Understanding employer expectations of students in different disciplines: A novel graph methodology

VERONICA SALM
LUKASZ GOLAB[1]
University of Waterloo, Waterloo, Canada

Co-operative education (co-op) programs enable students to gain real-world experience by alternating work and study terms. Understanding employers' expectations of students in different disciplines is therefore critical for success. To do this, the relationships between co-op job postings, required skills, and academic backgrounds can be modeled using graphs. This study proposes a novel graph methodology for analyzing employer expectations of co-op students by grouping academic backgrounds and either co-op job titles or required skills into graph clusters. A case study is used to demonstrate the effectiveness of this methodology. Over 13,000 co-op job postings are used to uncover key differences in employer expectations for students in each program, including the importance of communication for arts students who are expected to fill interdisciplinary roles and the prevalence of data science and highly specific technical skills in many academic programs.

Keywords: co-operative education, co-operative job postings, graphs, community detection, employer expectations

---

Co-operative education (co-op) is one of the nine types of work-integrated learning defined by Co-operative and Work-Integrated Learning Canada (CEWIL Canada). In co-operative education programs, students alternate between study terms and work terms. These programs enable students to gain valuable work experience while strengthening ties between institutions and employers (Eames & Coll, 2010; Thiel & Hartley, 1997).

By examining the co-operative roles advertised to students in different programs and the skills required for those roles, it is possible to assess the gap between employer expectations and skills taught to students. This understanding can help students better prepare for future employment in their field. Similarly, institutions can ensure that skills expected by employers are taught in relevant academic programs. With this motivation in mind, this study aims to answer the following questions:

1. What skills, abilities, and attributes do co-op employers expect from students with specific academic backgrounds?
2. What co-op placements do employers expect to be most appropriate for students with specific academic backgrounds?

When exploring relationships between attributes, such as skills and student backgrounds, it is helpful to model them as a network, or graph. Many real-world datasets can be modeled as graphs, including social networks, protein interactions, and job markets (Fortunato, 2010). For example, users in a social network might be modeled as nodes in a graph, where two or more users are connected by edges if they share a relationship (e.g., if they are friends in the social network). Analyzing the structure of these graphs can provide useful insights and reveal relationships. Though previous work has leveraged graph structures to understand the relationships between jobs and the skills they require (de

---

[1] Corresponding author: Lukasz Golab, lgolab@uwaterloo.ca

Groot et al., 2021; Jia et al., 2018), there is a notable gap in the use of graphs to understand what co-op employers expect of students with different backgrounds.

The purpose of this article is to propose a graph methodology for analyzing job postings to answer the above research questions. The methodology is described in the following sections and demonstrated using a case study analyzing over 13,000 unique co-op jobs posted to students in a large North American university. This secondary data analysis explores the types of jobs available to co-op students and the expectations of employers of students with different backgrounds, uncovering skills and roles that students should pursue to be successful in their chosen disciplines.

Permission for this secondary data analysis was granted by the university's office of research ethics (application number 43970) on January 11, 2022.

BACKGROUND

The following sections summarize two areas of research closely related to this work: graph-based studies of job markets and analyses of required skills in job postings.

Using Graphs to Model and Detect Communities in Job Markets

By identifying groups of similar nodes in a graph (e.g., users who are close friends in a social network), researchers can understand how communities differ and interact. A large body of previous work has developed techniques for community detection (Fortunato, 2010). The Louvain method (Blondel et al., 2008) is a popular approach for community detection which aims to maximize connections within communities while minimizing connections between. For example, Choi et al. (2021)use Louvain to analyze communities in a graph of employees to understand how they are connected based on the similar tasks they perform. Samek et al. (2021) 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.

Community detection has also been used to analyze competition in the co-op workforce. Jiang and Golab (2016) perform Louvain community detection to characterize competition for co-op interviews between students from different academic programs. Similarly, Toulis and Golab (2017) apply Louvain community detection to identify the most important co-op employee and employer communities. However, these studies focus on characterizing competition using interview data rather than identifying important skills expected by employers.

Other graph analyses have been used to study job markets, including graph representations for job recommendation (Giabelli et al., 2021; Shalaby et al., 2017; Upadhyay et al., 2021) and to understand online job markets (Collarana et al., 2018). However, these analyses typically require access to additional datasets with extensive user or employer information that is not available in this study.

Using Job Postings to Understand Required Skills

Previous work has analyzed job postings to understand the types of skills required by employers. Lyu and Liu (2021)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. Sodhi and Son (2010)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 and Mangeol (2020)analyze over nine million job postings from four US states using frequency to understand how skill demand varies for different states and occupations. Although these studies address interesting problems, they do not leverage the relationships between skills and other job posting attributes to answer the research questions posed by this study.

Other analyses have used graphs to understand the relationships between jobs and skills. Using a graph of skills and occupations, de Groot et al. (2021) find similar occupations and the most relevant skills per occupation group. However, this analysis does not use community detection, does not consider an employee's academic or work history background, and does not reveal the skills that most strongly differentiate each occupation, as it is possible for two occupation groups to have the same top skill. Jia et al. (2018) use a neural network to extract skills from artificial intelligence positions and then find in-demand skills by visualizing a graph of skills and occupations. However, like the previous study, this analysis compares skill requirements for different job types and does not consider employee backgrounds.

DATA OVERVIEW

A case study analyzing co-op job postings at a large North American institution is used to demonstrate the proposed graph methodology. The case study dataset consists of 13,377 unique filled co-op job descriptions. All jobs were posted over one year between the winter and fall terms in 2021.

Co-op Job Postings

Each job posting includes a unique identifier, the academic term during which it was filled (e.g., Winter of 2021), and the following text fields:

- Job Title: the name of the role being advertised, such as 'Software Developer'. There are 5,836 unique job titles in this dataset.
- Job Summary: typically includes information about the co-op employer, job, and why a prospective employee should apply.
- Job Responsibilities: main tasks and responsibilities of the job.
- Required Skills: abilities, skills, and attributes required of prospective employees.

Co-op Job Posting Clusters

As a unique feature of this dataset, co-op employers must select advertising clusters to target the students whose background they believe would be best suited to the given position. These clusters reflect the academic programs offered by this institution but enable the employer to advertise to groups of students without having to understand institution-specific programs. There are two types of clusters:

- Academic clusters target groupings of academic programs. These clusters allow the co-op employer to specify the type of student they believe would best fit the role. Each academic cluster is related to at least one of the six faculties: ENG (Engineering), MATH (Mathematics), ARTS (Arts), SCI (Science), ENV (Environment), or AHS (Health). The Mathematics faculty at this institution is unique because it includes programs typically included in other faculties (e.g., computer science).
- Thematic clusters reflect the function of the job (such as 'Finance and Investment') and the type of work that a co-op student 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 between academic and thematic clusters. Most employers specify at least one academic advertising cluster, as demonstrated by 97.8% of 2021 jobs (13,088) using at least one to advertise to students.

A list of all academic clusters in the dataset is provided in Table 1. The clusters are grouped by faculty (including a separate section for the three academic clusters that correspond to multiple faculties). Also included are 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%).

METHODS

This case study analyzes all 13,377 unique co-op jobs as described in the previous section. The following sections describe the preprocessing of each job posting, the conversion of the job postings into two graphs, and how communities of important job titles and skills are detected for different student disciplines.

Preprocessing

Required skills

Job postings indicate how co-op employers advertise themselves and what they expect from prospective student employees. However, these data present challenges. Many fields contain unstructured text written freely by the employer. This natural language is prone to exceptions, mistakes, and idiosyncrasies. Words with similar meanings may not have the same representation in the text. 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. The challenge lies in making a computer understand that these dissimilar phrases have the same meaning through a process called *parsing*.

When parsing unstructured text fields, standard natural language processing techniques are a good first step. The text can be converted to lowercase, punctuation removed, and special characters normalized. Text can then be divided into individual words and words with similar meanings converted to common representations called *tokens*. The job description parser developed by Chopra and Golab (2018) is designed to extract required skills from job postings using a combination of basic methods and additional domain-specific rules (for example, converting both 'Microsoft Office' and 'MS Office' to the token *msoffice*). This makes Chopra and Golab's parser appropriate for parsing the required skills sections of job postings in this dataset.

Applying the parser to the required skills section of each job posting produces a vocabulary of 6,313 unique tokens.

Co-op job posting clusters

In this dataset, all advertising clusters for a given job are stored as a single string, separated by semicolons. Therefore, the first step is to split the string on the semicolon to produce a list of clusters for each job. Finally, thematic clusters are filtered out (by removing any clusters that begin with 'Theme') to produce a list containing only the academic clusters for each job. After processing, there are 45 unique academic clusters.

TABLE 1: A list of 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	2,426	18.1
ARTS Economics	1,222	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	1,308	9.8
ARTS/MATH/SCI Chartered Professional Accounting	883	6.6
ARTS/SCI Psychology	441	3.3
ENG Software Engineering	5,117	38.3
ENG Electrical and Computer Engineering	4,649	34.8
ENG Mechanical and Mechatronics Engineering	2,679	20.0
ENG Systems Design and Biomedical Engineering	2,617	19.6
ENG Civil, Environmental and Geological Engineering	1,541	11.5
ENG Management Sciences	1,522	11.4
ENG Nanotechnology Engineering	1,176	8.8
ENG Chemical Engineering	1,095	8.2
ENG Architectural Engineering	595	4.4
ENG Architecture	508	3.8
ENV Business, Enterprise and Development	1,165	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	5,343	39.9
MATH Business	2,402	18.0
MATH Computing and Financial Management	1,768	13.2
MATH Statistics and Actuarial Science	1,623	12.1
MATH Applied Mathematics	1,421	10.6
MATH Pure Mathematics	940	7.0
MATH Combinatorics and Optimization	918	6.9
MATH Teaching	368	2.8
SCI Business	1,532	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

Job titles

No preprocessing is performed on job title text because the number of unique job titles is relatively small and preprocessing might remove important information (for example, abbreviations such as 'QA' or 'CS' might be incorrectly removed). Because there are only 5,836 unique job titles for 13,377 unique jobs, multiple unique jobs may have the same title (such as Software Developer). For jobs with duplicate titles, the lists of co-op job posting clusters for each job are concatenated before performing further analyses.

Community Detection in Bipartite Graphs

This analysis aims to determine the required skills and co-op job roles that are most important for students with specific academic backgrounds. When constructing graphs of these data, two sets of nodes naturally appear: the academic background of a student (e.g., computer science) and either job roles or skills. For a given job, a skill or job role is then linked to each academic background targeted by the employer. For example, if a job requires the skill 'Python' and advertises to a computer science student, an edge in the graph would be constructed to link the Python node to the computer science node. In the resulting graph, there are connections only between the two sets of nodes, with no direct links between skills/titles or between advertising clusters. This structure is called a *bipartite graph*.

Unfortunately, it is difficult to group two sets of nodes simultaneously. Existing solutions to this problem convert the bipartite graph to use only one type of node (Zweig & Kaufmann, 2011), to enable the use of powerful tools designed for simpler graphs. For example, here, a graph of skill nodes could be created where two skills are linked if they appear together in a job posting targeting the same academic cluster. However, doing this loses information stored in the bipartite version(Latapy et al., 2008). For example, with this study's bipartite graphs, it is possible to detect which academic clusters link specific skills and the strength of each individual connection. In the simplified version, the academic advertising clusters that caused the overlap are difficult to extract. Bipartite graphs are needed to answer this study's 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.

The problem of grouping two sets of nodes simultaneously is known as *co-clustering* (Rege et al., 2008). Co-clustering is appropriate for real-world problems that require capturing the relationships between two types of objects. Previous research has demonstrated the success of co-clustering either documents (Bisson & Hussain, 2008)or sentences (Cai et al., 2014)with words, producing better results than grouping words, sentences, or documents alone. Co-clustering has been used to understand the qualities users care about most strongly in video games (Raison et al., 2012), to improve theme-based summarization(Cai et al., 2014), to find teams of experts whose combined skills fulfill a given task(Farhadi et al., 2012), and for web log analysis(Xu et al., 2010). Co-clustering reveals the underlying relationships between the two sets of nodes more effectively than relying on one dimension alone.

Community detection provides a method of co-clustering bipartite graphs. The Leiden algorithm for community detection, introduced as an improvement to Louvain, uncovers communities more efficiently. However, most algorithms for community detection, including both Louvain (Blondel et al., 2008) and Leiden (Traag et al., 2019), are designed for one mode networks(Tackx et al., 2018). Some previous work has designed or adapted specialized methods for bipartite community detection(Rossetti et al., 2019; Tackx et al., 2018; Taguchi et al., 2020), but implementations are not readily available. Fortunately, the Python package Leidenalg supports modified community detection for bipartite graphs. It defines three resolution parameters: one for within each set of node and one for the edges between the nodes. By emphasizing the edges between the classes and de-emphasizing edges within the sets, it is possible to perform community detection effectively on bipartite graphs. Finally, because community detection forces the assignment of each node to a unique cluster, both sets of node are clustered together. This enables finding the skills or roles that are more linked to each academic cluster than to any other cluster.

Required skill graph

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 containing the same skill and cluster. Therefore, skills and clusters that often appear together will be more strongly linked. The result is a bipartite graph mapping academic advertising clusters to required skills. A small example skill graph (Figure 1) is constructed using only two job postings (Table 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.

TABLE 2: Job postings used to generate the example graph (Figure 1), including the title, original required skills section, tokens produced after parsing, and co-op job posting clusters.

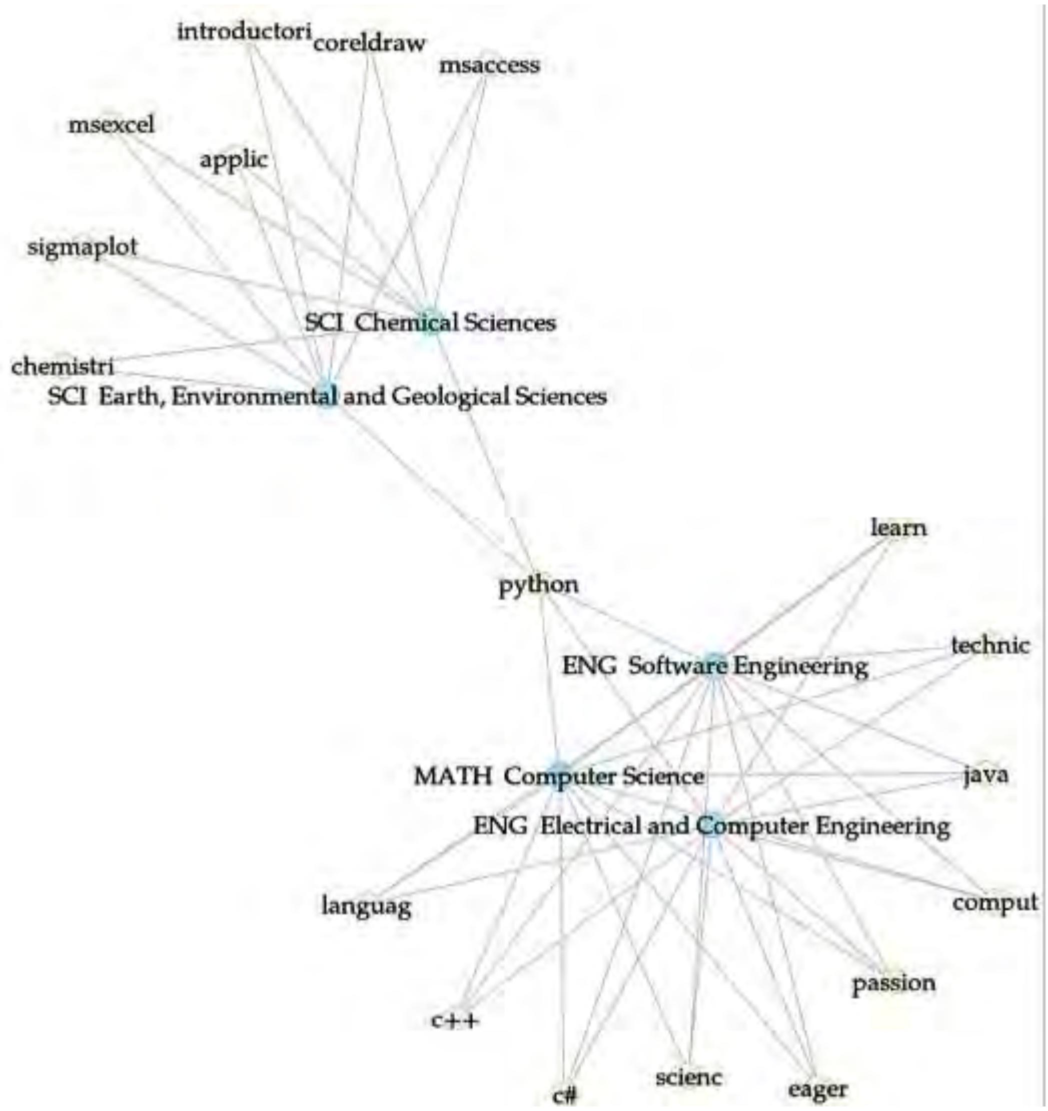
Job Title	Original Required Skills	Parsed Tokens	Co-op Job Posting Clusters
Full Stack Software Engineering Intern	- Must be taking a Computer Science or related technical degree. - Experience programming in at least one of these languages: Java, C#, Python, C++ - Eager to learn and passionate.	comput 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

Job title graph

In this graph, each node is either a job title or an academic cluster. An edge is added between an academic cluster node and job title node if at least one job with that title specifies the given academic cluster. Each time an academic cluster is found in the same job description as a given title, the edge weight is increased by one. The result is a bipartite graph mapping academic advertising clusters to job titles.

One might wonder why a bipartite graph was not also constructed between job titles and required skills. First, this graph would not provide information to answer this study's research questions, because it does not determine which skills are most important for students with specific academic backgrounds. 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 employers may use different titles for similar jobs, resulting in noisy data.

FIGURE 1: A small example bipartite graph, generated using the job postings in Table 1. 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.



Community Detection

For both graphs, community detection is performed using the Leiden algorithm using the Leidenalg Python package after tuning as specified in the documentation.

RESULTS

This analysis produces two graphs:

1. Skill Graph: the graph between academic clusters and skills has 6,359 nodes and 76,901 edges.
2. Job Title Graph: the graph between academic clusters and co-op job titles has 5,881 nodes and 26,993 edges.

Performing community detection resulted in 10 communities for the skill graph and 12 communities for the job title graph. Each community was given a descriptive name based on the included program groupings. The communities are summarized in Table 3 (Skill graph communities) and Table 4 (Job title communities). The communities are summarized in the sections below, and the three largest skill and job title communities can be explored in more detail in Appendix A Tables 1-3 and Appendix B Tables 1-3.

Skill Graph Communities

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

- Of these skills, 67 were only listed in jobs that never advertised to any academic cluster (that is, they used thematic clusters only). As a result, they did not form edges to any academic clusters and could not become part of a larger community.
- The remaining 16 nodes corresponded to skill terms such as *urban*, *spatial*, *configur* (configure), and *climat* (climate). These skills appeared infrequently in the dataset and were evenly linked to a wide variety of academic clusters, so they were not assigned to any particular cluster. It is likely that this is the result of a co-op employer targeting all 45 academic clusters regardless of whether they are appropriate, as occurred in 10 unique jobs in the dataset (0.07%).

TABLE 3: A summary of the 10 skill graph communities, including identification number, manually assigned name, and number and percentage of nodes and edges in the community.

Community ID	Name	# of Nodes	% of Nodes	# of Edges	% of Edges
1	Arts and Humanities	949	14.92%	6,584	8.56%
2	Computer Science, Software, and Hardware Engineering	1,982	31.17%	6,682	8.69%
3	Business	591	9.29%	2,328	3.03%
4	Environment and Geological Sciences	711	11.18%	2,658	3.46%
5	Math, Statistics, and Optimization	600	9.44%	2,189	2.85%
6	Architecture, Pharmacy, and Teaching	453	7.12%	1,288	1.67%
7	General Sciences	453	7.12%	1,192	1.55%
8	Finance and Accounting	245	3.85%	456	0.59%
9	Health	236	3.71%	403	0.52%
10	Management Sciences	56	0.88%	55	0.07%

Community 1: Arts and Humanities includes eight of the 10 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.

Community 2 includes Computer Science, Software, and Hardware Engineering students. The skills most specific to this community 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).

Community 3: Business 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 mathematical and statistical concepts such as matrices and bias. When examining this community, the presence of the term *diplomat* was initially unclear. However, manual inspection found that these students are expected to be diplomatic team players who carefully consider the interests of stakeholders.

Community 4 includes students from Environment and Geological Sciences. In addition to using environmental terminology more frequently (e.g., water, geology, conservation) when advertising to students from this community, employers also expect them to be familiar with AutoCAD, a program used to create 2D and 3D drawings.

Community 5 includes students in Math, Statistics, and Optimization. Employers expect students in this community to be familiar with the Society of Actuaries and programming languages such as Typescript, HTML5, Racket, and Ruby.

Community 6: Architecture, Pharmacy, and Teaching students have a common requirement for machine learning and AI skills: algorithms, TensorFlow, PyTorch, along with other technological skills such as Rhinoceros 3D, Javascript, and concepts like object-oriented programming, and application programming interfaces.

Community 7: General Sciences includes students in biology, chemistry, physics, and nanotechnology. They 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.

Community 8 includes Finance and Accounting students. In addition to tools such as Visual Basic for Applications (VBA) and concepts such as auditing and risk management, finance and accounting students are uniquely expected to be lifelong learners and focus on extracurriculars.

Community 9: Health includes public health and kinesiology students, who are expected to understand anatomy and kinesiology. They must also be dependable and be prepared for multi-disciplinary work.

Community 10: Management Sciences students 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 and Robotic Process Automation (RPA).

Job Title Communities

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

- Of these nodes, 151 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 a larger community.

- The remaining two 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 were not assigned to any particular cluster.

Community 1 includes Computer Science and Software Engineering students. Employers expect these students to fill jobs involving software, mobile, firmware, full stack, backend, and automation development.

Community 2 involves Math, Statistics, Optimization, and Teaching. In general, these students are suited for roles focusing on mathematics and finance. However, employers also expect to hire these students for teaching assistant positions, data science, and software development.

TABLE 4: A summary of the job title graph communities, including identification number, manually assigned name, and number and percentage of nodes and edges in the community.

Community ID	Name	# of Nodes	% of Nodes	# of Edges	% of Edges
1	Computer Science and Software Engineering	1,286	21.87%	3,160	11.71%
2	Math, Statistics, Optimization, and Teaching	566	9.62%	1,795	6.65%
3	Arts and Humanities	466	7.92%	1,730	6.41%
4	Business	498	8.47%	1,317	4.88%
5	Mechanical, Biomedical, and Nanotechnology Engineering	664	11.29%	1,243	4.60%
6	Environment and Geological Sciences	409	6.95%	1,212	4.49%
7	Finance, Accounting, and Economics	455	7.74%	873	3.23%
8	Architecture and Environmental Engineering	448	7.61%	711	2.63%
9	Biology, Chemistry, and Pharmacy	442	7.52%	740	2.74%
10	Health, Recreation, and Kinesiology	241	4.10%	393	1.46%
11	Management Sciences	148	2.52%	147	0.54%
12	Physics	105	1.79%	104	0.39%

Community 3: Students in Arts and Humanities 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.

Community 4: Business students are expected to fill roles associated with operations, customer success, sales, marketing, and human resources.

Community 5 includes students in Mechanical, Biomedical, and Nanotechnology Engineering. Although employers expect these engineering students to fill engineering roles related to their programs, they also might fill roles involving software development, quality assurance, and design. Job titles such as 'iOS Developer,' 'Game Programmer,' and 'Junior Developer' demonstrate that programming skills are important for all engineers (not only those in programs explicitly related to software or hardware).

Community 6 is Environment and Geological Sciences**.** This is another case where employers appear to expect technological skills from students not explicitly enrolled in computer science or software engineering. For example, these students are targeted for compiler software engineering and big data development roles in addition to geospatial intelligence and geomatics.

Community 7: Finance, Accounting, and Economics students are most targeted for business, finance, and accounting roles. There is a notable absence of technology related roles in this community, perhaps indicating that finance does not rely as heavily on technological skills.

Community 8: Architecture and Environmental Engineering students are generally expected to fill architectural, project management, construction, and online learning assistant roles. Although technology related roles do not appear in the top 15, there are roles such as 'Data Software Engineering' and 'Software Developer' within the top 30.

Community 9: Biology, Chemistry, and Pharmacy students are targeted for varied roles including 'Production Worker,' 'Math Tutor,' and 'Lab Assistant.' They are often considered a good fit for research and educational positions.

Community 10 includes Health, Recreation, and Kinesiology. Unsurprisingly, most of the roles that target these students are related to health care, animal care, and community-oriented roles. Many positions also involve customer service work that centers around interfacing with people.

Community 11 is Management Sciences. This community includes only one academic cluster (ENG - Management Sciences) and includes roles related to logistics, reliability, and maintenance. However, there are also numerous information technology related roles, including software and web application development.

Community 12 is Physics. Jobs targeting physics students form their own community, where roles often involve concepts such as medical physics and 2D image sensing. However, there are also some other trends: research and development, engineering, software, and machine learning.

DISCUSSION

When analyzing the skill and job title communities, several key insights emerge:

- Many co-op students are expected to possess highly specific technical skills.
- Data science and machine learning are important for a variety of disciplines.
- Arts students are expected to possess more people skills (especially communication) and fill interdisciplinary co-op roles compared to students in traditionally technology-oriented fields.

These insights are expanded in the following sections.

Required Skill Expectations

Students in traditionally technology related fields (Community 2) are expected by co-op employers to possess significant technical knowledge. Many other student groups also require highly specific technical knowledge, including the Management Sciences and General Sciences groupings. However, it is also common for co-op 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 analyzing skills in the energy sector, Lyu and Liu (2021) found that soft skills were more frequently required than hard skills in the past decade, although general computer skills was one of the two most valuable skills (along with products and marketing). This is one example where the approach taken by this work provides more insight into the most strongly required skills for specific occupations than this simple frequency analysis. Lyu and Liu were not able to determine whether soft skills were more important for energy sector jobs than for other types of jobs, or which skills were most closely linked to these occupations. In contrast, this community detection analysis determined that soft skills (such as communication) were most valued by co-op students in Arts and Humanities (Community 1).

Data science, artificial intelligence, and machine learning skills appear to be in demand for co-op students in a diverse range of fields: Computer Science, Software, and Hardware Engineering as well as Architecture, Pharmacy, and Teaching. After manually inspecting co-op jobs from the Architecture, Pharmacy, and Teaching community, which seemed to be an unusual grouping, 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, and machine learning tools. Previous work by Jia et al. (2018) found that Python and Linux are the most necessary key skills for all jobs in AI. This aligns with Community 6 in the skill graph, where Linux was one of the most strongly linked skills for students in Architecture, Pharmacy, and Teaching. Notably, although Python was not most strongly linked to this group specifically, other skills such as PyTorch and Tensorflow (Python add-ons commonly used for machine learning) were present, giving more information about which specific flavors of Python were the most important. Jia et al. found that PyTorch and Tensorflow were required for deep learning positions in particular, which may indicate that deep learning has become more important for AI jobs since 2018.

Job Role Expectations

In general, some program groupings are expected by co-op 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 technology roles targeted to co-op students not in traditionally technology related programs. For example, students in biology, chemistry, and pharmacy are linked to 'Sensors Research Samurai', in environment and geological sciences are linked to 'AI Developer', and students in business to 'Digital Communications Assistant'. Students in mathematics and statistics are expected to fill roles in data science more strongly than any other group. 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, co-op employers might simply expect finance students to have a specific financial skill set, where technology skills are secondary.

The job title communities reveal the roles most closely associated with different student backgrounds quickly and automatically. In contrast, the work of Brüning and Mangeol (2020), relies on manual interpretation of frequency information to analyze the jobs expected of students graduating from sociology programs. While this analysis provides useful insights for sociology majors in four US states, it is limited in scope, does not compare the roles to those needed for students with other backgrounds, and would be manually intensive to repeat for other disciplines. The graph-based community detection methodology used in this analysis provides similar insights without the need for extensive manual interpretation of the results, and different disciplines can be compared at a glance.

Comparison to Work-Integrated-Learning Literature

In the WIL literature, the closest prior work in terms of the methodology is by Jiang and Golab (2016) and Toulis and Golab (2017). These two studies analyzed communities in graphs in which edges represented competition for co-op jobs, i.e., students interviewing for the same placement or co-op employers interviewing the same students. In contrast, the graphs used in this analysis are based on skill and job title relationships for the purpose of understanding co-op employers' expectations of students in different academic programs.

In the context of skill or job role expectations, Chopra and Golab (2018) proposed a clustering method to understand the different types of co-op roles available to students. While their clustering method could be applied separately to co-op postings advertised to different disciplines, the resulting clusters may overlap. In contrast, by including academic disciplines directly in the graph structure, the methodology presented in this paper finds representative skills and job titles corresponding to employer expectations in different disciplines. Moreover, studies have been published on important skills from the point of view of WIL students and employers (Basson et al., 2023; Lisa et al., 2019), but these studies also do not focus on employer expectations of WIL students with different academic backgrounds.

Finally, recent work addressed issues in the context of matching co-op students with employers, by analyzing how students and employers rank their options (Chopra & Golab, 2022). While employer expectations of students from different academic disciplines were not discussed, an interesting direction for future work informed by this paper is to analyze how employers who advertise a co-op placement to a specific academic program rank students from other programs.

PRACTICAL IMPLICATIONS

To ensure that the best candidate is matched to each co-op job, it is important to understand which skills and job roles are most relevant and in demand. Although finding popular skills and job roles might help some students better navigate the co-op job market, these frequency-based analyses do not explore employer expectations of students with specific backgrounds. The graph-based community detection methodology presented in this work provides new insights that are valuable to institutions, employers, students, and researchers. Institutions are under pressure to keep curricula up to date with employer expectations, especially in applied programs (such as engineering). This analysis can uncover insights about what roles employers expect from students with each background (for example, that Architecture, Pharmacy, and Teaching students were linked by their eligibility for AI-related roles). Institutions can use this information to 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. Similarly, current students can learn where they might potentially have an advantage over students in other occupations and help them recognize areas for development in their skills. This is especially important for students in programs where the career path is not clearly defined, such as sociology (Brüning & Mangeol, 2020), to help them transition into the labour market. Prospective students might also 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.

Unlike previous work analyzing required skills in job postings (de Groot et al., 2021), this analysis did not rely on predefined skill or occupation taxonomies. Instead, groupings of skills, job roles, and even groups of academic programs were extracted from the data, providing unique insights into how co-op employers view different academic programs and what they expect from these students. This work could be used to assess the quality of existing taxonomies by comparing their roles and skills to see if they align with the communities constructed in this analysis using employer expectations.

Additionally, follow-up research could analyze the experiences of co-op students with respect to the skills required in their internship experience. Finally, researchers and practitioners can leverage this graph methodology to co-cluster interesting attributes of job descriptions in other datasets.

CONCLUSION

This study presented a graph methodology for exploring relationships between pairs of attributes in job description datasets. By constructing a bipartite graph and grouping similar nodes into communities, this methodology reveals the most strongly related attributes. A case study consisting of over 13,000 co-operative job postings used this methodology to investigate the relationships between academic advertising clusters, used by co-operative employers to target students with specific backgrounds, and job description attributes: required skills or job roles. The resulting communities revealed differences in employers' expectations for students with different academic backgrounds, with specific technical skills required for many programs. These findings provide direction for institutions to better prepare students for future success and demonstrate the effectiveness of the proposed methodology, which researchers and practitioners can use to investigate their own job posting datasets.

REFERENCES

Basson, M., Du Plessis, T., & Brikn, R. (2023). Visual representation of the mismatch between industry skills demand and higher education skills supply. *International Journal of Work Integrated Learning, 24(*1), 117-139.

Bisson, G., & Hussain, F. (2008). Chi-sim: A new similarity measure for the co-clustering task. *Seventh International Conference on Machine Learning and Applications*,. 211–217.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *10*, P10008.

Brüning, N., & Mangeol, P. (2020). What skills do employers seek in graduates?: Using online job posting data to support policy and practice in higher education. *OECD Education Working Papers,* 231. https://doi.org/10.1787/bf533d35-en

Cai, X., Li, W., & Zhang, R. (2014). Combining co-clustering with noise detection for theme-based summarization. *ACM Transactions on Speech and Language Processing (TSLP*), *10*(4), 1–27.

CEWIL Canada. (2021). CEWIL resource hub: Co-operative education. https://cewilcanada.ca/CEWIL/CEWIL/Resources/Resource-Hub/Co-operative%20Education.aspx

Choi, J., Foster-Pegg, B., Hensel, J., & Schaer, O. (2021). Using graph algorithms for skills gap analysis. *2021 Systems and Information Engineering Design Symposium (SIEDS),* 1–6.

Chopra, S., & Golab, L. (2018). Job description mining to understand work-integrated learning. *Proceedings of the International Conference on Educational Data Mining (EDM),* 32-43.

Chopra, S., & Golab, L. (2022). Analyzing ranking strategies to characterize competition in the co-operative education job market. *International Journal of Work Integrated Learning, 23*(4), 543-564.

Collarana, D., Galkin, M., Lange, C., Scerri, S., Auer, S., & Vidal, M.-E. (2018). Synthesizing knowledge graphs from web sources with the MINTE ̂++ Framework. *The Semantic Web–ISWC 2018: 17th International Semantic Web Conference*, 359–375.

de Groot, M., Schutte, J., & Graus, D. (2021). Job posting-enriched knowledge graph for skills-based matching. *arXiv*, Article 2109.02554.

Eames, C., & Coll, R. K. (2010). Cooperative education: Integrating classroom and workplace learning. In S. Billett (Ed.), *Learning through practice: Models, traditions, orientations and approaches* (pp. 180–196). Springer.

Farhadi, F., Hoseini, E., Hashemi, S., & Hamzeh, A. (2012). Teamfinder: A co-clustering based framework for finding an effective team of experts in social networks. *2012 IEEE 12th International Conference on Data Mining Workshops,* 107–114.

Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, *486*(3-5), 75–174.

Giabelli, A., Malandri, L., Mercorio, F., Mezzanzanica, M., & Seveso, A. (2021). Skills2Job: A recommender system that encodes job offer embeddings on graph databases. *Applied Soft Computing, 101,* Article 107049.

Jia, S., Liu, X., Zhao, P., Liu, C., Sun, L., & Peng, T. (2018). Representation of job-skill in artificial intelligence with knowledge graph analysis. *2018 IEEE symposium on product Compliance Engineering-Asia (ISPCE-CN),* 1–6.

Jiang, Y., & Golab, L. (2016). On competition for undergraduate co-op placements: a graph mining approach. *Proceedings of the International Conference on Educational Data Mining (EDM),* 394-399.

Latapy, M., Magnien, C., & Del Vecchio, N. (2008). Basic notions for the analysis of large two-mode networks. *Social Networks, 30*, 31–48.

Lisa, E., Hennelova, K., & Newman, D. (2019). Comparison between employers' and students' expectations in respect of employability skills of university graduates. *International Journal of Work Integrated Learning, 20*(1), 71-82

Lyu, W., & Liu, J. (2021). Soft skills, hard skills: What matters most? Evidence from job postings. *Applied Energy, 300*, Article 117307.

Raison, K., Tomuro, N., Lytinen, S., & Zagal, J. P. (2012). Extraction of user opinions by adjective-context co-clustering for game review texts. *Advances in Natural Language Processing: 8th International Conference on NLP*, 289–299.

Rege, M., Dong, M., & Fotouhi, F. (2008). Bipartite isoperimetric graph partitioning for data co-clustering. *Data Mining and Knowledge Discovery, 16*, 276–312.

Rossetti, G., Milli, L., & Cazabet, R. (2019). CDLIB: a python library to extract, compare and evaluate communities from complex networks. *Applied Network Science, 4*, 1–26.

Samek, L., Squicciarini, M., & Cammeraat, E. (2021). The human capital behind AI: Jobs and skills demand from online job postings. *OECD Science, Technology and Industry Policy Papers*, 120.

Shalaby, W., AlAila, B., Korayem, M., Pournajaf, L., AlJadda, K., Quinn, S., & Zadrozny, W. (2017). Help me find a job: A graph-based approach for job recommendation at scale. *IEEE International Conference on Big Data*, 1544–1553.

Sodhi, M. S., & Son, B.-G. (2010). Content analysis of OR job advertisements to infer required skills. *Journal of the Operational Research Society, 61*, 1315–1327.

Tackx, R., Tarissan, F., & Guillaume, J.-L. (2018). ComSim: A bipartite community detection algorithm using cycle and node's similarity. In C. Cherifi, H. Cherifi, M. Karsai, & M. Musolesi (Eds.), *Complex networks & their applications* (pp. 278–289). Springer.

Taguchi, H., Murata, T., & Liu, X. (2020). Bimlpa: community detection in bipartite networks by multi-label propagation. *Proceedings of NetSci-X 2020: Sixth International Winter School and Conference on Network Science*, 6, 17–31.

Thiel, G. R., & Hartley, N. T. (1997). Cooperative education: A natural synergy between business and academia. *SAM Advanced Management Journal, 62*(3), 19.

Toulis, A., & Golab, L. (2017). Graph mining to characterize competition for employment. *Proceedings of the 2nd International Workshop on Network Data Analytics*, 1-7.

Traag, V. A., Waltman, L., & Van Eck, N. J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. *Scientific Reports, 9*, Article 5233.

Upadhyay, C., Abu-Rasheed, H., Weber, C., & Fathi, M. (2021). Explainable job-posting recommendations using knowledge graphs and named entity recognition. *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 3291–3296.

Xu, G., Zong, Y., Dolog, P., & Zhang, Y. (2010). Co-clustering analysis of weblogs using bipartite spectral projection approach. In R. Setchi, I. Joranov, R. J. Howlett & L. C. Jain (Eds.), *Knowledge-based and intelligent information and engineering systems* (pp. 398–407). Springer.

Zweig, K. A., & Kaufmann, M. (2011). A systematic approach to the one-mode projection of bipartite graphs. *Social Network Analysis and Mining, 1*, 187–218.

APPENDIX A

For the three largest communities in the skill graph, the following tables list all academic clusters and the top 10 skills. The degree and weighted degree of each node within its community is reported. 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 represent how strongly each node is linked to other nodes in its community. The top skill representatives were selected after sorting by weighted degree (from high to low) within the given community. This represents how strongly the given skill is connected to its community. Tables 1-3 correspond to skill graph communities 1-3 in the main text.

TABLE 1: 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	ARTS Social Sciences	824	19,236
N/A	ARTS English Language and Literature	799	15,724
N/A	AHS Recreation and Leisure Studies	669	12,819
N/A	ARTS Humanities	791	12,509
N/A	ARTS Political Science	722	12,035
N/A	ARTS Sociology and Legal Studies	733	11,700
N/A	ARTS/SCI Psychology	719	11,111
N/A	ARTS Languages and Cultures	672	8,149
N/A	ARTS Fine and Performing Arts	655	7,322
communic	'communication skills', 'communication', 'verbal communication skills', 'communicate'	9	2,920
excel	'excellent', 'excel', 'excellence', 'excels', 'excelled'	9	2,484
team	'team', 'team player', 'teams', 'teamwork'	9	2,208
knowledg	'knowledge', 'knowledgeable', 'knowledgable'	9	1,542
written	'written'	9	1,512
manag	'management', 'manage', 'managing', 'manager', 'managers'	9	1,257
creativ	'creative', 'creativity', 'creatively'	9	1,210
timemanag	'time management', 'manage time', 'time manager', 'time manage', 'manages time'	9	1,198
msoffic	'microsoft office', 'ms office', 'msoffice', 'current ms office', 'ms office365'	9	1,156
attentiontodetail	'attention to detail', 'attention to details', 'high attention to detail', 'amazing attention to detail'	9	1,121

TABLE 2: 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	ENG Software Engineering	1,623	6,483
N/A	ENG Electrical and Computer Engineering	1,585	6,171
N/A	MATH Computer Science	1,510	5,987
N/A	ENG Systems Design and Biomedical Engineering	998	3,209
N/A	ENG Mechanical and Mechatronics Engineering	966	3,040
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

TABLE 3: 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	ARTS Business	537	1,506
N/A	MATH Business	481	1,239
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

APPENDIX B

For the three largest communities in the job title graph, the following tables list all academic clusters and the top 10 job titles. The degree and weighted degree of each node within its community is reported. 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 represent how strongly each node is linked to other nodes in its community. The top representatives were selected after sorting by weighted degree (from high to low) within the given community. This represents how strongly the given title is connected to its community. Tables 1-3 correspond to skill graph communities 1-3 in the main text. Note that identifying information, such as specific course numbers or employer names, is replaced with [REDACTED] in the results below for privacy reasons.

TABLE 1: The academic advertising clusters and top 15 job titles from Community 1 in the job title graph, sorted by weighted degree within the community.

Node ID	Degree in Community	Weighted Degree in Community
ENG Software Engineering	1,135	1,966
MATH Computer Science	1,123	1,947
ENG Electrical and Computer Engineering	902	1,547
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

TABLE 2: The academic advertising clusters and top 15 job titles from Community 2 in the job title graph, sorted by weighted degree within the community.

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

TABLE 3: The academic advertising clusters and top 15 job titles from Community 3 in the job title graph, sorted by weighted degree within the community.

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