The Digitization of College Labor Markets - A Study of the Labor Intermediary, Handshake

Christine Chen

chric16@uci.edu

January 2020

The Digitization of College Labor Markets - A Study of the Labor Intermediary, Handshake

Christine Chen

January 2020

Abstract

Online labor intermediaries have become one of the most widely utilized means of job search within college campuses. Newly developed technologies facilitating job search through filtering methods and preference indicators aid the matching process between employers and job candidates. In doing so, these technologies strive to overcome application congestion, enabling employers to find and respond to qualified applicants within a timely manner. However, filtering may also raise concerns regarding the likelihood of application review based upon a candidate's profile. Our empirical study analyzes Internet & Software industry employers, job ads, and job applications within Handshake, the online labor intermediary deployed by the University of California Irvine. We find that employers are more likely to review applications submitted by candidate profiles that indicate a 'full match' with employer preferences (major, GPA, school year, U.S. work authorization), as well as those for part-time positions, and those within large companies. These findings explore the role of newly developed features and filtering methods within online labor markets.

JEL Code: O33, J40
Keywords: Labor Economics, Preference Signaling, Digitization of Work, Matching Markets, Technology

1 Introduction

Labor intermediaries provide a platform for employers to share information about their company, job vacancies, and desirable employee qualifications. It also serves as a platform for job seekers to share their own experiences and backgrounds. Through inbuilt competitive search and matching algorithms, job seekers can set filters to search and apply to specific types of job postings. In ideal conditions, the vast number of workers and employers participating in the online labor intermediary would create "thick markets", an idealistic market condition discussed by Niederle, Roth, and Sonmez (2007). Thick markets increase diversity within candidate pools and the probability that there exists a suitable match between a job seeker and job posting. While labor market intermediaries support these optimal matches, the ease of submitting job applications could consequentially result in "congestion externalities" and an over-saturation of job applications, preventing employers from thoroughly reviewing applications and resumes in a timely manner. The solution is advanced technology that optimizes competitive search and efficiently clears the online labor market. In recent years, newly developed features within online labor intermediaries have done just that. These platforms provide employers and job candidates with the ability to create online profiles and signal their preferences as a means of communication in order to facilitate job matches and overcome congestion within the online labor market.

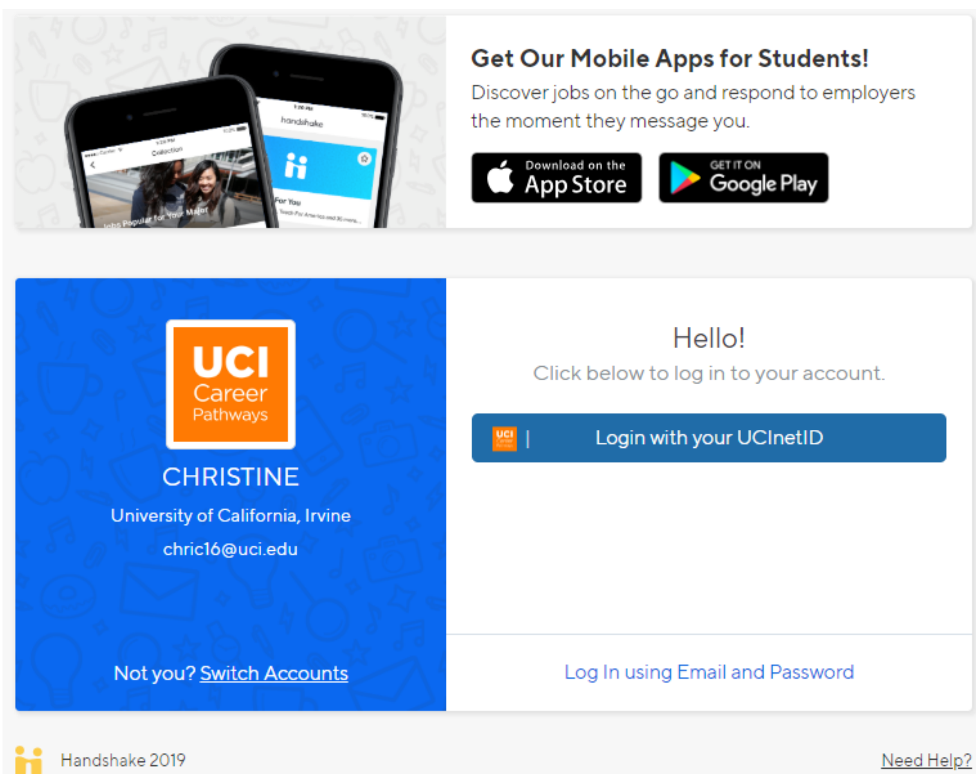
Our study focuses on the college labor market within the online labor intermediary platform currently deployed by the Division of Career Pathways within the University of California Irvine: Handshake. Handshake is a career platform for college students, alumni, and employers within a university's network. Overseen by university career centers across the United States, Handshake facilitates job search and matching processes between employers and applicants. As of 2019, 33,000 enrolled college students attended the University of California Irvine. Within the time span of our study, 12,004 of these students had utilized Handshake's services by submitting at least 1 job application through the online job portal. Through Handshake, students can create online profiles and identify their academic backgrounds, professional experiences, and skills. Similarly, employers can create online profiles and identify their professional industry, company size, job posting, job type, salary type, and applicant preferences. Students and employers can search for potential matches, according to their profiles and preferences. What raises interesting discussion is how preference indicators influence application turnouts and whether the online intermediary effectively facilitates job search so that employers are able to review applications to find and respond to qualified candidates within a timely manner.

In our study, we reference prior research conducted in labor economics, preference signaling, digitized job search, and matching markets. Discussions within these fields provide context to online job search. These discussions lead to the key focus of our study: role of newly developed features and filtering methods within online labor markets. Within the Data section, we examine 3 data sets analyzed to study Internet & Software employers, job ads, and candidate applications. Within the Methods section, we describe the models constructed to identify key features linked to specific application turnouts. Online job search involves numerous external factors, including in-built job recommendation systems, resume parsing technology, and web links to external applications. While our study is limited to the data collected through the Handshake job portal, we acknowledge additional factors that influence job search outside of the scope of our study. For instance, preference indicators may discourage candidates from applying to a job ad. If the candidate does not submit an application, this is not captured in the data. Instead, our study merely focuses on the applications that were submitted. Using this data, we analyze the status of each application submission based on features within Handshake's job portal.

1.1 Navigating the Handshake Website

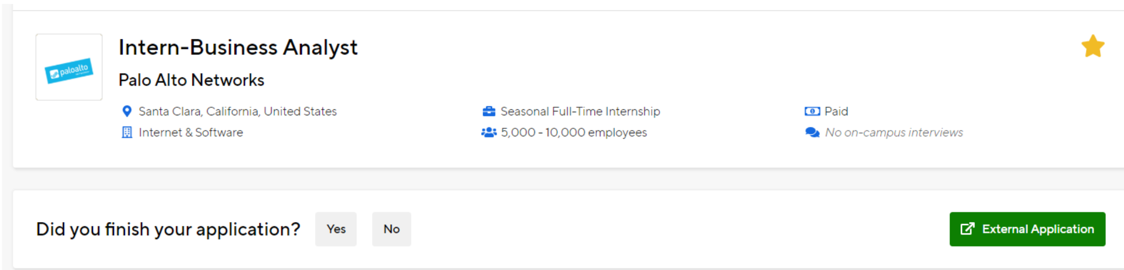
To utilize Handshake's services, the student must first activate their university account. After logging in with their registered UC Irvine Net-ID, issued by the university administration, the student is verified as either an undergraduate student, graduate student, or recent alumni who had enrolled at the University of California Irvine. The student has the option of uploading their resume, posting their work experiences, volunteer work, awards, and skills to personalize their online profile. Basic information, such as student name, major, and year, is preset through the student's university profile.

Handshake Sign In

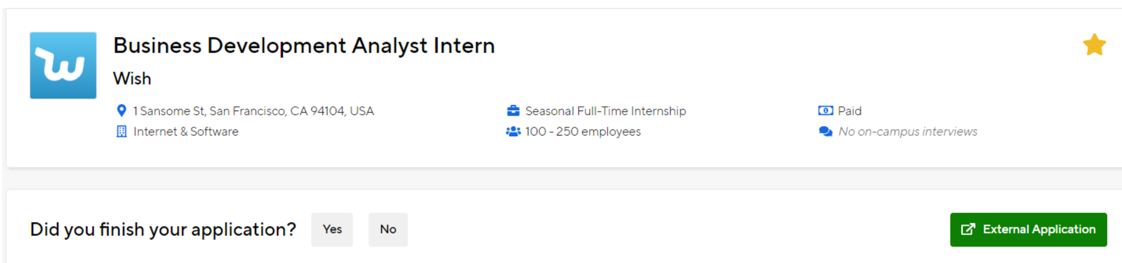


Navigating the homepage, the student will see job ads posted by registered employers. To search for a job ad, the student can set filters that return search results meeting their own search criteria. These filters include job type, employer preferences, work authorization, industry, job function, major, and employer name. The student can directly search for their sought-after role by typing a position title or company name into the Handshake search bar. After clicking on a job posting, the student will see a description of the job posting and the employer's profile, which indicates the employer's location, industry, job type, company size, salary type, and whether they hold on-campus interviews. Companies can also choose to include a separate link to an external application.

Handshake Employer Profile 1

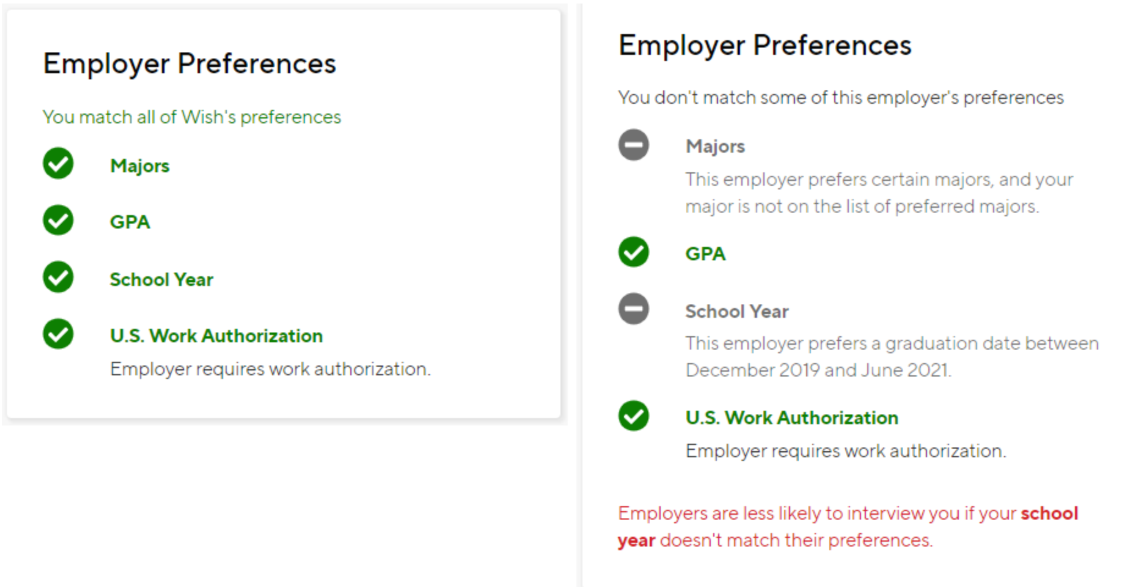


Handshake Employer Profile 2



The student will see whether they meet the 4 basic preferences set by the employer: major, grade point average, school year, and U.S. work authorization. These indicated preferences are met if the student's major is listed as a preferred field of study by the employer, meets the employer's minimum GPA threshold, falls into the employer's preferred year in college, and has legal authorization to work in the United States. It should be noted that students can manually log into their Handshake profile and set their U.S. work authorization status. If the student meets all of the employer's basic preferences, they are considered "fully qualified", which is indicated in their application profile. If a single preference is not met, they are "not fully qualified", which is marked with an indicator visible to both the student and employer. Students are able to apply regardless of their qualification status.

Handshake Employer Preferences 1



Once the student submits an application, the employer has the ability to review his or her application. Once an application is reviewed, it no longer has a ”pending” status. If the employer wishes to contact the applicant, they can email the applicant or send a message through Handshake. Handshake's platform does not capture communication beyond this point. If the employer decides to move forward with the application, they can log into Handshake, take down the job posting, and change the status of the application to ”hired”, ”primary”, ”alternate”, ”reviewed”, or ”declined”. Applications left unopened have a ”pending” status. Further explanation regarding application statuses is discussed in the Data Section.

2 Literature Review

Research literature regarding online job markets first arose when the the world wide web launched in the early 1990's. Online job listings and websites have become what are now known as “labor market intermediaries.” Some of the most well-known labor market intermediaries today include Indeed, CareerBuilder, Glassdoor, ZipRecruiter, and Monster. These online career platforms allow employers and job seekers to create online profiles, indicate their preferences, and search for matches. Through in-built matching and collaborative filtering algorithms, the online search engine returns a list of job postings that align with the indicated search criteria sought by the job seeker. These types of search engines operate as collaborative filters. Ekstrand, Riedl, and Konstan (2010) describe collaborative filtering methods as a means of utilizing a variety of algorithmic processes, including baseline predictors, user-to-user, item-to-item, dimension reduction, probabilistic methods, and hybrid recommendation systems, to generate a list of items that best fit the search parameters provided. The determination of these results is based on a similarity score. Many times, this is a correlation measure between the search criteria and the job ad. Because Handshake's online search engine allows students to set filters and enter in search queries, the orientation and display of generated results could in itself, influence the candidate's application process. In addition, Handshake's search engine utilizes a hybrid recommendation system that assists users by analyzing numerous data points, such as jobs that students with similar backgrounds have applied to, jobs that may interest students based on their search history and preferences, as well as jobs related to positions previously submitted by the applicant.

As noted in prior research, the effects of competitive search and online job boards have transformed the environment of today's labor market. Through channels of worker-firm communication, Autor (2001) states that the internet will change the way matches are made. Because the hiring process has moved to the internet, the dynamics of supply and demand have and will continue to shift within the labor market. In addition to these changes, Stevenson (2009) discusses how information availability, reduced costs, and increased rates of information exchange will alter job search behavior. Due to explicit information exchange, the ease of digitized search potentially lowers the rate of frictional unemployment. The sophisticated, personalized methods for finding suitable jobs and employees, as discussed by Mang (2012), confirms that matches resulting from online labor intermediaries perform more efficiently compared to job postings via newspaper, job agencies, and other channels. This may be due to the availability of information provided by both the employee and employer.

Shimer and Smith (2001), Rogerson, Shimer, and Wright (2005), and Wright, Kircher, Julien, Guerieri (2017) have modeled the process of job search through competitive and directed search theory. Because employers and job seekers reveal information about their own preferences through online search, the probability distribution of generated results deviate from what would have otherwise been a uniform distribution under random search, since all job ads would have been equally visible to the job seeker had he or she been simply presented with a list of job ads. The nature of competitive and directed search alter and influence the search results presented to a potential job candidate. This influences the job search and application process for candidates, as they will most likely apply to the top search results returned from the algorithm.

Preference signaling and information asymmetry can also influence the quality of matches and types candidates who choose to apply. As shown by Banfi and Roldan (2017), explicit wages were used to primarily target low-skilled candidates for low-skilled jobs. Meanwhile, employers strategically posted implicit-wage jobs to signal wage negotiations for high-skilled workers. Economists have studied labor intermediaries from various perspectives to understand how information asymmetry and realistic conditions deviate from an ideal model where both candidates and employers have perfect information. Studies conducted by Coleman (1991) and Kelso and Crawford (1982) assume perfect information symmetry to derive a stable equilibrium where workers have discrete choices among jobs. On the contrary, studies by Faberman and Kudlyak (2014) have studied the labor markets under conditions holding information symmetry to be unrealistic. They found that job applicant heterogeneity, information asymmetry, and congestion externalities led to prolonged search duration and greater search intensity.

Studies in job descriptions and employer profiles have found that language can influence the composition of job applicant pools. Marinescu and Wolthoff (2012) concluded that job titles explain more than 80% of applicant education levels and experiences. Cohen and Klepper (1992) and Borovickova and Shimer (2018) have also attributed corporation size and technological progress as driving factors that influence job search and how applicants perceive wages and jobs. However, the composition of applicant pools may continue to evolve according to its prior state and interactions between new and existing applicants. Mang (2012) and Seabright and Sen (2015) have found that the marginal effect of adding candidates into the applicant pool can vary by the total number of applications, as well as interactions between the cost of application, language used in the job posting, quality of the applicant (high ability versus low ability), and diversity between candidates.

Because there are a multitude of factors and interactions that influence applicant perception and motivation, our study will segregate job applications by industry and take a closer look into application turnouts within the Internet & Software industry, due to its substantially greater population size on Handshake. We will analyze the interactions between applicants and employers through the job application process to better understand the online labor market landscape. By identifying and analyzing applications within the industry that receives the greatest number of applicants, we hope to learn about the relationship between the types of applicant pools targeted by employers and the ones who apply. This may provide further insight into the externalities, either positive or negative, that arise from online labor market intermediaries.

3 Data

The data section covers all expired job ads posted on Handshake between the time that the labor intermediary platform was first introduced to the University of California Irvine to the date of data exportation (May 5, 2017 – November 6, 2018). This provides a 550 day time frame. Although Handshake operates across multiple universities, our study solely looks at job ads posted through the UC Irvine Handshake job portal. Job seekers can navigate Handshake's website by setting filters to find specific job postings or employers by indicating their desired job type, role, location, and wage type. Job positions posted on Handshake are only visible to enrolled UC Irvine undergraduates, graduate students, and recent alumni. In our study, we will refer to 3 data sets collected from Handshake: an Employer data set, Job Ads data set, and Applications data set. By combining each data set and joining tables, we can map applicants to job postings and link job postings to employer profiles.

3.1 Employers

The Employers data set contains 16,001 registered employers who had posted job ads on Handshake. After data cleansing, 4 categorical variables were used for our analysis: Employer ID, Employer Name, Employer Industry, and Employer Size. The data set used covers 62 unique employer industries. Employer Size is represented in intervals: 50-100, 100-250, 250-1000, 1000-5000, 5000-10,000, 10,000-25,000, 25,000+ employees, or not reported (NA). Unreported Employer Size is due to numerous factors, including the assertion of invalid data values or lack of information within the employer's profile. Because there was a random distribution of missing Employer Size data, it holds no meaningful significance in this study. Small to mid-sized firms had fewer than 25,000 employees, and large firms had over 25,000 employees. In our study we combined employers into 3 groups: those that had less than 25,000 employees, 25,000+ employees, and those that did not report Employer Size.

Employers Summary

N	**16001**		
No	**Variable**	**Values**	**Frequency of Valid**
1	**Employer Name**	1. Marcus & Millichap	5
		2. Social Security Administration	5
		3. UBS Financial Services	3
		4. University of California Irvine	3
		5. 99 Ranch Market	2
		[15925 others]	15983
2	**Employer Size**	1. 50-100	1719 (20.9%)
		2. 100-250	1667 (20.2%)
		3. 250-1000	1972 (24.0%)
		4. 1000-5000	1335 (16.2%)
		5. 5000-10000	446 (5.4%)
		6. 10000-25000	439 (5.3%)
		7. 25000+	654 (7.9%)
		8. NA	7769 (48.55%)
3	**Employer Industry**	1. Internet & Software	1353 (8.5%)
		2. Healthcare	1072 (6.7%)
		3. Non-Profit	827 (5.2%)
		4. K-12 Education	795 (5.0%)
		5. Higher Education	682 (4.3%)
		6. Other Industries	641 (4.0%)
		7. Advertising, PR, & Marketing	610 (3.8%)
		8. Manufacturing	585 (3.7%)
		9. Other Education	533 (3.3%)
		10. Legal & Law Enforcement	448 (2.8%)
		[52 others]	8375 (52.6%)

A majority of employers that post jobs on Handshake operate within the Internet & Software, Healthcare, Non-Profit, K-12 Education, and Higher Education industries. Employers that did not report an industry were labeled as "Missing".

3.2 Job Ads

The Job Ads data set represents the 52,073 job ads posted on Handshake from May 5, 2017 through November 6, 2018. All of these positions had expired at the time of the study, and no longer accepted additional applications. The Job Ads summary lists the 7 variables contained in the data set: Job Title, Applicants, Job Type, Employer, Employer Industry, Explicit Wage, and Salary Type. The number of applicants for a single job posting ranged from 0 to 525. The data distribution was heavily left-skewed, with most jobs receiving few to no applications. Job type is separated into 3 categories: Full-Time, Part-Time, and Seasonal. Whether a job posting reports a wage or not is indicated by the variable, "Explicit Wage". This value is represented by a logical value (True or False). Reported job salaries vary between hourly, annual, and contract wages. Because there is no clear indicator of how wages are classified or the accuracy of reported wages, we do not factor these values into our study. Instead, we evaluate which job postings report explicit wages and which do not. This variable is used to indicate whether an employer had utilized a means of wage signaling on applicant behavior. The variable, "Salary Type" indicates whether a job is paid or unpaid. This is a variable set by employers when they post the job on Handshake.

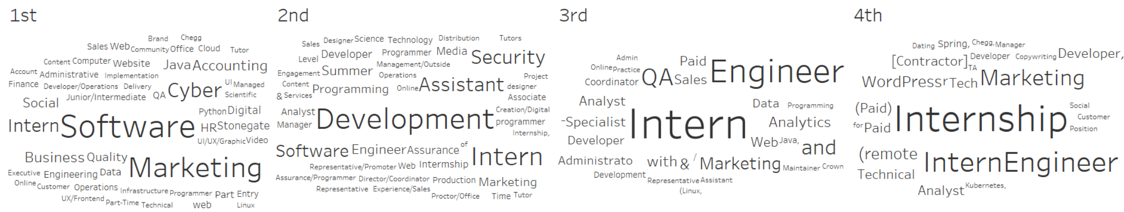
Job Ads Summary

N	52,073		
No	**Variable**	**Values**	**Frequency (% of Valid)**
1	**Status**	1. alternate	67 (0.1%)
		2. declined	6136 (5.4%)
		3. hired	264 (0.2%)
		4. pending	99532 (87.4%)
		5. primary	1070 (0.9%)
		6. reviewed	6800 (6.0%)
2	**Fully Qualified**	1. FALSE	45108 (39.6%)
		2. TRUE	68761 (60.4%)
3	**Employer Name**	1. University of California	2192 (1.9%)
		2. Moboware Inc.	1368 (1.2%)
		3. University of California	1364 (1.2%)
		4. Tesla	1085 (1.0%)
		5. Amazon	927 (0.8%)
		6. Dell, Inc.	902 (0.8%)
		7. Siemens Corporation	834 (0.7%)
		8. University of California	792 (0.7%)
		9. Microsoft	788 (0.7%)
		10. Volt Workforce Solutions	776 (0.7%)
		[4113 others]	102841 (90.3%)
4	**Title**	1. Software Engineer	1789 (1.6%)
		2. Software Engineer Intern	1591 (1.4%)
		3. Student Assistant	838 (0.7%)
		4. Marketing Intern	761 (0.7%)

	5. Software Engineer Intern	644 (0.6%)
	6. Software Development Intern	547 (0.5%)
	7. Administrative Assistant	529 (0.5%)
	8. Junior Software Engineer	520 (0.5%)
	9. Office Assistant	495 (0.4%)
	10. Accounting Intern	480 (0.4%)
	[8856 others]	105675 (92.8%)
5 Employment Type	1. Full-Time	65291 (57.3%)
	2. Part-Time	-42.60%
	3. Seasonal	23 (0.0%)
6 Employer Industry	1. Internet & Software	22849 (20.1%)
	2. Higher Education	14172 (12.5%)
	3. Accounting	6281 (5.5%)
	4. Electronic & Computer Hardwar	6176 (5.4%)
	5. Healthcare	4816 (4.2%)
	6. Investment/Portfolio Managem	4111 (3.6%)
	7. Manufacturing-Other	3671 (3.2%)
	8. K-12 Education	3391 (3.0%)
	9. Other Industries	3311 (2.9%)
	10. Human Resources	2852 (2.5%)
	[50 others]	41985 (37.0%)
7 Salary Type	1. Paid	106802 (93.9%)
	2. Unpaid	6979 (6.1%)
8 Employer Size	1. 50-100	6243 (7.9%)
	2. 100-250	8497 (10.7%)
	3. 250-1000	10473 (13.2%)
	4. 1000-5000	14801 (18.7%)
	5. 5000-10000	4882 (6.2%)
	6. 10000-25000	6856 (8.6%)
	7. 25000+	27504 (34.7%)
9 Explicit Wage	1. FALSE	68580 (60.2%)
	2. TRUE	45289 (39.8%)

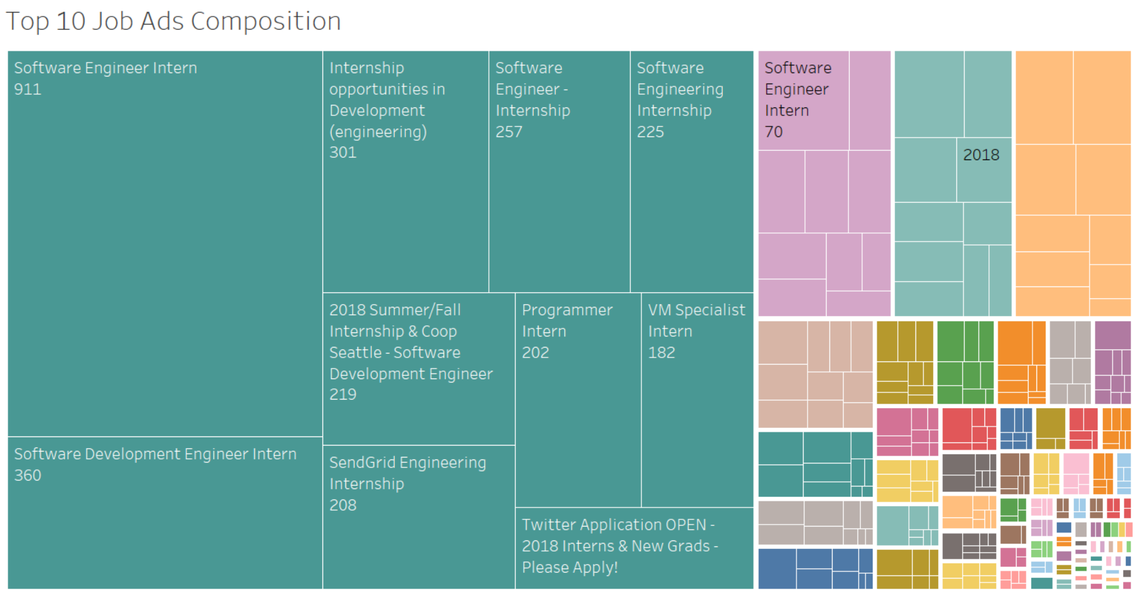
Across the entire Job Ads data set, the most common job titles for Handshake job ads are as follows: Software Engineer, Administrative Assistant, Sales Development Representative, Marketing Intern, and Account Manager. For our study, we solely look at job postings within the Internet & Software industry. Using this criteria, we create four word clouds by separating the job title into its first 4 words, a method introduced by Marinescu and Wolthoff (2012) in their study of the influence and explanatory power of job titles on wage variance and the job matching process. As discussed in their paper, the first 4 words contain key information regarding the type of job role.

Job Ads Summary



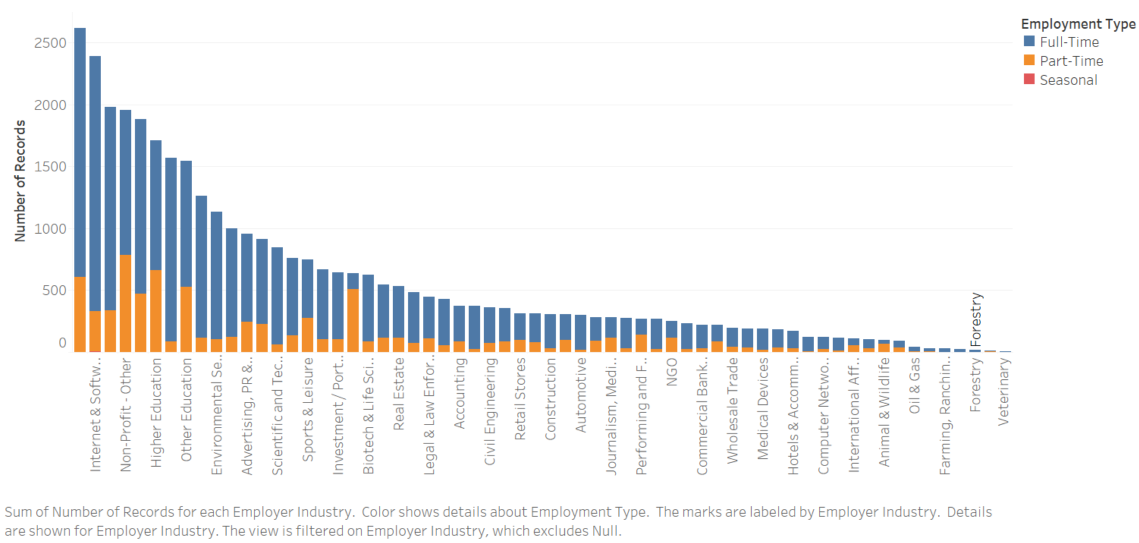
The size of each word corresponds to the frequency in which the word appears in job titles within the Internet & Software industry. As shown in the Internet & Software job title word clouds, some of the most common job positions were Software Development Engineers, Software Development Interns, Marketing Interns, and Cyber Security Interns.

Composition of Job Applications by Major



The tree map breaks down the applications submitted to job ads within the Internet & Software industry. The aqua-green block represents applications submitted by students studying Computer Science. This block is broken down further to indicate the number of job applications received based on job ad title. A majority of students who had applied to Software Engineer Intern positions within handshake had studied Computer Science.

K-12 Job postings dominate job board



A majority of job ads are full-time positions (77.3%) while 22.7% are part-time positions. Job ads within K-12 Education dominated the job board, most of which were full-time positions. Jobs within the Internet & Software, Non-Profit, and Higher Education industries follow respectively. A majority of job ads are paid positions (92.7%).

3.3 Applications

The Applications data set contains information about the applicants who had applied to job ads on Handshake between May 5, 2017 – November 6, 2018. The entire Applications data set contains 113,869 application records, of which 12,004 are unique applicants. Each record represents a student and the job ad they had applied to, identified by a unique application ID, employer ID, and job ID. We find that the number of applicants vary across job positions on Handshake, ranging between 0 to 525 applicants for a single position.

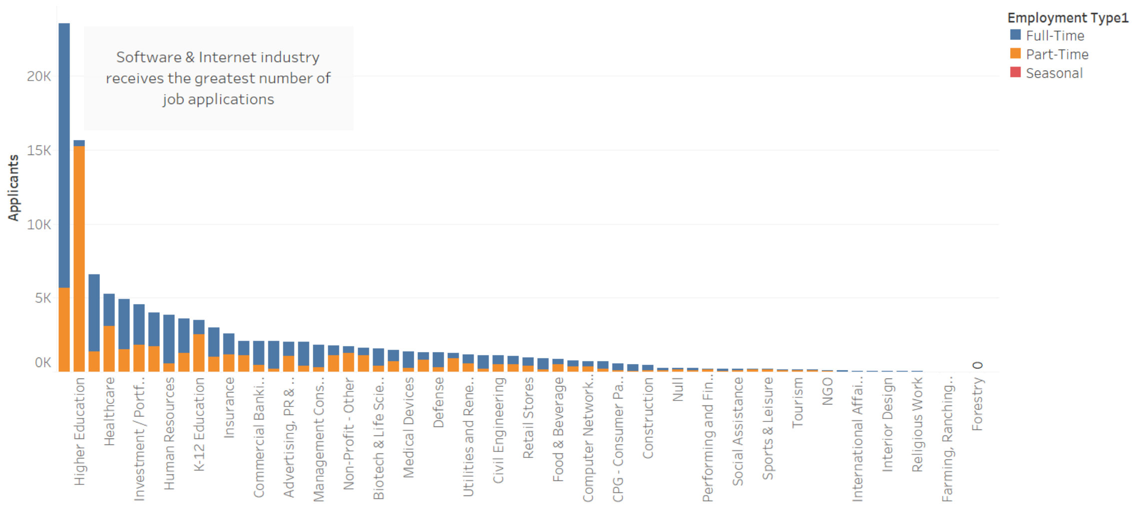
After cleaning the data set, we decide to keep 8 values for the study: Status, Fully Qualified, Employer Name, Title, Job Type, Employer Industry, Salary Type, and Explicit Wage. The Internet & Software industry received 20.1% of applications submitted through Handshake, which is 7.6% greater than Higher Education, the next leading industry. Due to this large gap, we choose to solely focus on job applications submitted to the Internet & Software industry. Applicants who are identified as 'fully qualified' fulfill the employer's major, grade point average, school year, and U.S. work authorization preferences. All application data is collected through the UC Irvine Handshake job portal.

Application Summary

N	113869		
No	Variable	Values	Frequency (% of Valid)
1	Status	1. alternate	67 (0.1%)
		2. declined	6136 (5.4%)
		3. hired	264 (0.2%)
		4. pending	99532 (87.4%)
		5. primary	1070 (0.9%)
		6. reviewed	6800 (6.0%)
2	Fully Qualified	1. FALSE	45108 (39.6%)
		2. TRUE	68761 (60.4%)
3	Employer Name	1. University of California	2192 (1.9%)
		2. Moboware Inc.	1368 (1.2%)
		3. University of California	1364 (1.2%)
		4. Tesla	1085 (1.0%)
		5. Amazon	927 (0.8%)
		6. Dell, Inc.	902 (0.8%)
		7. Siemens Corporation	834 (0.7%)
		8. University of California	792 (0.7%)
		9. Microsoft	788 (0.7%)
		10. Volt Workforce Solutions	776 (0.7%)
		[4113 others]	102841 (90.3%)
4	Title	1. Software Engineer	1789 (1.6%)
		2. Software Engineer Intern	1591 (1.4%)
		3. Student Assistant	838 (0.7%)
		4. Marketing Intern	761 (0.7%)
		5. Software Engineer Intern	644 (0.6%)
		6. Software Development Intern	547 (0.5%)
		7. Administrative Assistant	529 (0.5%)
		8. Junior Software Engineer	520 (0.5%)
		9. Office Assistant	495 (0.4%)
		10. Accounting Intern	480 (0.4%)
		[8856 others]	105675 (92.8%)
5	Employment Type	1. Full-Time	65291 (57.3%)
		2. Part-Time	-42.60%
		3. Seasonal	23 (0.0%)
6	Employer Industry	1. Internet & Software	22849 (20.1%)
		2. Higher Education	14172 (12.5%)
		3. Accounting	6281 (5.5%)
		4. Electronic & Computer Hardwa	6176 (5.4%)
		5. Healthcare	4816 (4.2%)
		6. Investment/Portfolio Managem	4111 (3.6%)
		7. Manufacturing-Other	3671 (3.2%)
		8. K-12 Education	3391 (3.0%)
		9. Other Industries	3311 (2.9%)
		10. Human Resources	2852 (2.5%)
		[50 others]	41985 (37.0%)
7	Salary Type	1. Paid	106802 (93.9%)
		2. Unpaid	6979 (6.1%)
8	Employer Size	1. 50-100	6243 (7.9%)
		2. 100-250	8497 (10.7%)
		3. 250-1000	10473 (13.2%)
		4. 1000-5000	14801 (18.7%)
		5. 5000-10000	4882 (6.2%)
		6. 10000-25000	6856 (8.6%)
		7. 25000+	27504 (34.7%)
9	Explicit Wage	1. FALSE	68580 (60.2%)
		2. TRUE	45289 (39.8%)

As shown in the Handshake applicant data set, the Internet & Software industry received the greatest number of applications from UC Irvine students. Of these 22,849 applications, 64.79% were full-time and 35.21% were part-time positions. Job ads within Higher Education, Electronic & Computer Hardware, Healthcare, and Accounting industries also received a high number of applications.

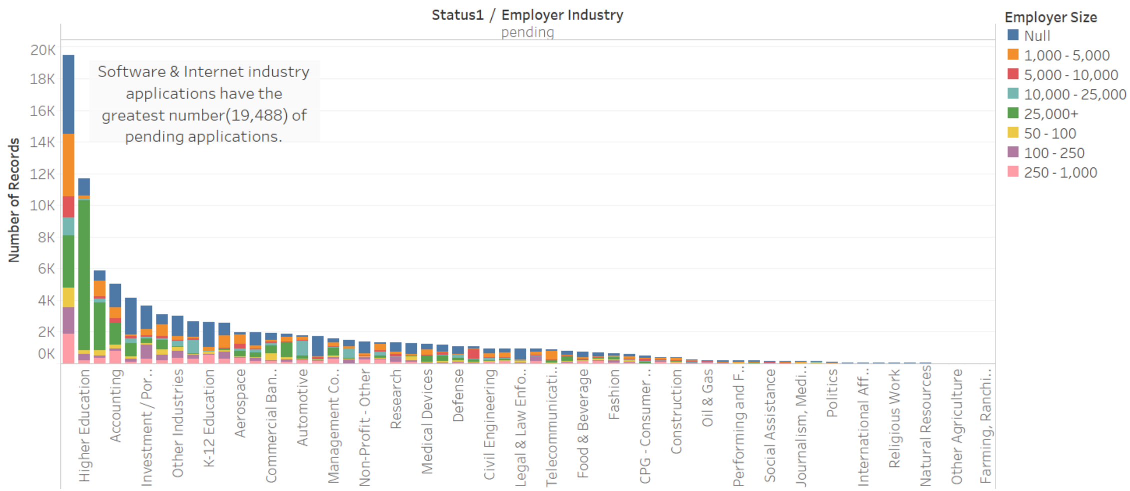
Software Internet Job postings receive the most applicants



In our model, we categorize job ads by the number of applications they receive. This is represented through intervals. If the job ad receives 0 applications, its 'application' variable is set to 'none'. Job ads receiving less than 10 applications are set to 'low'; those between 10 and 50 are set to 'middle'; those between 50 and 100 are set to 'mid-high'; those with over 100 applications are set to 'high'.

A major concern falls within the applications that are left pending. Handshake job ads can contain links to external company websites, where applicants can apply directly to the job ad. As a result, applicants could be hired or denied through third party websites for the positions they had applied for. Across all industries, 87.4% of Handshake applications that were submitted directly through Handshake's job portal have a pending status. Only 6% of applications are reviewed and 5.4% are declined. Of the applications that are reviewed, 0.9% of applicants are primary candidates for hire and 0.1% are alternate candidates. If the primary candidate does not accept the job, the employer will hire the alternate candidate. Our study re-categorizes 'primary' and 'alternate' status applications as 'reviewed' applications because the employer must review these applications to determine whether a candidate is their primary or alternate choice. In addition, applications with primary and alternate statuses comprise 0.1% of the sample size, which could lead to over-fitting within our study.

Pending Applications across Employer Size and Industry



The applicant data set is broken down into 4 categories: those that were reviewed, hired, denied, or left pending through the Handshake job portal. Within the Internet & Software industry, we find that 19,488 of the 22,849 (85.29%) applications have a 'pending' status. This issue is considered in the various approaches taken to model the odds that an application will be reviewed.

4 Methods

In this section, we evaluate various statistical models to analyze factors within the applicant profile and job posting. Factors such as employer preferences, basic qualification match, as well as employer size and industry can all influence the odds of application review.

To map the applications submitted to job ads posted on Handshake, we join the employer, job ads, and applications data sets. Then we separate the applications by industry, as each industry has its own unique set of skilled labor. For our study, we analyze the industry that received the greatest number of applications: Internet & Software, which received 22,849 applications over an 18 month time span. We collapse the original 6 application status categories (hired, declined, reviewed, primary, alternate, and pending) into 4 application statuses: hired, declined, reviewed, and pending. 'Primary' and 'alternate' application are re-categorized under 'reviewed' applications because employers must first review the application before determining whether a job candidate is their primary or alternate choice. If the employer reviews these applications and does not provide a final offer or rejection, these applications are simply left with a 'reviewed' status. In total, there are 19,488 pending, 1,703 declined, 1,617 reviewed, and 41 hired applications.

4.1 Multinomial Logistic Regression

Our first model addresses the discrete nature of the data within our study and calculates the odds of each application status against pending application. Given the 6 categorical dependent variables used to model application turnout, we run a multinomial logistic regression to study the differences in application statuses by job type, employer size, salary type, full match, applicant group size, and explicit wage. To run a multinomial logistic regression, we must have a large sample size, and there must not exist multi-collinearity within the sample. In addition, we must meet the Independence of Irrelevant Alternatives (IIA) assumption, which states that adding or deleting alternative outcomes will not alter the odds of the remaining outcomes. Considering this assumption, we acknowledge a potential change in log odds within application outcomes if an alternative outcome is added or deleted. Although we do not ultimately choose to model our study with the multinomial logistic regression for this reason, there is value in studying the relationship between the log odds of each outcome. To build this model, we consider 6 factors: Job Type, Employer Size, Salary Type, Fully Qualified, Applicant Group Size, and Explicit Wage.

4.1.1 Model

The multinomial logistic regression model considers 4 'Status' categories: Reviewed, Declined, Hired, and Pending. The base model represents a pending candidate who had applied to a full-time paid position for a firm with less than 25,000 employees. This position does not have an explicit wage and received less than 10 applications. The multinomial logistic regression compares the base model to the other application status outcomes (Reviewed, Hired, or Declined). We conduct this test by running the 'multinom' function from the nnet package in R to obtain coefficient parameter estimates.

$$\begin{aligned}\ln\left(\frac{P(Status = k)}{P(Status = Pending)}\right) =& \beta_{k0} + \beta_{k1} JobType + \beta_{k2} EmployerSize + \beta_{k3} SalaryType + \\ & \beta_{k4} FullyQualified + \beta_{k5} ApplicantGroupSize + \beta_{k6} ExplicitWage\end{aligned} \tag{1}$$

where k = {Hired, Declined, Reviewed}

Job Type can take on 1 of 3 values: Full-time, Part-time, and Seasonal. Employer Size is a binary variable: less than 25,000 employees or 25,000+. Salary Type is a binary variable: paid or unpaid. Full match is a binary variable: True or False. Applicant Group Size can take on 1 of 4 values: low (less than 10), middle (10-50), mid-high (50-100), or high (100+). Explicit wage is also a binary variable taking on one of two values: True or False.

$$\ln\left(\frac{Z_i}{Z = Pending}\right) = \Phi(Z_{i=1,2,3}) \tag{2}$$

$$\begin{aligned}Z_1 &= \beta_{10} + \beta_{11}\text{Job Type} + \beta_{12}\text{Employer Size} + \beta_{13}\text{Salary Type}...\beta_{16}\text{Explicit Wage}\\ Z_2 &= \beta_{20} + \beta_{21}\text{Job Type} + \beta_{22}\text{Employer Size} + \beta_{23}\text{Salary Type}...\beta_{26}\text{Explicit Wage}\\ Z_3 &= \beta_{30} + \beta_{31}\text{Job Type} + \beta_{32}\text{Employer Size} + \beta_{33}\text{Salary Type}...\beta_{36}\text{Explicit Wage}\end{aligned}$$

4.2 Logistic Regression

A detailed look into the different application statuses provides insight into significant discrepancies in our data. It should be noted that 87.4% of job applications remain 'pending'. Given this discrepancy between the number of applications that received a status versus those left pending, we decided that the best approach would be incorporating dimension reduction into our model and re-categorize 'hired','declined', and 'reviewed' applications as 'non-pending' applications. This creates 2 groups: pending vs. non-pending applications. Doing so accounts for disproportionate sample sizes, which mitigates the risk of making a type I error and concluding statistical significance based on random chance. With the re-categorization of application outcomes, our research addresses a broader question: What factors influence the odds that an application will receive a status update?

In order to address this question, we run a logistic regression. The logistic regression model can only be applied to independently distributed observations that follow a distribution from an exponential family. In addition, errors must be independent, and the data must have large sample approximations used to estimate the parameters using maximum likelihood estimation (MLE). The data used in the logistic model fulfill the same conditions set by the multinomial logistic regression model. In addition, the errors are independent and normally distributed. With these conditions met, our model compares pending applications to non-pending applications using the same 6 variables: Job Type, Employer Size, Salary Type, Fully Qualified, Applicant Group Size, and Explicit Wage.

4.2.1 Model

To compare pending applications against non-pending applications, we re-code application statuses originally labeled as 'Hired', 'Declined', or 'Reviewed' status as a 'Non-Pending' application, denoted by a '0'. Pending applications are re-coded as '1'. If employers gain a higher utility (ε) for reviewing, hiring, or declining a job application, they will view the application, which changes its status from 'pending' to 'non-pending'. This model is represented as a logistic regression equation.

$$\begin{aligned} \varepsilon > &P(Pending = 1 | \beta_0 + \beta_1 JobType + \beta_2 EmployerSize + \beta_3 SalaryType \\ &+ \beta_4 FullyQualified + \beta_5 ApplicantGroupSize + \beta_6 ExplicitWage) \end{aligned} \tag{3}$$

Using the generalized linear model (glm) function in R, which allows for error distributions that deviate from the normal distribution, we run a logistic regression on application status using a "binomial" family. The base model represents an application for a full-time job for a firm with less than 25,000 employees that does not post an explicit wage. We run the logistic regression by using the generalized linear model (glm) function in R, with a binomial family. This ensures the outcome variable, pending or non-pending, is interpreted as a binary variable.

$$\begin{aligned} \ln\left(\frac{P(Status = Pending)}{P(Status = Non - Pending)}\right) = &\beta_0 + \beta_1 JobType + \beta_2 EmployerSize + \beta_3 SalaryType + \\ &\beta_4 FullyQualified + \beta_5 ApplicantGroupSize + \beta_6 ExplicitWage \end{aligned} \tag{4}$$

5 Results and Discussion

This section presents our findings and analyzes the variables correlated with application turnout. Our initial model runs a multinomial logistic regression to model the factors influencing 'Application Status' and analyze the relationship between each factor and application status. To account for the disproportionate number of applications left pending, we re-categorize the data based on their statuses and run a logistic regression model to analyze the applications that received a status update. Through each model, we gain better insight into significant variables that influence application status.

5.1 Multinomial Logit Model

We first run a multinomial logistic regression to model how factors within the candidate's profile and job ad influence application statuses. The coefficients are interpreted as the log odds of either the declined, hired, or reviewed applications compared to pending applications. The base model represents a pending candidate who had applied to a full-time paid position for a firm with less than 25,000 employees.

Internet & Software Multinomial Logit Status Results

	Dependent Variable		
	Declined	**Hired**	**Reviewed**
Covariates			
Job Type			
Part-Time	1.316***	1.128***	-0.253***
	0.068	0.411	0.066
Seasonal	-8.133	-2.683	-10.063
	39.600	16.938	64.475
Employer Size			
25,000+ employees	-1.603***	3.823***	-2.313***
	0.189	1.055	0.248
Not Reported (NA)	1.417***	3.688***	1.740***
	0.069	1.035	0.062
Salary Type			
Unpaid	-0.979***	0.851	0.613***
	0.224	0.812	0.167
Fully Qualified			
TRUE	-0.247***	1.568***	0.188***
	0.059	0.533	0.059
Applicant Group Size			
Middle	-0.450***	0.221	-0.029
	0.098	0.664	0.081
Mid-High	-0.710***	1.103	-0.606***
	0.119	0.672	0.111
High	0.469***	0.393	0.320***
	0.091	0.665	0.083
Explicit Wage			
TRUE	-0.897***	-0.286	-0.056
	0.078	0.373	0.062
Constant	-3.144***	-11.151***	-3.186***
	0.091	1.243	0.084
Akaike Inf. Crit.	19593.590	19593.590	19593.590

Note: *p<0.1; **p<0.05; ***p<0.01

The results of our preliminary multinomial logistic regression are simply used to investigate the log odds between each possible outcome. They show that part-time employment, employer size, salary type, fully qualified, applicant group size, explicit wage, and the intercept are all statistically significant for declined applicants. Part-time employment, employment size, salary type, fully qualified, and the intercept are statistically significant for hired applicants. Employer size, salary type, fully qualified, mid-high and high applicant group size, and the intercept are statistically significant for applicants whose applications were reviewed. Each significant variable is associated with a change in the log odds of being declined, hired, or reviewed compared to pending by the coefficient value. Considering that 85.29% of Internet & Software applications submitted directly through Handshake's job portal have a pending status, it's easy to see that there's a disproportionate population of declined, hired, reviewed, and pending applications. Our solution was to re-combine these variables into 2 separate groups: pending vs. non-pending.

5.2 Logistic Regression Model

To address the large number of pending applicants, we separate the joined application data set into 2 groups: pending vs. non-pending applications. Then, we construct and run a logistic regression model using the glm function in R with a binomial family. Using a generalized linear model (glm) allows for non-normal dependent variables. The results from the summary output of our model are shown below.

Internet & Software Logistic Regression Status Results

	Pending
Covariates	
Job Type	
Part-Time	-0.9961*
	0.4152
Seasonal	10.1143
	500.0843
Employer Size	
25,000+ employees	-3.9234***
	1.0525
Not Reported (NA)	-3.3897**
	1.0354
Salary Type	
Unpaid	-0.9046
	0.8110
Fully Qualified	
TRUE	-1.5808**
	0.5326
Applicant Group Size	
Middle	-0.2679
	0.6642
Mid-High	-1.2279
	0.6723
High	-0.2384
	0.6673
Explicit Wage	
TRUE	0.1348
	0.3728
Constant	11.2518***
	1.2425
Akaike Inf. Crit.	534.14

Note: *p<0.1; **p<0.05; ***p<0.01

From the results, we find that the intercept, part-time employment, employment size, and fully qualified applications have statistically significant outcomes. If an application is submitted for a part-time position, its log odds of having a pending status are expected to change by a factor of 0.3693. For an application submitted to a company with over 25,000 employees, its log odds of pending are expected to change by a factor of 0.0198. For an application submitted to a company that does not report their company size (NA), its log odds of pending are expected to change by a factor of 0.0337. For an application that is fully qualified for a job position, its log odds of remaining a pending application decrease by 0.2058. From the results we find that applications that fully meet employer preferences, those for part-time positions, and those submitted to large companies with 25,000+ employees or do not reveal their company size see a decrease in the log odds of having a pending status. This means that these applications were more likely to receive job status update, whether that means they were hired, declined, or reviewed by the employer.

Because we simply use the multinomial logistic regression as a tool to assess the relationship between covariates and application status, we find that both models present consistent results. Across both models, there is statistical significance in the following variables: part-time positions, company size, and fully qualified applicants. Given the small number of applications within each separate category of non-pending applications, there may be more explanatory power in the logistic regression model, which collapses application statuses into 2 categories: pending vs. non-pending. Although we could lose information in doing so, it would still be the optimal model for this study as it mitigates the risk of running into a type I error from using heavily disproportionate sample sizes.

6 Conclusion

Our findings show that within the Internet & Software industry, the 'Job Type', 'Employer Size', and whether an applicant is 'Fully Qualified' have significant influence on application status. Running a logistic regression, we find that the log odds of pending applications decrease in 3 scenarios: part-time positions, companies with 25,000+ employees, and fully qualified applicants. In other words, the odds that an application gets reviewed increases if the application is submitted for a part-time position, to an employer with over 25,000+ employees, or if the applicant is fully qualified based on employer preferences set on Handshake.

Our findings could be attributed to numerous factors. Part-time positions, which comprise 35.21% of the 22,849 applications submitted within the Internet & Software industry could have higher odds of being marked as non-pending due to the nature of the work, the faster hiring process for part-time workers, or simply because they comprise a smaller percentage of all applications submitted through Handshake. Meanwhile, the odds of having an application marked as non-pending may increase for applications submitted to large employers due to the availability of human resource services that have the technological capability of processing job applications in bulk, thereby allowing large firms to provide a response to an applicant within a timely manner. The increased odds that a fully qualified application is marked as non-pending could be attributed to the screening process offered through Handshake's profile matching algorithm. By utilizing this tool, employers are able to process applications more quickly and provide a response to students' applications.

Further research into online labor market activity could provide a better measure of the economic gains and costs of digitized job search. These measures could be weighed and determined through a deeper analysis into factors such as employer response rate, a reduction in application congestion, and the creation of stable matches between job candidate and job position. The contribution of this study could assist career counselors and students by providing insight into the digitization of job search, and how to build a profile that increases the likelihood that a submitted application will be reviewed. For instance, with newly developed algorithms that parse hundreds upon thousands of applicant profiles, only a certain subset that score a minimum percentage of keyword matches will have the opportunity to be reviewed. While the result of these filtering methods could benefit candidates who manage to score a high match rate, the scoring system could inadvertently place qualified candidates who somehow hadn't managed to reach this arbitrary marker at a disadvantage. This raises the question of whether the digitization of job search could lead to inequalities within online labor intermediaries, considering how the determination of whether a job application gets reviewed depends on qualities of the candidate profile rather than the application submitted.

The role of labor intermediaries has become a key component of job search, as it determines the initial interactions between job candidates and employers. Algorithmic matching and filtering methods can serve as a means of communication between employer and job candidate without there ever being explicit interaction. As a result, the impact of newly developed technologies vary across participants depending on the information they reveal to the employer and job candidate. All of these factors have the ability to influence the ways in which job applications are processed and evaluated. A deeper understanding of the externalities resulting from technological advancements in job search involved require further research into online labor markets.

Declarations

Availability of data and material

The datasets generated and/or analysed during the current study are not publicly available due to the confidentiality of student data collected through the Division of Career Pathways at the University of California Irvine, but are available from the corresponding author on reasonable request.

Competing interests

Not Applicable

Funding

Not Applicable

Authors' contributions

Christine Chen is the sole contributor of this study.

Acknowledgements

This research project was undertaken at the University of California Irvine. Data was provided by the Division of Career Pathways and completed within the Economics Honors Program.

References

[1] **Acemoglu, D.; Autor, D.** "Lectures in Labor Economics." *Labor Economics MIT Press.*

[2] **Autor, D.** "Economics of Labor Market Intermediation: An Analytic Framework." Working paper, National Bureau of Economic Research, September 2008.

[3] **Autor, D.** "Wiring the Labor Market." *Journal of Economic Perspectives*, 2001, Volume 15, Number 1, pp. 25–40.

[4] **Banfi, S.; Roldan, B. V.** "Do High-Wage Jobs Attract more Applicants? Directed Search Evidence from the Online Labor Market." Working paper, University of Chile, Department of Industrial Engineering, May 2017.

[5] **Banfi, S.; Choi, S.; Roldan, B. V.** "Deconstructing Job Search Behavior." Working paper, Institute of Labor Economics, October 2018.

[6] **Cohen, W. M.; Klepper, S.** "The Tradeoff between Firm Size and Diversity in the Pursuit of Technological Progress."*Small Business Economics*, March 1992, Vol. 4, No. 1, pp. 1-14.

[7] **Coleman, J. S.** "Matching Processes in the Labor Market." *Acta Sociologica*, January 1991, Vol. 34, No. 1, pp. 3-12.

[8] **Ekstrand, M. D.; Riedl, J. T.; Konstan, J. A.** "Collaborative Filtering Recommender Systems." *Foundations and Trends in Human–Computer Interaction*, 2010, Vol. 4, No. 2, pp. 81–173.

[9] **Faberman, R. J.; Kudlyak, M.** "What does online job search tell us about the labor market?" *FRB Chicago Economic Perspectives*. March 2016.

[10] **Kelso, A. S.; Crawford V. P.** "Job Matching, Coalition Formation, and Gross Substitutes." *Econometrica*, November, 1982, Vol. 50, No. 6, pp. 1483-1504.

[11] **Marinescu, I.; Wolthoff, R.** "Opening the Black Box of the Matching Function: the Power of Words." Working paper, The National Bureau of Economic Research, August 2016.

[12] **Niederle, M.; Roth, A. E.; Sonmez, T.** "Matching." *The New Palgrave Dictionary of Economics*, February 2007, 2nd edition.

[13] **Rogerson, R.; Shimer, R.; Wright, R.** "Search-Theoretic Models of the Labor Market: A Survey." *Journal of Economic Literature*, December 2005, Vol. 43, pp. 959–988.

[14] **Seabright, P.; Sen, A.** "Statistical Externalities and the Labour Market in the Digital Age." *Elsevier*, July 2015.

[15] **Shimer, R.; Smith, L.** "Matching, Search, and Heterogeneity." *Advances in Macroeconomics*, 2001, Vol. 1, Issue 1, Article 5.

[16] **Stevenson, Betsey.** "The Internet and Job Search." *The National Bureau of Economic Research*, November 2009, pp. 67 - 86.

[17] **Wright, R.; Kircher, P.; Julien, B.; Guerrieri, V.** "Directed Search: A Guided Tour." Working paper, National Bureau of Economic Research, September 2017.