced”, “challenges facing”
57	fulli	-1.165	“fully”
58	error	-1.160	“errors”, “error”
59	test	-1.152	“testing”, “test”, “tests”, “tested”, “testings”
60	pursu	-1.152	“pursue”, “pursuing”, “pursues”, “pursued”, “pursueing”
61	judgement	-1.150	“judgement”, “judgements”

62	encourag continu	-1.148	“encouraged to continue”, “encourage her to continue”, “encourage him to continue”, “encourage you to continue”, “encourage to continue”
63	client	-1.143	“client”, “clients”
64	attempt	-1.136	“attempt”, “attempting”, “attempts”, “attempted”
65	xd ask	-1.130	“xd ask”, “xd asking”, “xd asked”, “xd asks”, “xd i asked”
66	listen	-1.130	“listening”, “listen”, “listens”, “listener”, “listened”
67	work tri	-1.130	“work and try”, “work try”, “work to try”, “working on trying”, “work and tried”
68	intuit	-1.129	“intuition”, “intuitive”, “intuit”, “intuitively”, “intuitions”
69	activ listen	-1.126	“active listening”, “active listener”, “actively listening”, “actively listens”, “actively listened”
70	consist	-1.122	“consistently”, “consistent”, “consistency”, “consistantly”, “consisted”
71	verbal	-1.116	“verbal”, “verbally”, “verbalize”, “verbalized”, “verbalizing”
72	object	-1.113	“objectives”, “objective”, “object”, “objects”, “objectively”
73	successful	-1.111	“successfull”
74	care	-1.096	“care”, “carefully”, “careful”, “cares”, “caring”
75	communic skill	-1.093	“communication skills”, “communications skills”, “communication skill”, “communicative skills”, “communications skill”
76	grow profession	-1.084	“grow professionally”, “grow as a professional”, “grow her professional”, “grow his professional”, “growing professionally”
77	consult	-1.078	“consulting”, “consultant”, “consultants”, “consult”, “consultation”
78	opportun demonstr	-1.078	“opportunities to demonstrate”, “opportunity to demonstrate”, “opportunities that she has demonstrated”, “opportunity for you to demonstrate”, “opportunities and demonstrated”
79	social	-1.072	“social”, “socially”, “socialize”, “socializing”, “socials”
80	xd	-1.072	“xd”
81	becam	-1.067	“became”
82	quick	-1.065	“quickly”, “quick”, “quickness”
83	mark	-1.063	“marking”, “mark”, “marked”, “marks”, “markedly”
84	persu	-1.062	“persue”, “persuing”, “persued”, “persues”
85	ergonom	-1.061	“ergonomics”, “ergonomic”, “ergonomically”

86	opinion	-1.055	“opinions”, “opinion”, “opinionated”
87	control	-1.048	“control”, “controls”, “controller”, “controlled”, “controllers”
88	affect	-1.045	“affect”, “affected”, “affecting”, “affects”, “affection”
89	opportunit	-1.040	“opportunites”
90	situat	-1.039	“situations”, “situation”, “situational”, “situated”, “situate”
91	complet	-1.039	“completed”, “complete”, “completing”, “completion”, “completes”
92	undertak	-1.033	“undertake”, “undertaking”, “undertakes”, “undertakings”
93	concis	-1.023	“concise”, “concisely”, “conciseness”, “concision”
94	xd br	-1.022	“xd br”, “xd same as above br”, “xd and br”, “xd so br”, “xd these are br”
95	louder	-1.021	“louder”
96	agre	-1.021	“agreed”, “agree”, “agrees”, “agreeing”
97	present skill	-1.017	“presentation skills”, “presentation skill”, “presentations skills”, “presenting skills”, “presentational skills”
98	vs	-1.016	“vs”
99	op	-1.007	“op”, “ops”
100	qa	-1.007	“qa”, “qaing”