The Complexity of the Future of Work

by

Morgan Ryan Frank

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in Media Arts and Sciences

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2019

Signature redacted

Author ....
Program in Media Arts and Sciences
April 20, 2019

Signature redacted

Certified by ....
Iyad Rahwan
Associate Professor of Media Arts and Sciences
Thesis Supervisor

Signature redacted

Accepted by ....
Tod Machover
Academic Head, Program in Media Arts and Sciences

The Complexity of the Future of Work

by

Morgan Ryan Frank

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
on April 20, 2019, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in Media Arts and Sciences

Abstract

Rapidly advancing cognitive technologies, such as artificial intelligence (AI), have the potential to drastically impact modern society and to shape the future of work. Accordingly, policy makers and researchers seek forecasts into technological change and labor trends, including growing job polarization and income inequality as well as decreasing career mobility and spatial mobility for workers. Although a given technology impacts demand for only a narrow set of workplace skills, modern empirical work relies on coarse labor distinctions between cognitive and physical or routine and non-routine work to explain employment trends. In this dissertation, I explore the complex ways that skills and employment undergird aggregate labor dynamics in the US. As a motivating example, I demonstrate how simple measures for skills within a labor market contribute to the differential impact of automation across US cities of different sizes. I build on this motivation to address methodological barriers through a refined model of workplace skills and their interdependencies, thus connecting microscopic workplace connections to macroscopic labor trends. I perform an unsupervised analysis of specific workplace skills as a skills network whose aggregate and refined topology grant new insights into job polarization and workers' career mobility. Since these inter-skill connections predict career mobility, I construct a map of US occupations that captures worker transition rates between employment opportunities and, in combination with urban employment data, predicts workers' spatial mobility. These refined models that connect workplace skills to both inter-city and intra-city dynamics enable new insights and new input data sources for real-time labor trends at the level of specific technologies and specific workplace skills. I conclude by exploring one novel and potentially useful source of input information: the evolution of scientific AI research. The analyses in this dissertation provide new tools to policy makers designing viable worker retraining programs, offer new insights to individual workers navigating their careers, and present new measures for economic resilience in the face of changing technology.

Thesis Supervisor: Iyad Rahwan
Title: Associate Professor of Media Arts and Sciences

This doctoral thesis has been examined by the following committee members:

Signature redacted

Professor Iyad Rahwan . . . .

Thesis Supervisor

Associate Professor of Media Arts and Sciences

AT&T Career Development Professor of Media Arts and Sciences

Signature redacted

Professor Alex 'Sandy' Pentland . . . .

Thesis Committee

Professor of Media Arts and Sciences

Toshiba Professor

Media Lab Entrepreneurship Program Director

Signature redacted

Professor Erik Brynjolfsson . . . .

Thesis Committee

Schussel Family Professor of Management Science

Professor of Information Technology

Director of The MIT Initiative on the Digital Economy

Acknowledgments

I have many people to thank for enabling my academic journey thus far. This is only the surface.

First, thank you to my loving parents, Stephen and Ellen Frank. You first taught me to balance pushing myself with having fun. Through literal and metaphorical peaks and valleys, your love and support for me have been rock solid and provided the foundation I rely on. Thank you for being the first to see my intellectual curiosity even from a very young age; even though my role in science involves fewer test tubes and bubbling liquids than I expected back then.

I am lucky to have many friends who have advised me and guided me through the opening stages of my academic journey. Thank you Professor Doug Dickey (UVM) and Professor Jeff Dinitz (UVM) for first pointing me towards academia during an extremely uncertain time in my life when I was unsure that I would even complete a bachelor degree. I am forever grateful that you started me on this path and for your continued support of my adventures since then. You saw in me the potential to love science.

I am immensely grateful to my masters advisers, Professor Chris Danforth (UVM) and Professor Peter Dodds (UVM). Thank you for allowing my forced entry into the Computational Story Lab. My time with you showed me what academia is all about, and your guidance led me to share your love for science. Further, and more importantly, you both taught me to pursue a full and complete life in addition to scientific pursuits; these lessons have proven to be an essential asset and advantage over my peers in my adventures since leaving home. Because of you, I know how to do science and have fun.

Thank you to Professor Iyad Rahwan (MIT, Max Planck Institute) for your invaluable guidance through my PhD. You taught me to dream bigger and to ask bigger questions; these lessons have forever altered how I see science and my role in contributing to it. Especially, thank you for teaching me to engage with new academic communities and to constantly consider the larger picture as I pursue bodies of work over one-off projects. Finally, thank you for your unwavering support of all things not academic during my time as your student; my family is extremely grateful for your flexibility, support, and trust in me.

Finally, I want to thank my loving wife, Catherine Westbom. Thank you for letting me drag

you across the country and back to "try-out" an industry job. You give me the daily support I rely on to pursue my goals. You make me feel like anything is possible and that I can dream as big, or as small, as I want. You make me laugh when I am too serious, and you make me focus when I am too distracted. You do all of these things even while you pursue your own daring ambitions. I am constantly in awe of you and your goals, and I am lucky that you are on my side while I pursue my career.

Contents

THIS PAGE INTENTIONALLY LEFT BLANK

List of Figures

THIS PAGE INTENTIONALLY LEFT BLANK

List of Tables

Chapter 1

Introduction

1.1 Motivation

Although the technology of the Industrial Revolution expanded the capability of unskilled labor, artificial intelligence (AI) and other modern technology may augment cognitive work while substituting for routine physical work (e.g. robotics in manufacturing). In turn, this divide among skill sets threatens to exacerbate growing wealth disparity and job polarization already occurring in the US. To highlight this, consider two burgeoning AI technologies: autonomous vehicles and machine learning. On the one hand, experts expect autonomous vehicles will diminish employment for truck drivers, which is the most common job among workers in the US according to the US Bureau of Labor Statistics. On the other hand, the prevalence of the technology industry and demand for programmers have increased with the maturation of machine learning techniques and enriched data sets. Why do these two technologies impact their relevant workers so differently? Generalizing from this comparison, automation research is dominated by contrasting perspectives: will AI automate US employment, or, to the contrary, will the efficiency gains of AI lead to new employment, wealth, and opportunity for all?

As AI technologies reach *general purpose* status, they have the potential to reshape labor demands, career opportunities, and the distribution of employment in the US. Yet researchers and policy makers maintain competing forecasts of the labor trends that result from each new wave of technology. Echoing concerns from the past, some decry technology for its deleterious effect on employment for unskilled workers, while others praise increasing productivity of skilled workers with labor substitution as a necessary side-effect to "creative destruction." It seems that the competing narratives of technological substitution and augmentation are *both* true for different subsets of workers; so how can we weave these elements together into a single cohesive framework? And can we efficiently identify the specific features that determine the type of exposure a worker will experience? Dear reader, are you *skilled* or *unskilled*?

1.2 Statement of Problem

Existing literature on technological change and the future of work highlights the crucial role of workplace tasks and skills. Occupations are best understood from their constituent skill requirements, and technology alters the demand for human labor capable of specific tasks or skills. Accordingly, analytical explanations of technological change and labor trends often rely on distinctions between skill sets as a divide between cognitive workers and physical workers.

These coarse labor categories have indeed proven useful, but there may be further insight to be gained if we refine our focus on workplace skills. In other words, relying only on coarse labor categories may leave important features unresolved. If we instead focus on granular skill categories, then it becomes increasingly difficult to rely on analytical analysis alone. Therefore, from an empirical and data-driven perspective, how can we investigate the role of workplace skills in explaining exposure to technological automation and the opportunities available to workers based on their skills and abilities? Generalizing further, based on the interdependencies between workplace skills, where do bottlenecks exist to career mobility? As cities increasingly act as centers for innovation and economic productivity in the US, how do skills and employment relate to urbanization and the economic well-being of cities?

1.3 Conceptual Framework for Study

In this dissertation, I argue that we should refine our analysis down to the atomic units of labor: workplace tasks and skills. This idea builds on existing concepts; foundational work on job matching theory [156] highlights the importance of *skill matching* when workers seek to fulfill the skill requirements of job vacancies. Since then, labor economists have suggested that occupational job titles are best understood as abstract bundles of task and skills requirements [24, 136]. Leading descriptions of US job polarization (e.g. the "hollowing of the middle class") rely on coarse labor distinctions between high-skill *cognitive* and low-skill *physical* workers and distinctions between routine and non-routine work [7]. In chapter 2, I provide a detailed literature review of the competing narratives of technological change and discuss specific areas where refined research of workplace skills could improve models of the future of work.

Why have researchers restricted themselves to coarse labor categories? Relying on simple labor categories makes analytical descriptions of labor trends tractable but can hinder predictive modelling and empirical analysis by leaving important factors unresolved. To advance, we must refine the resolution with which we study types of labor to observe specific workplace tasks and skills. Interdependencies between these atomic units of labor (e.g. skill complementarity) undergird several aggregated labor dynamics that we study today, including career mobility and worker reskilling or up-skilling. Combined with heterogeneous employment across different regions and cities, these connections between skill sets have the potential to describe the labor trends that shape the entire US labor system.

Further, these refinements will enable researchers to consider technological change with improved precision and specificity by relating specific examples of technology to specific alterations in skill demand. Technology is often designed to perform a specific task (e.g. consider a robot arm on a manufacturing line) or to solve a specific class of problems (e.g. different machine learning algorithms and the "No Free Lunch" theorem [217]). Accordingly, these new labor studies may connect the microscopic impact of technology to macroscopic labor trends, such as job polarization. In chapter 3, I present an example of this where the specific skill requirements of US job titles grant new insights into the differential impact of automation across US cities.

The observations in chapter 3 motivate a deeper investigation into specific workplace skills and their role in aggregate job polarization. Therefore, I examine the patterns of co-occurrence of skills pairs in chapter 4. Surprisingly, our unsupervised data-driven analysis of skills reveals a high degree of skill polarization that respects key features of the so-called "hollowing of the middle class." Firstly, this analysis offers an empirical validation of popular analytical descriptions of US job polarization. However, this analysis allows me to go a step further by using the specific connections between skills to describe previously unresolved labor dynamics, such as worker transition rates between job titles based on skill requirements.

Since workers move between employment opportunities with similar skill sets, I further leverage granular skill requirements to relate pairs of US job titles in chapter 5 as a map to worker career mobility. Again, this unsupervised data-driven analysis of skills and jobs provides an accurate description of US job polarization. However, in combination with urban employment distributions, I demonstrate how skills data can also model the spatial mobility of workers in addition to their career mobility between job titles. In line with existing anecdotal evidence, this analysis shows how skills and employment can inform spatial mobility models.

These analyses demonstrate a new framework for labor, AI, and the connectivity of specific types of labor. My multidisciplinary approach is enabled by recently available sources of skills data in combination with tools from network science and complex systems. Labor economics lays the foundation for this investigation and provides powerful methods to describe labor systems at equilibrium. However, a world with constant technological change is never really at equilibrium. Fortunately, the data-driven analysis offered in this dissertation may help to describe out-of-equilibrium behaviors by connecting microscopic elements of the US labor system to macroscopic labor trends.

In particular, I believe the insights laid out in this document describe pieces of a new pipeline for forecasting the future of work (see chapter 2). In addition to new models for aggregate labor trends, models that incorporate granular workplace skills enable us to consider new sources of input data. In the final chapter 6, I examine one such data source: the evolution of AI research itself. Understanding the trajectory of future AI technologies from active research has the potential to inform models of technological change, innovation, and the future of work.

1.4 Research Questions

Many major US labor trends have been described analytically and with some empirical investigation. However, absent from the literature is a connection between these macroscopic labor trends and the microscopic workplace tasks and skills that differentiate job titles and individual workers. In this thesis, I demonstrate empirical analysis to address the following questions on employment, mobility, and the future of work:

1. Is the expected impact of automation equal across cities?
2. What factors, including differences in employment and differences in supported skill sets, predict the differential impact of automation across US cities?
3. Do interdependencies between workplace skills predict aggregate labor trends and, in particular, job polarization in the US?
4. Do skill interdependencies predict workers' transitions between job titles?
5. Does the inclusion of skills data and employment data improve predictions of spatial mobility between US cities?
6. Are social scientists–the researchers who study social and societal dynamics–keeping pace with AI research?
7. Is industry's role growing within AI research?

To answer questions 1 and 2, I combine existing automation estimates with employment distributions to assess the expected impact of automation in each city in the foreseeable future. I then compare this measure to other features describing the employment distribution in cities. Additionally, I consider the distribution of supported skill sets in each city.

To answer questions 3 and 4, I leverage skills data from the US Bureau of Labor Statistics to identify co-occurring skill requirements across job titles. I represent these pairwise skill relationships as a network and compare the topology of this network to patterns of career mobility and aggregate labor trends. To answer question 5, I project these skill relationships into a job network that relates job titles according to their shared skill requirements and relate the topology of the job network to spatial mobility patterns between US cities.

To answer questions 6 and 7, I perform a bibliometric analysis of academic publications from the last 60 years to study which academic fields have been influential in AI research, and vice versa. In particular, I identify patterns within the AI research community that may contribute to a growing gap between AI research and the social sciences.

1.5 Significance of the Study

This study offers several unsupervised data-driven analyses of the impact of technological change, the factors that contribute to workers' career mobility, and the differences in urban labor markets. These methods and results bolster and complement existing analytically-driven analyses in the literature. Further, this study opens up several future directions of inquiry:

- Which skills are exposed to technology?
- Can we predict occupational skill redefinition over time?
- Similar to ecological resilience, can we measure the connectivity between jobs and/or skills within a labor market to infer its economic resilience?
- How can we incorporate real-time data reflecting the changing demand for workplace skills (e.g. online job postings and worker profiles) to improve the resolution of forecasts into the future of work?

1.6 Limitations

1.6.1 Causally relating skills to career success

Although this study identifies several ways to connect granular workplace skills to aggregate labor trends, this study does not focus on causal identification. Rather, much of the goal in this study is to demonstrate a proof of concept: workplace skills and their interdependencies matter and improve predictions of the labor dynamics we care about, such as workers' spatial and career mobility. Seeing that certain skill sets relate to increased mobility may identify potential causal relationships, but it does not confirm them.

Resolving these causal dependencies represents an extremely valuable direction for future work with the potential to further inform policy decisions. For examplc, imagine being a policy maker attempting to design a retraining program for workers in your city. Your goal is to give truck drivers, who may see depressed employment with the adoption of autonomous vehicles, the skills they need to obtain stable employment. The analysis reported in this study significantly informs this effort by highlighting skills that may be more attainable to a worker based on their existing skills and abilities and, in particular, identifying skills that are "far away." For instance, although demand for software developers may be increasing, teaching truck drivers to program may not offer a scalable solution if the other skills required by software developers are not immediately obtainable based truck drivers' skills. While this study significantly informs this effort, it does not completely identify which skills actually cause a worker to succeed as a software developer.

1.6.2 Which skills will be automated by technology?

This analysis shows the insights granted by the interdependencies between specific workplace skills, but it does not offer a forecast into the automatability of specific skills. Nor does this study tackle the huge problem of forecasting occupational skill redefinition with changing technology. However, this is an area of future work that is enabled by the focus on specific workplace skills offered in this analysis. Next steps in this direction include identifying which skills produce augmenting or substitution effects to workers when demand for those skills is impacted by automation.

If skill demands change, then how does the skill requirements of a given job title change in response (e.g. see the story of ATMs and bank tellers in chapter 2)?

1.6.3 How is technology changing over time?

This question is immensely import to forecasts of the future of work based on technological change, but this task is made difficult by the diverse examples of new technology we see today. Accordingly, this study does not tackle this problem. However, this study does offer methods and results from a labor perspective that enables researchers to effectively leverage this type of insight once it is gained.

Chapter 2

A Framework for Studying the Future of Work

Note: This chapter summarizes thoughts on forecasting the future of work presented in [94].

2.1 A history of automation and the rise of Artificial Intelligence

Rapid advances in artificial intelligence (AI) and automation technologies have the potential to significantly disrupt labor markets. While AI and automation can augment the productivity of some workers, they can replace the work done by others, and will likely transform almost all occupations at least to some degree. Rising automation is happening in a period of growing economic inequality, raising fears of mass technological unemployment and a renewed call for policy efforts to address the consequences of technological change.

Several barriers inhibit scientists from measuring the effects of AI and automation on the future of work. These barriers include the lack of high quality data about the nature of work (e.g. the dynamic requirements of occupations), lack of empirically-informed models of key micro-level processes (e.g. skill substitution, human-machine complementarity), and insufficient understanding of how cognitive technologies interact with broader economic dynamics and institutional mechanisms (e.g. urban migration, international trade policy). Overcoming these barriers requires improvements in the longitudinal and spatial resolution of data, as well as refinements to data on workplace skills. These improvements will enable multidisciplinary research to quantitatively monitor and predict the complex evolution of work in tandem with technological progress.

Artificial Intelligence is a rapidly advancing form of technology with the potential to drastically reshape U.S. employment [61, 152]. Unlike previous technologies, examples of AI have applications in a variety of highly-educated, well-paid, and predominantly-urban industries [59], including medicine [30, 74], finance [80], and information technology [216]. With AI's potential to change the nature of work, how can policy makers facilitate the next generation of employment opportunities? Studying this question is made difficult by the complexity of economic systems and AI's differential impact on different types of labor.

While technology generally increases productivity, AI may diminish some of today's valuable employment opportunities. Consequently, researchers and policy makers worry about the future of work in both advanced and developing economies worldwide. As an example, China is making AI-driven technology the centerpiece it's economic development plan [112]. Automation concerns are not new to AI, and examples date back even to the advent of written language. In ancient Greece (*circa* 370 B.C.), Plato's *Phaedrus* [221] described how writing would displace human memory and reading would substitute true *knowledge* with mere *data*. More commonly, historians point to the Industrial Revolution and the riots of 19th century Luddites [127] as examples where technological advancement led to social unrest. Examples from the recent past echo these concerns:

"Labor will become less and less important... More workers will be replaced by machines. I do not see that new industries can employ everybody who wants a job"
- Wassily Leontief (1952) [141], winner of the 1973 Nobel Prize in Economics

"In the past, new industries hired far more people than those they put out of business. But this is not true of many of today's new industries... Today's new industries have comparatively few jobs for the unskilled or semiskilled, just the class of workers whose jobs are being eliminated by automation."

- The Automation Jobless, Time Magazine (1961)

"Automation provides us with wondrous increases of production and information, but does it tell us what to do with the men the machines displace? Modern industry gives us the capacity for unparalleled wealth - but where is our capacity to make that wealth meaningful to the poor of every nation?"

- U.S. Attorney General Robert F. Kennedy (1964) [128]

And yet, despite these longlasting and oft-recurring concerns, society underwent profound transformations, the economy continued to grow, technology continues to advance, and workers continue to have jobs. Given this history of concern, what makes human labor resilient to automation? And is AI a fundamentally new concern from technologies of the past?

Answering these questions requires a detailed knowledge that connects AI to today's workplace skills. Each specific technology alters the demand for specific types of labor, and, thus, the varying skill requirements of different job titles can obfuscate technology's impact. In general, depending on the nature of the job, a worker may be augmented by technology or in competition with it [24, 56, 57]. For example, technological advancements in robotics can diminish wages and employment opportunities for manufacturing workers [8, 18]. However, technological change does not necessarily produce unemployment, and, in the case of AI, cognitive technology may actually augment workers [24, 44]. For instance, machine learning appears to bolster the productivity of software developers while also creating new investment and manufacturing opportunities (e.g. autonomous vehicles). Complicating matters further, the skill requirements of occupations do not remain static, but instead change with changing technology [44, 78].

In the remainder of this chapter, I describe how these competing dynamics combined with insufficient data might allow contrasting perspectives to coexist. In particular, we argue that the limitations into data about workplace tasks and skills restricts the viable approaches to the problem of technological change and the future of work. I offer suggestions to improve data collection with the goal of enriching models for workplace skills, employment, and the impact of AI. Finally, I suggest new insights that improved data could provide in combination with a new methodological focus on resilience and forecasting.

2.2 Contrasting Perspectives on the Future of Work

2.2.1 Doomsayer's Perspective

Technology improves to make human labor more efficient, but large improvements may yield deleterious effects for employment. This obsoletion through labor substitution leads many to worry about "technological unemployment"[1] and motivates efforts to forecast AI's impact of jobs. One study assessed recent developments in AI to conclude that 47% of current U.S. employment is at high risk of computerization [98], while a contrasting study, employing a different methodology, concluded that a less alarming 9% of employment is at risk [16]. Similar studies have looked at the impact of automation on employment in other countries and reached sobering conclusions: automation will affect 35% of employment in Finland [170], 59% of employment in Germany [62], and 45% to 60% of employment across Europe [54]. Critics have complained that prospective studies lack validation, but retrospective studies also find that robotics are diminishing employment opportunities in U.S. manufacturing [8, 43] (although, not in Germany [75]).

2.2.2 Optimist's Perspective

Optimists suggest that technology may substitute for some types of labor but that efficiency gains from technological augmentation outweighs transition costs [26, 87, 97, 123, 160], and, in many cases, technology increases employment for workers who are in not direct competition with it [21, 44] (although recent follow-up work suggests these are temporary gains [43]). Furthermore, the skill requirements of each job title are not static and actually evolve over time to reflect evolving labor needs. For example, workers may require more social skills because those skills remain difficult to automate [78]. Even if technology depresses employment for some types of labor, it can create new needs and new opportunities through "creative destruction" [9, 31]. For instance, the replacement of equestrian travel with automobiles spurred demand for new roadside amenities,

[1] *"We are being afflicted with a new disease of which some readers may not yet have heard the name, but of which they will hear a great deal in the years to come–namely, technological unemployment. This means unemployment due to our discovery of means of economising the use of labour outrunning the pace at which we can find new uses for labour."* - John Maynard Keynes

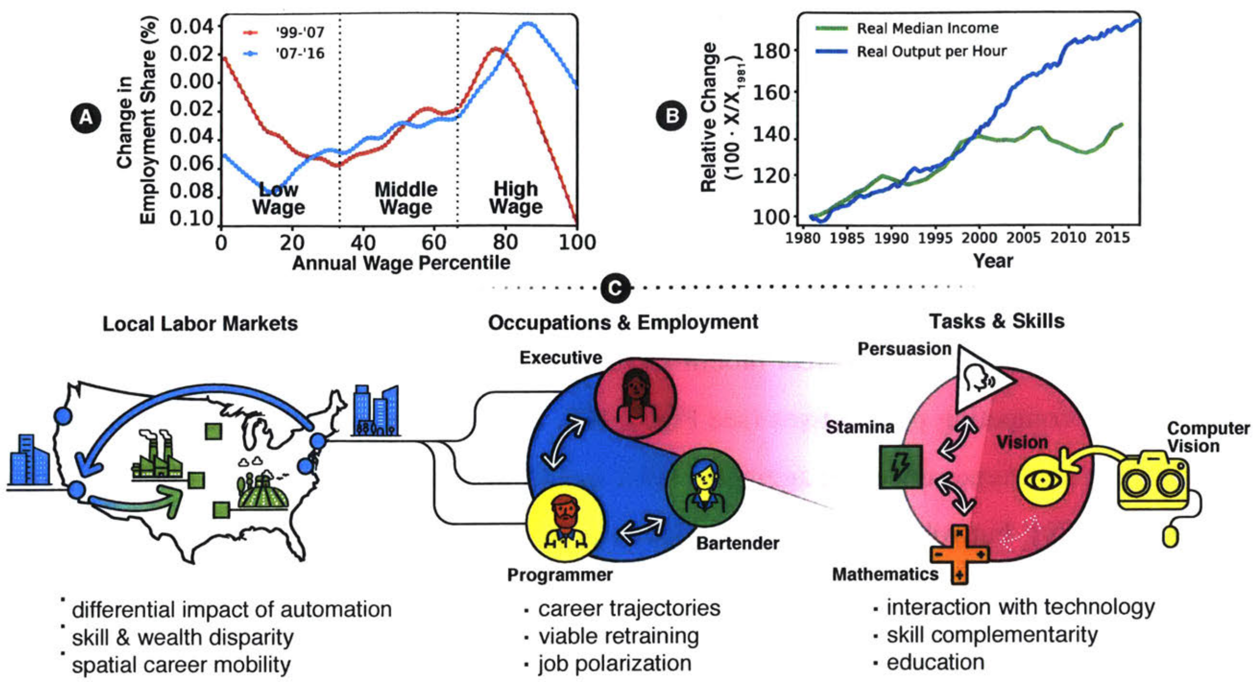


Figure 2-1: Motivating and describing a framework to study the technologies impact on workplace skills. **(A)** Following [7], we use ACS national employment statistics to compare the change in employment share (y-axis) of occupations according to their average annual wage (x-axis) during two time periods. Employment share is increasing for low- and high-wage occupations at the expense of middle-wage occupations. **(B)** Following [57], we use data from the Federal Reserve Bank of St. Louis to compare U.S. productivity (real output per hour) and workers' income (real median personal income), which have traditionally grown in tandem. The efficiency gains of automating technologies are thought to contribute to this so called "Great Decoupling" starting around the year 2000. **(C)** A framework for studying technological change, workplace skills, and the future of work as multi-layered network. (Left) Cities and rural areas represent separate labor markets, but workers and goods can flow between them. (Middle) Each location can be represented as an employment distribution across occupations. Connections between occupations in a labor market represent viable job transitions. Job transitions are viable if workers of one job can meet the skill requirements of another job (i.e. "skill matching" [156]). (Right) Workers' varying skill sets represent bundles of workplace skills that tend to be valuable together. Skill pairs that tend to co-occur may identify paths to career mobility. Technology alters demand for specific workplace skills, thus altering the connections between skill pairs. As an example, machine vision software may impact the demand for human labor for some visual task. These alterations can accumulate and diffuse throughout the entire system as aggregate labor trends described in (A) & (B).

such as motels, gas stations, and fast food [125].

2.2.3 Unifying Competing Viewpoints

On one hand, multiple dynamics accompany technological change and create uncertainty about the future of work. On the other hand, experts agree that occupations are best understood as abstract bundles of skills [24, 136] and that technology directly impacts demand for specific skills instead of acting on whole occupations all at once [18, 21, 44, 142]. Therefore, a detailed framework that connects specific skill types to career mobility [10, 24] and to whole urban workforces [136] may help to unify competing perspectives (see Fig. 2-1C). Existing studies have argued theoretically that different skill types underpin aggregate labor trends, such as job polarization [18] and urban migration [1, 89], but robust empirical validation is made difficult by the specificity of modern skills data and its temporal sparsity.

2.3 Barriers to Forecasting the Future of Work

In this section, we identify barriers to our scientific modelling of technological change and the future of work. Along with each barrier, we propose a potential solution that could enable improvement in forecasting labor trends. We provide a summary of these barriers and solutions in Table 2.1.

2.3.1 Barrier: Sparse Skills Data

Forecasting automation from AI requires skills data that keeps pace with rapidly advancing technology (e.g. Moore's Law [203], robots in manufacturing [8], and patent production [207, 208, 219]). While skill types inform the theory of labor and technological change [7, 24, 152, 214], standard labor data focuses on aggregate statistics, such as wage and employment numbers, and can lack resolution into the specifics that distinguish different job titles and different types of work. For example, previous studies have empirically observed a "hollowing" of the middle skill jobs described by increasing employment share for *low-skill* and *high-skill* occupations at the expense

BARRIER	POTENTIAL SOLUTION
Sparse Skills Data	• adaptive skill taxonomies • connect susceptiple skills to new technology • improve temporal resolution of data collection • use data from career web platforms
Limited Modeling of Resilience	• explore out-of-equilibrium dynamics • identify workplace skill inter-dependencies • connect skill relationships to worker mobility • relate worker mobility to economic resilience in cities • explore models of resilience from other academic domains
Places in Isolation	• labor dependencies between places (e.g. cities) • identify skillsets of local economies • idenitfy heterogeneous impact of technology across places • use inter-city connections to study national economic resilience

Table 2.1: Tabulating the current barriers to forecasting the future of work along with proposed solutions.

of *middle-skill* occupations [18, 21] (reproduced in Fig. 2-1A). These studies use skills to explain labor trends, but are limited empirically to measuring annual wages instead of skill content directly. While wages may correlate with specific skills, wage alone fails to capture the defining features of an occupation, and models focused on only cognitive and physical labor fail to explain responses to technological change [7].

As another approach, data on educational requirements can add resolution to employment trends [19, 23, 107]. For instance, jobs that require a bachelor's degree may identify cognitive workers who are less susceptible to automation. Ideally, educational institutions train workers to possess valuable skills that lead to higher wages [129]. However, looking at education and wages alone have proven insufficient to explain stagnating returns on education [18,34,69] and slow wage growth despite increases in national productivity [56, 57, 142] (see Fig. 2-1B).

Improving data on the skills required to perform specific job tasks may provide better insights than wages and education alone. For example, previous studies have considered occupations as routine or non-routine and cognitive or physical [7, 13, 49, 81, 91, 120, 121, 150, 204], or looked at specific types of skills in relation to augmentation and substitution from technology [24, 142].

Increasing a labor model's specificity into workplace tasks and skills might further resolve labor trends and improve predictions of automation from AI. As an example, consider that civil engineers and medical doctors are both high-wage cognitive non-routine occupations requiring many years of higher education and additional professional certification. Yet, these occupations require distinct workplace skills that are largely non-transferable, and these occupations are likely to interact with different technologies. Wages and education—and even aggregations of workplace skills—may be too coarse to distinguish occupations and, thus, may obfuscate the differential impact of various technologies and complicate predictions of changing skill requirements. In turn, these shortcomings may help explain the variability in current automation predictions (see section 2.2).

2.3.2 Solutions to Improve Skills Data

While publicly available skills data are limited, the U.S. Department of Labor's O*NET database has seen recent use in labor research (e.g. [95, 98, 142]). O*NET offers many benefits including a detailed taxonomy of skills and more regular updates than preceding datasets. In 2014, O*NET began to receive partial updates twice annually, which is a considerable improvement on the Dictionary of Occupational Titles, which was published in four editions in 1939, 1949, 1965, and 1977 with a revision in 1991. However, employment trends and changing demand for specific tasks and skills might change faster than O*NET's temporal resolution and skill categorization can capture. Complicating matters further, advances in AI and machine learning may be changing the nature of automation thereby altering the types of tasks that are affected by technology [59, 60].

Furthermore, studies often use O*NET data to construct aggregations of skills, such as Information Input or Mental Processes [136], rather than focusing on skills at the their most granular level. Methodological choices aside, O*NET's relatively static skill taxonomy poses its own problems as well. For instance, according to O*NET, the skill "Installation" is equally important to both Computer Programmers and to Plumbers, but, undoubtedly, workers of these occupations are performing very dissimilar tasks when they are *installing* things on the job (see Fig. 2-2A and SI Section 1 for calculation). More generally, any static taxonomy for workplace skills is not ideal for a changing economy: should Mathematics and Programming be two separate workplace skills

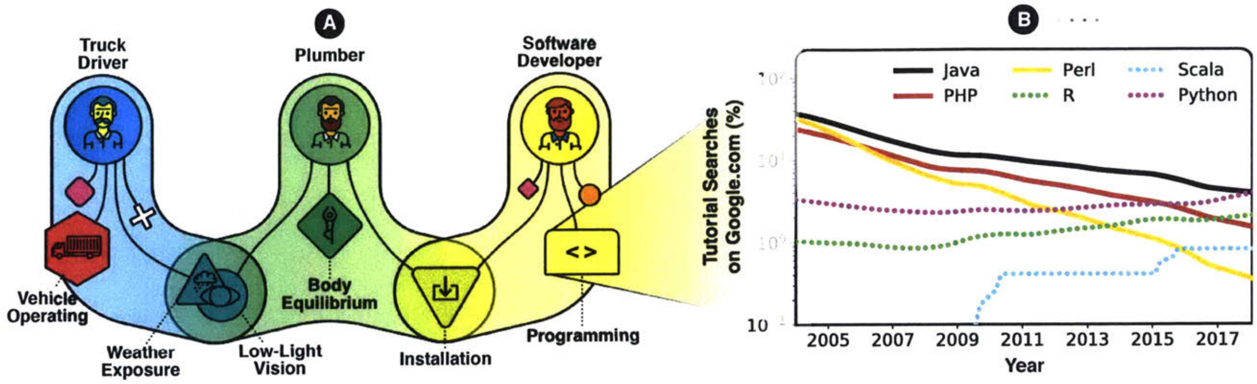


Figure 2-2: Since the skill requirements of occupations may inform opportunities for career mobility, abstract skill data may obfuscate important labor trends. **(A)** We use O*NET data to identify the characteristic skill requirements for Truck Drivers, Plumbers, and Software Developers (see section A.1 for calculation). Individual skills may be unique to an occupation (e.g. Operating Vehicles) or shared between occupations (e.g. Low-Light Vision). The skill Installation is required by both Plumbers and Software Developers, but this skill may not mean the same thing to workers in these two occupations. Programming is a skill required by Software Developers, but the coarseness of this skill definition may hide important dynamics brought on by new technology, including AI. **(B)** For example, we provide the percentage of Google searches for coding tutorials by programming language. Trends are smoothed using locally weighted scatter plot smoothing (see section A.2 for calculation). The Python programming language is widespread in the field of machine learning. Therefore, the increased ubiquity of AI and, in particular, machine learning may contribute to Python's steady growth in popularity.

given they are both computational? Or, on the other hand, is Programming too broad given the variety of existing software and programming languages? Perhaps it is more appropriate to specify programming tasks or specific programming languages (see Fig. 2-2B for example), especially given the rapid development of AI and machine learning. Likely, the correct abstraction is situation dependent, but O*NET data offer limited flexibility.

Granular skills data will help elucidate the micro-scale impact of AI and other technologies in labor systems. For instance, the specifications of recent patents might suggest automatable types of labor in the near future [207, 208, 219], thus elucidating the impact of technological change at the granularity of workplace specific tasks and skills. The distribution of skill categories within occupations and over individuals' careers can reveal how occupational skill requirements evolve. As an example, consider that occupations, such as Software Developer, dynamically change the skill requirements in job listings (e.g. "programming" in the 1990s vs. "python," "Java," "kubernetes," etc. today) to reflect the tools and required specialization of the time. Understanding the dynamics of specific skills combined with the incomes within occupations can capture the marginal value of different skills despite the dynamic nature of work.

Online career platforms offer an example of the empirical possibilities facilitated by non-traditional and new data sources. These websites collect real-time data that reflect labor dynamics in certain industries. Data from workers' resumes can improve our understanding of education and careers, as well as identifying workers' transitions between occupations and skill sets. Additionally, job postings capture fluctuations in labor demands and demonstrate changes in demand for specific skills. Combined, these two sources of skills data offer an adaptive granular view into the changing nature of work that may detail where labor disconnects exist. Access to these private data sources is currently restricted and typically requires a data sharing agreement that protects personally identifiable information and other proprietary information. Of course, personal privacy and issues of representative sampling are inherent to these data, but increased access could meaningfully augment currently available open data on employment and workplace skills. One potential solution is to construct a secure environment for the sharing of detailed skills and career data that is similar to the recent Social Science One partnership (see `https://socialscience.one`).

2.3.3 Barrier: Limited Modeling of Resilience

Recent studies show that historical technology-driven trends may not capture the AI-driven trends we face today. Consequently, some have concluded that AI is a fundamentally new technology [59, 60]. If the trends of the past are not predictive of the employment trends from current or future technologies, then how can policy makers maintain and create new employment opportunities in the face of AI? What features of a labor market lead to generalized labor resilience to technological change?

It is difficult to construct resilient labor markets because of the uncertainty around technology's impact on labor. For instance, designing viable worker retraining programs requires detailed knowledge of the local workforce, fluency with current technology, and an understanding of the complex dependencies between regional labor markets around the world [213, 220]. Technology typically performs specific tasks and may alter demand for specific workplace skills as a result. These micro-scale changes to skill demand can accumulate into systemic labor trends including occupational skill redefinition, employment redistribution (e.g. job creation and technological unemployment), and geographic redistribution (e.g. worker migration). Forecasting these complex effects requires a detailed understanding of the pathways along which these dynamics occur.

As an emblematic example of these complex dynamics, consider the competition between human bank tellers and automated teller machines (ATMs) (described in [44]). Unexpectedly, national employment for bank tellers rose with the adoption of ATMs. One explanation is demand elasticity: as ATMs decreased the operating cost of bank branches, more bank branches opened nationwide to meet rising consumer demand. Another more complicated reason is the accompanying shift in fundamental skill requirements from clerical ability to social and persuasive skills used by salespeople and customer service representatives. The story of bank tellers and ATMs is only fully captured by connecting the job-level changes in occupational skill composition with the system-level dynamics of demand brought on by increased efficiency. Accordingly, an updated framework for labor and AI must capture the interactions of microscopic workplace skills in combination to produce macroscopic labor trends, such as employment shifts, job polarization, and workers' spatial mobility (for example, see Fig. 2-3B).

2.3.4 Solutions for Fostering Resilient Labor

Existing theory of the matching process between job seekers and job vacancies [156] provides a stylized description of the matching process that lacks resolution into skill demand. Mapping the space of skill inter-dependencies (e.g. Fig. 2-1D) could inform training and job assistance programs by identifying which types of work—and which locations—may experience augmentation and/or substitution with new technology. The detailed skill requirements of occupations determine the career mobility of individual workers, thus, changes to the demand for certain skills have the potential to redefine viable career trajectories and worker flow between occupations (e.g. middle layer of Fig. 2-1D). Therefore, mapping the relationships between jobs and skills that produce

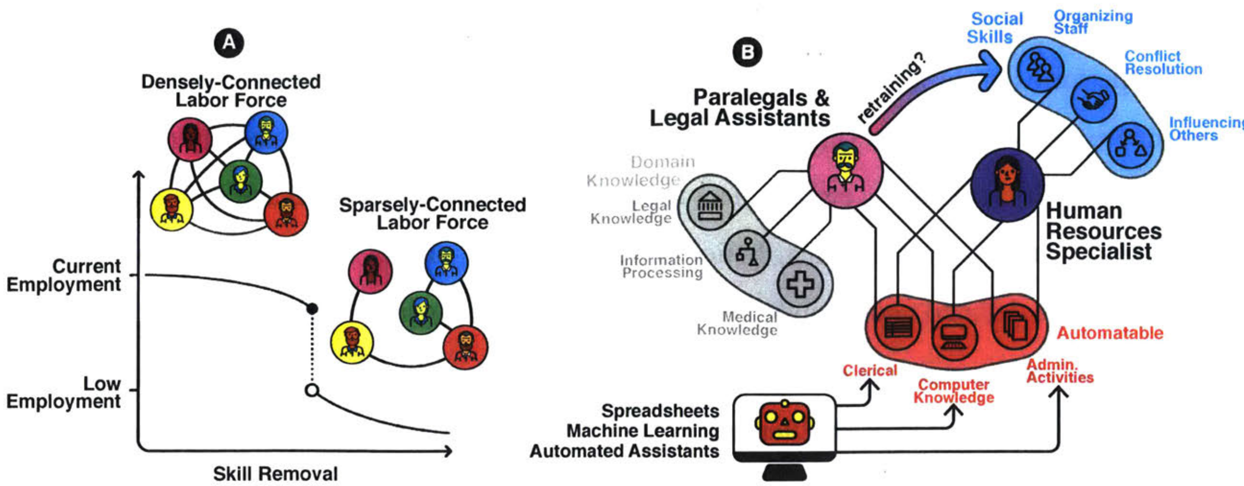


Figure 2-3: Skill complementarity may define the structural resilience of a workforce and inform worker retraining programs. **(A)** As in climatology and ecology, the structural pathways constraining labor dynamics could determine the resilience of a labor market to changing labor skill demands. In this example, we connect occupation pairs with high skill similarity because skill similarity might indicate easier worker transitions between job titles. Borrowing from research on ecological systems [100], the density of connections between occupations could determine "tipping points" for aggregate employment in cities. **(B)** With recent concerns of automation [76, 199], which jobs might be suitable for Paralegals and Legal Assistants if employment for these jobs diminishes? Better resolution into skill requirements could help identify occupations that rely on similar skills, but also rely on skills that are removed from competition with technology. In this example, we identify characteristic skills using the O*NET database to find that Paralegals rely on many shared workplace skills with Human Resource Specialists. Human Resource Specialists rely on social skills, which are not easily automated [78]. See section A.1 for skill calculations.

employment opportunities is a vital step for policy makers in the face of technological change.

In related domains, tools from network science have already provided new insights into modeling (and minimizing) systemic risk in global credit [190] and financial industries [33], forecasting the future exports of national economies [111, 115, 117], mapping worker flows between industries [163] and firms [109], and charting the changing industrial composition of cities [158, 192, 194]. Therefore, identifying the pathways along which labor dynamics (e.g. how skills determine workers' career mobility) occur may provide similarly useful insights into the impact of AI on labor. Similar methods have been used to measure ecological resilience based on the structure of mutualistic inter-species interactions [100, 155]. These methods often rely on the size and density of inter-connected entities to estimate systemic resilience to species removal—perhaps analogous to diminishing demand for a skill with new technology (e.g. Fig. 2-3A).

Mapping skill dependencies will require appropriate data handling methods. The ideal skills data should reflect the dynamic nature of skill representation, and so, the methods we use to detect, categorize, and measure the demand for skills must be adaptable as well. Perhaps ironically, advanced AI techniques may help. Tools from machine learning (ML) and natural language processing (NLP) may capture the latent structure in complex high-dimensional data, thus making them ideal tools for the proposed application (and other applications in econometrics [157]). For example, NLP may be used to process historical skills data from the Dictionary of Occupational Titles into a format akin to the modern O*NET data. ML can be used on longitudinal job postings data to identify trends in skill demands that may reflect changes in technological ability. Combining these modern computational methods with relevant sources of data may foster new insights into labor dynamics at a high temporal resolution. In turn, these methodological improvements can bolster labor forecasts and policy makers' ability to respond to real-time labor trends.

2.3.5 Barrier: Places in Isolation

The impact of AI and automation will vary greatly across geography, which has implications for the labor force, urban-rural discrepancies, and changes in the income distribution [20]. The study of AI and automation are largely focused on national employment trends and national wealth disparity. However, recent work demonstrates that some places (e.g. cities) are more susceptible to technological change than others [8, 95]. Occupations form a network of dependencies which constrain how easily jobs can be replaced by technology [192, 196]. Therefore, the health of the aggregate labor market may depend on the impact of technology on specific urban and rural labor markets.

Although technological change alters demand for specific workplace tasks and skills, current skills data mask the specific skill sets that comprise and differentiate the workforces of different geographies. In part, this is because skills data from nationwide surveys, such as the O*NET database, average over the regional variability in the required skills of workers with shared job titles. For example, software developers seeking employment in Silicon Valley may need to advertise more specific skill sets than similar employees in a shallower labor market (following the division of labor theory). Exacerbating this trend, the same AI-technologies that augment high-wage cognitive employment are more abundant in large cities, while the physical low-wage tasks that are most readily replaced by robotics are more abundant in small cities and rural communities. This observation suggests that national wealth disparity is reflected in the wealth disparity between large and small cities akin to wage inequality across individuals.

2.3.6 Solutions for Studying Technology's Impact on Spatial Dependencies

Improved models for spatial inter-dependencies require more granular skills data (discussed above) and new insights into the mechanisms that create today's cross-sectional geographic trends. For instance, how do university towns, where people gain valuable cognitive skills, contribute to the productivity of large cities? Do these economic connections help explain why university towns perform surprisingly well compared to similarly sized cities according to socio-economic indicators (including exposure to automation [95])?

Furthermore, just as internal connectivity determines urban economic resilience [194], so too can the connections between U.S. cities underpin the economic health of the national economy [208]. For instance, an interruption in the supply chain of well-educated cognitive workers may stifle an urban economy that normally attracts skilled workers. Therefore, it behooves policy makers to understand the connections between their local labor market and other urban labor markets in order to assess the resilience of their local economy. Since employment opportunities are central in people's decision to relocate [1] and skill matching is essential to the job matching process [156], understanding the constituent skill sets in cities can inform models for the spatial mobility of workers and improve our understanding of career mobility and career incentives.

2.4 Technological change and an improved data-pipeline

Artificial intelligence has the potential to reshape skill demands, career opportunities, and the distribution of workers among industries and occupations in the U.S., and in other developed and

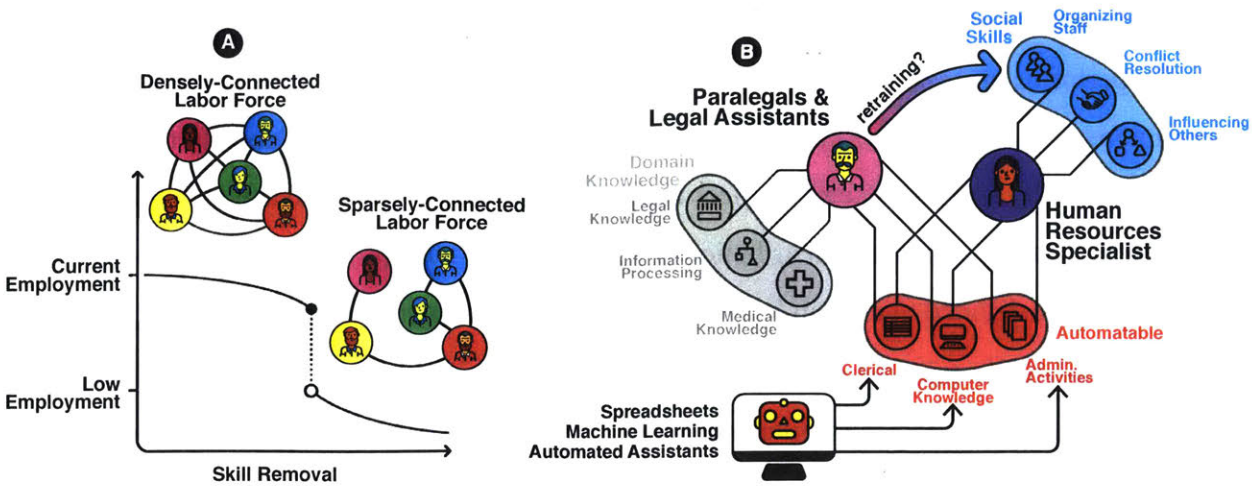


Figure 2-3: Skill complementarity may define the structural resilience of a workforce and inform worker retraining programs. **(A)** As in climatology and ecology, the structural pathways constraining labor dynamics could determine the resilience of a labor market to changing labor skill demands. In this example, we connect occupation pairs with high skill similarity because skill similarity might indicate easier worker transitions between job titles. Borrowing from research on ecological systems [100], the density of connections between occupations could determine "tipping points" for aggregate employment in cities. **(B)** With recent concerns of automation [76, 199], which jobs might be suitable for Paralegals and Legal Assistants if employment for these jobs diminishes? Better resolution into skill requirements could help identify occupations that rely on similar skills, but also rely on skills that are removed from competition with technology. In this example, we identify characteristic skills using the O*NET database to find that Paralegals rely on many shared workplace skills with Human Resource Specialists. Human Resource Specialists rely on social skills, which are not easily automated [78]. See section A.1 for skill calculations.

employment opportunities is a vital step for policy makers in the face of technological change.

In related domains, tools from network science have already provided new insights into modeling (and minimizing) systemic risk in global credit [190] and financial industries [33], forecasting the future exports of national economies [111, 115, 117], mapping worker flows between industries [163] and firms [109], and charting the changing industrial composition of cities [158, 192, 194]. Therefore, identifying the pathways along which labor dynamics (e.g. how skills determine workers' career mobility) occur may provide similarly useful insights into the impact of AI on labor. Similar methods have been used to measure ecological resilience based on the structure of mutualistic inter-species interactions [100, 155]. These methods often rely on the size and density of inter-connected entities to estimate systemic resilience to species removal—perhaps analogous cess to unstructured skills data from resumes and job postings along with new indicators for recent technological change (e.g. patent data) and models for both inter-city and intra-city labor dependencies will enable new and promising techniques for understanding and forecasting the future of work. This improved data collection will enable the use of new data-driven tools, including machine learning applications and systemic modeling that more accurately reflects the complexity of labor systems. New data will lead to new research that enriches our understanding of the impact of technology on modern labor markets.

Chapter 3

Small cities face greater automation

Note: This chapter summarizes findings from [95].

3.1 Motivation

The city has proven to be the most successful form of human agglomeration and provides wide employment opportunities for its dwellers. As advances in robotics and artificial intelligence revive concerns about the impact of automation on jobs, a question looms: How will automation affect employment in cities? Here, we provide a comparative picture of the impact of automation across U.S. urban areas. Small cities will undertake greater adjustments, such as worker displacement and job content substitutions. We demonstrate that large cities exhibit increased occupational and skill specialization due to increased abundance of managerial and technical professions. These occupations are not easily automatable, and, thus, reduce the potential impact of automation in large cities. Our results pass several robustness checks including potential errors in the estimation of occupational automation and sub-sampling of occupations. Our study provides the first empirical law connecting two societal forces: urban agglomeration and automation's impact on employment.

3.2 Background: Automation & Urbanization

Cities, which accommodate over half of the world's population [1], are modern society's hubs for economic productivity [17, 137, 153] and innovation [45,46,218]. Since job migration is the leading factor in urbanization [1, 184], policy makers are increasingly concerned about the impact of Artificial Intelligence and automation on employment in cities [8, 67, 167]. While researchers have investigated automation in national economies and individual employment, it remains unclear *a priori* how cities naturally respond to this threat. In a world struggling between localism and globalism, a question emerges: *How will different cities cope with automation?* Answering this question has implications on everything from urban migration to investment, and from social welfare policy to educational initiatives.

To construct a comparative picture of automation in cities, our first challenge is to get reliable estimates of how automation impacts workers. Existing estimates are wide ranging. Frey and Osborne [98] estimate that 47% of U.S. employment is at "high risk of computerization" in the foreseeable future, while an alternative OECD study concludes a more modest 9% of employment is at risk [16]. Note that these results do not tell us about the impact of automation in cities as they are presented at a national level. Differences in these predictions arise from discrepancies over two main skill dynamics: the substitution of routine skills, and complementarity of non-routine and communication skills [19, 24]. Additionally, technology-driven efficiency may redefine the skill requirements of occupations and actually increase employment in low-skilled jobs [42, 44].

Nevertheless, even if we take current estimates of the *absolute* risk of computerization of jobs with skepticism, these estimates can provide useful guidance about *relative* risk to different cities that is robust to errors in the estimates provided by [98] and [16]. We can interpret the 'risk of computerization' estimates as an educated guess about which occupations will experience greater adjustment due to machine substitution of a large portion of their content. These adjustments represent a significant cost to an urban system from both technological unemployment and expensive worker retraining programs.

A priori, it is not obvious whether large cities will experience more or less impact from automation. On one hand, an influx of occupational diversity explains the wealth-creation, innovation, and success of cities [105, 113, 171, 180]. On the other hand, cities connect people with greater efficiency [171, 198]. This enables a greater division of labor that increases overall productivity [46, 202, 209] through occupational specialization. However, the division of labor may facilitate automation as it identifies routine tasks and encourages worker modularity. If these modular jobs are at greater risk of computerization, then more workers may be impacted by automation in large cities. These observations pose a puzzle: *are the forces of diversity, specialization, and the division of labor shaping a city's ability to accommodate automation?*

Here, we undertake a comparative examination of cities while measuring the relative impact of automation on employment. We also contextualize these measurements through a detailed analysis of the skill composition of different cities. Note that *impact* includes unemployment, but may also manifest itself through the changing skill demands of occupations as automation diminishes the need for individual types of skills [42, 44]. In light of imminent automation technology, we highlight a complicated relationship between labor diversity and specialization in cities, and discover that small cities are susceptible to the negative impact of automation.

3.3 Materials and Methods

3.3.1 Data Sets

The U.S. Bureau of Labor Statistics (BLS) data identifies the employment distribution of about 700 different occupations across each of 380 U.S. metropolitan statistical areas (MSAs) and combined statistical areas (CSAs. We refer to both CSAs and MSAs as "cities") in 2014. We consider MSAs in isolation only when they are not part of a CSA. CSAs have arisen as the best approximation for determine cities [14, 45, 46, 166, 168, 185, 186]. The resulting list of occupations considered in this study represents 99.99% of national employment according to the Occupational Employment Statistics (OES) data produced annually by BLS. From these employment distributions, we calculate the probability of a worker in city m having job j according to

$$p_m(j) = \frac{f_m(j)}{\sum\limits_{j \in Jobs_m} f_m(j)}, \tag{3.1}$$

where $Jobs_m$ denotes the set of job types in city m according to BLS data, and $f_m(j)$ denotes the number of workers in city m with job j.

For each occupation, the BLS O*NET dataset details the importance of 230 different workplace skills, such as Manual Dexterity, Finger Dexterity, Complex Problem Solving, Time Management, and Negotiation. BLS obtains this information through several separate surveys which group the raw O*NET skills into the following categories: Abilities, Education/Training/Experience, Interests Knowledge, Skills, Work Activities, and Work Context. We normalize the raw survey responses to obtain a value between 0 (irrelevant to the occupation) and 1 (essential to the occupation) indicating the absolute importance of that skill to that occupation. We refer to these values of skill importance as *raw skill values*.

3.3.2 Measures for Specialization and Diversity

We assess the specialization or diversity of the employment distribution in city m by calculating the normalized Shannon entropy. Shannon entropy [138], an information theoretic measure for the expected information in a distribution, can be normalized according to

$$H_{job}(m) = -\sum_{j \in Jobs_m} p_m(j) \cdot \frac{\log(p_m(j))}{\log\left(|Jobs_m|\right)}. \tag{3.2}$$

This quantity measures the predictability of an employment distribution given the set of unique occupations in a city. The measure is maximized when the distribution is least predictable (i.e. the distribution is uniform). Therefore, the denominator of $\log(|\text{Jobs}_m|)$ normalizes the entropy score so that we can compare the distributions of jobs in cities with different sets of job categories (see subsection B.2.1 for further discussion). The values for normalized Shannon entropy lie between 0 (specialization) and 1 (diversity). Normalized Shannon entropy has been used in a variety of fields, including virology [66], climatology [215], and city science [82].

For a given occupation, we normalize each raw skill value by the sum of the values to obtain the relative importance of each skill to that occupation (denoted $p_j(s)$). Similar to above, we measure the normalized Shannon entropy of the relative skill distribution of job j according to

$$H_j = -\sum_{s \in Skills_j} p_j(s) \cdot \frac{\log(p_j(s))}{\log\left(|Skills_j|\right)}, \tag{3.3}$$

where $Skills_j$ denotes the set of O*NET skills with non-zero importance to job j. We employ normalized Shannon entropy here to facilitate a fair comparison of relative skill distributions between jobs which may have received the same raw O*NET value for a given skill, but have different numbers of non-zero raw O*NET skills.

We obtain a distribution of relative skill importance for a city according to

$$p_m(s) = \sum_{j \in Jobs_m} p_j(s) \cdot p_m(j), \tag{3.4}$$

where $p_m(s)$ is the relative importance of skill s in city m. Again, we use normalized Shannon entropy to assess the skill specialization in a city according to

$$H_{skill}(m) = -\sum_{s \in Skills_m} p_m(s) \cdot \frac{\log(p_m(s))}{\log\left(|Skills_m|\right)}, \tag{3.5}$$

where $Skills_m$ represents the set of O*NET skills with non-zero importance in city m.

These aggregate skill distributions for a city may obfuscate the specialization of skills through the relative abundance of jobs in that city. For example, the city-level aggregation of skills may appear diverse, while the jobs within the city are actually specialized. The Theil entropy [182] of a city is a multi-level information theoretic measure defined by

$$T_m = \sum_{j \in Jobs_m} p_m(j) \cdot \frac{H_{skill}(m) - H_j}{H_{skill}(m)}. \tag{3.6}$$

$T(m) = 1$ indicates that each job specializes in exactly one skill, and $T(m) = 0$ indicates that the specialization of skills among jobs is equal to the specialization of skills on the city-level aggregation. We do not observe any jobs relying on exactly one skill, and so we expect the Theil entropy of any given city to be well below 1. We present $1 - T_m$ throughout the study for easy comparison to Shannon entropy. Note that both normalized Shannon entropy and Theil entropy are unit-less measures do to the normalizations employed; we therefore do not focus on their range of values across cities, but instead we focus on the relationship between labor specialization/diversity and other urban indicators.

3.4 Results: Exploring the Differential Impact of Automation in Cities of Different Sizes

3.4.1 The Expected Job Impact of Automation in Cities

We estimate automation's expected impact on jobs in cities according to

$$E_m = \sum_{j \in Jobs} p_{auto}(j) \cdot share_m(j), \tag{3.7}$$

where $Jobs$ denotes the set of occupations, $share_m(j)$ denotes the employment share (as a percentage) in city m with occupation j according to the U.S. Bureau of Labor Statistics (BLS), and $p_{auto}(j)$ denotes the probability of computerization for occupation j as estimated by [98] (see section B.3 for more details). We can interpret E_m as the expected percentage of total employment in city m subject to computerization. Each city should expect between one-half and three-quarters of their current employment to be affected in the foreseeable future due to improvements in automation (see Fig. 3-1A. Also note that this estimate differs from [98] which focused on national statistics). While this calculation omits potential job creation or job redefinition which typically accompany innovation [15, 147], it highlights the differential impact of automation across cities and smooths potential noise in the predicted automation of individual jobs. Expected job impact may represent employment loss or changes in the type of work performed by those workers (e.g. see [8,42,44]), which, in turn, may not produce changes in net employment.

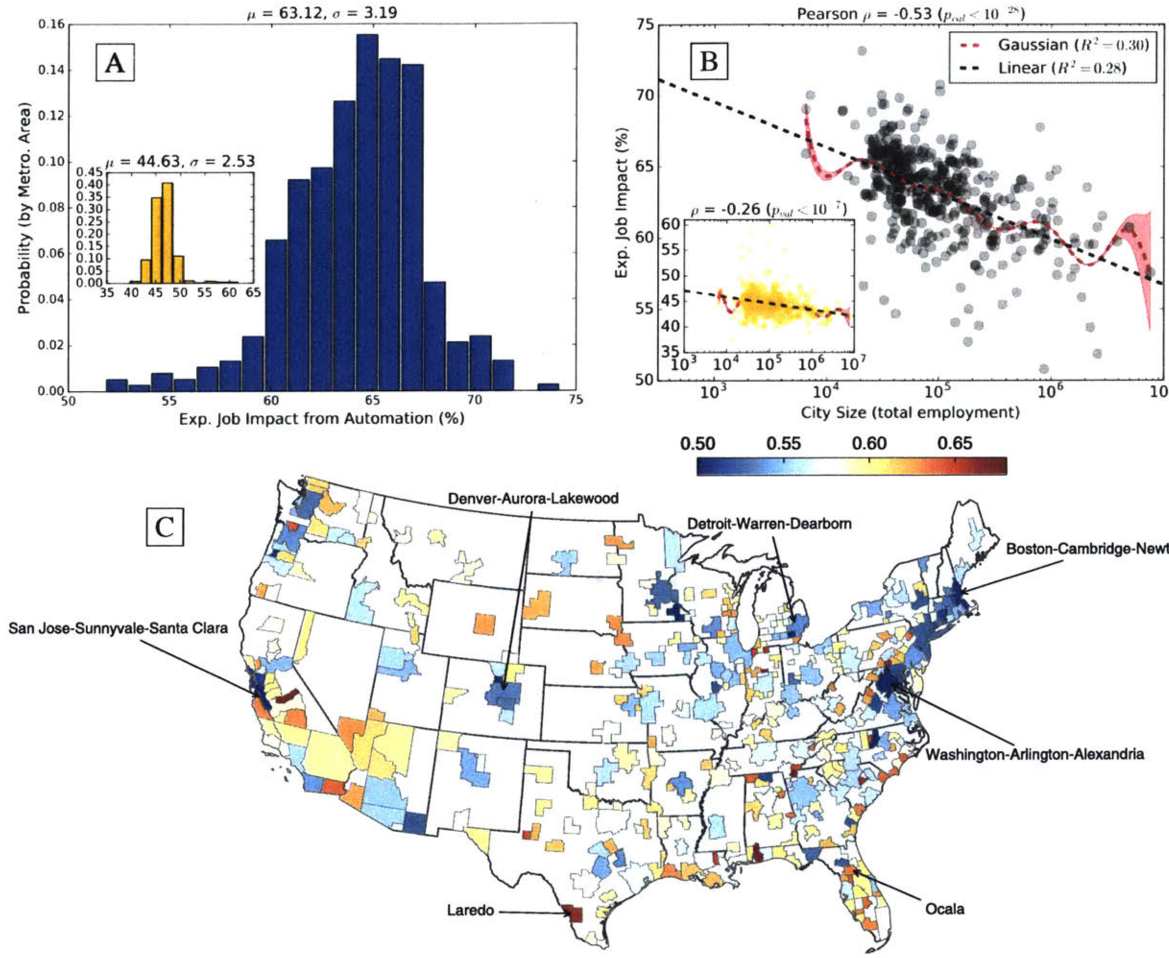


Figure 3-1: The impact of automation in U.S. cities. **(A)** The distribution of expected job impact (E_m) from automation across U.S. cities using estimates from [98]. (Inset) The distribution using alternative estimates [16]. **(B)** Expected job impact decreases logarithmically with city size using estimates from [98]. We provide the line of best fit (Slope $= -3.215$) with Pearson correlation to demonstrate significance (title). We also provide a Gaussian kernel regression model with its associated 95% confidence interval. (Inset) Decreased expected job impact with increased city size is again observed using alternative estimates [16] (best fit line has slope -1.24, Pearson $\rho = -0.26$, $p_{val} < 10^{-7}$). **(C)** A map of U.S. metropolitan statistical areas colored according to expected job impact from automation.

What differentiates cities' resilience to automation? Figure 3-1B demonstrates that expected job impact decreases according to $E_m \propto -3.2 \times \log_{10}(\text{city size})$, which suggests that larger cities are more resilient to the negative effects of automation. This relationship is significant with a Pearson correlation $\rho = -0.53$ ($p_{val} < 10^{-28}$), and shows that laborers in smaller cities are susceptible to the impact of automated methods ($R^2 = 0.28$). We confirm our finding using separate conservative skill-based estimates of the automatability of jobs [16] (Pearson $\rho = -0.26$ ($p_{val} < 10^{-7}$) and $E_m \propto -1.24 \times \log_{10}(\text{city size})$. See Fig.3-1B inset and subsection B.3.2). Despite the conservative nature of these alternative probabilities, we again observe increased resilience with city size. Furthermore, we demonstrate in the subsection B.3.1 that the observed negative trend relating city size to expected job impact from automation is robust to errors in the probabilities of computerization (i.e. p_{auto}) produced by [98] and robust to random removal of occupations from the analysis.

3.4.2 Labor Specialization in Large Cities

We explore the mechanisms underpinning resilience to automation by examining the most distinctive characteristics of urban economies: diversification and specialization. In particular, how does labor diversity, or specialization, mediate the relationship between city size and the expected job impact from automation? Since automation typically targets workplace skills [16], we consider the O*NET skill dataset, which relates occupations to their constituent workplace tasks and skills, in addition to employment data. For large cities, specialization (i.e. decreased Shannon entropy) appears in the employment distributions across occupations (Fig. 3-2A) and, separately, in the aggregate distributions of skills (Fig. 3-2B). Additionally, we use Theil entropy to measure the proportion of specialized jobs (in terms of skills) in comparison to the skill specialization of the city on whole. Figure 3-2C demonstrates an increasing proportion of specialized jobs in large cities (i.e. $1 - T_m$ decreases). See Materials and Methods for calculations of entropy measures.

In Figure 3-3, we examine eight regression models attempting to model the differential impact of automation across cities. In model 1, we first examine a baseline model using only generic urban variables, including city size (denoted $size_m$), median household income ($income_m$), the percent of population with a bachelor's degree ($bachelor_m$), per capita GDP (GDP_m), and the number of

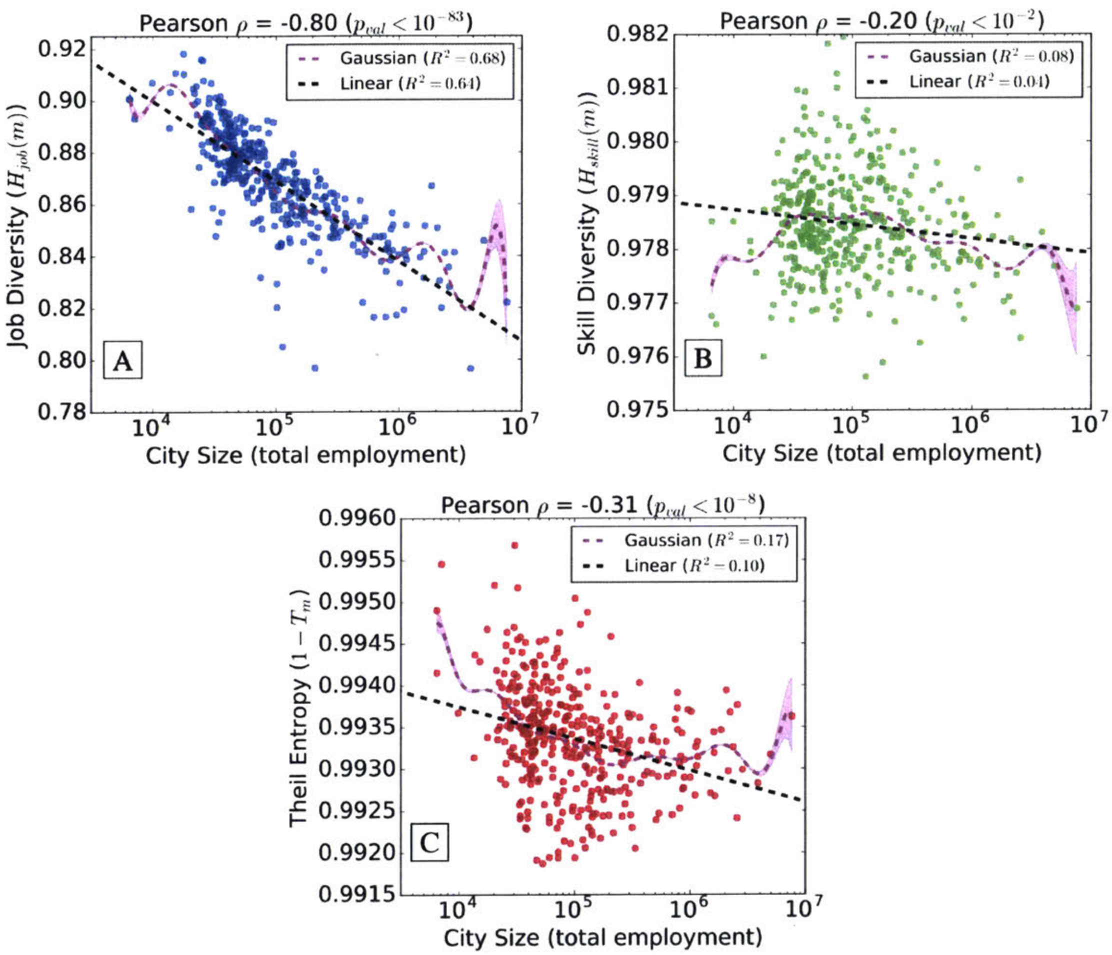


Figure 3-2: Large cities reveal increased occupational specialization through both job and skill distributions. **(A)** Shannon entropy of job distributions, $H_{job}(m)$, decreases with city size. **(B)** Shannon entropy of the O*NET skill distributions, $H_{skill}(m)$, decreases with city size. **(C)** Theil entropy, T_m, reveals the proportion of specialized jobs increases with city size. For plots (A), (B), & (C), we provide the line of best fit for reference, and we provide a Gaussian kernel regression model with its associated 95% confidence interval.

unique job titles ($jobs_m$). This generic model captures 53% of the variance in expected impact from automation across U.S. cities. Models 2, 3, & 4 use the information theoretic measures in three separate linear regression models to reveal that skill specialization (i.e. $H_{skill}(m)$) is the most predictive of expected job impact in cities ($R^2 = 0.20$) in the absence of other urban variables.

In models 5, 6, & 7, we demonstrate that the inclusion of each of the specialization measures produces models accounting for additional variance in expected impact over the use of generic variables alone. In particular, the inclusion of skill specialization (i.e. $H_{skill}(m)$) yields a model accounting for 60% of the variance in job impact (see model 6). Finally, we include all variables in a single model (see model 8) which produces the most predictive model accounting for 66% of the variance across cities (see Fig. 3-3A&B). We confirm the stability of our regression results by alternatively training the regression model on half of the cities and measuring the performance of the regression on the remaining cities as validation (see section B.4).

Each model that we tested yielded statistically significant coefficient estimates (note that variables were standardized before regression) and the inclusion of our labor specialization metrics yielded models with improved predictive power. Furthermore, we performed a formal mediation analysis that is presented in the subsection B.3.5. However, these observations should not be taken as conclusive evidence that labor specialization, or diversity, is causally related to the expected impact from automation in a city. This is due, in part, to the colinearity between variables used in model 8. For example, the city size coefficient ($size_m$) changes sign across the models in our analysis because of the strong relationship between city size and labor specialization, which we demonstrate in Figure 3-2.

The residuals between the actual and modelled values according to model 8 highlight notably resilient cities (given the model), such as Boulder, C.O. and Warner Robins, G.A., and notably susceptible cities, such as Napa, C.A. and Carson City, N.V. (see Fig. 3-3C). Examining these cities more closely may allow urban policy experts with a nuanced understanding of the policies in these cities to more easily identify causal mechanisms. The predictive power of this model and its reliance on workplace skills justifies our inclusion of skills data in addition to occupation data, and motivates us to characterize urban resilience to automation from the skills in cities.

A

Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variable				Coefficient (Standard Error)				
$size_m$	0.009				-0.016	0.011	0.012	-0.013
	$(< 10^{-4})$				$(< 10^{-4})$	$(< 10^{-4})$	$(< 10^{-4})$	$(< 10^{-4})$
$income_m$	−0.016				−0.013	−0.013	−0.015	−0.012
	$(< 10^{-5})$				$(< 10^{-5})$	$(< 10^{-5})$	$(< 10^{-5})$	$(< 10^{-5})$
$bachelor_m$	−0.005				-0.005	-0.002	-0.004	-0.002
	(10^{-5})				$(< 10^{-5})$	$(< 10^{-5})$	$(< 10^{-5})$	$(< 10^{-5})$
GDP_m	0.001				0.003	-0.001	-0.001	0.001
	$(< 10^{-5})$				$(< 10^{-5})$	$(< 10^{-5})$	$(< 10^{-5})$	$(< 10^{-5})$
$jobs_m$	−0.015				-0.002	-0.017	-0.020	0.001
	$(< 10^{-4})$				$(< 10^{-4})$	$(< 10^{-4})$	$(< 10^{-4})$	$(< 10^{-4})$
$H_{job}(m)$		0.006			-0.012			-0.010
		$(< 10^{-5})$			$(< 10^{-4})$			$(< 10^{-5})$
$H_{skill}(m)$			0.014			0.009		0.019
			$(< 10^{-5})$			$(< 10^{-5})$		$(< 10^{-5})$
$(1 - T_m)$				-8×10^{-5}			-0.005	0.012
				$(< 10^{-5})$			$(< 10^{-4})$	$(< 10^{-5})$
Sample Size	302	302	302	302	302	302	302	302
p-value	$< 10^{-10}$	0.0003	$< 10^{-10}$	0.96	$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$	$< 10^{-10}$
R^2	0.534	0.0429	0.197	$< 10^{-5}$	0.570	0.60	0.557	0.660
Adjusted R^2	0.534	0.0429	0.197	$< 10^{-5}$	0.570	0.60	0.557	0.660

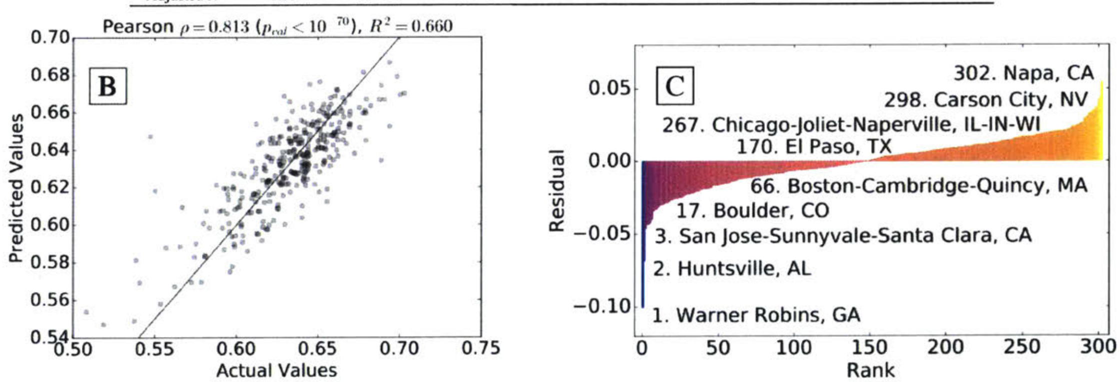


Figure 3-3: Labor specialization can model expected job impact (E_m) in cities. **(A)** A multiple linear regression analysis for predicting E_m that considers generic urban indicators including $\log_{10}$ city total employment ($size_m$), median annual household income ($income_m$), percentage of population with a bachelor's degree ($bachelor_m$), $\log_{10}$ GDP per capita (GDP_m), and the number of unique occupations ($jobs_m$). All variables have been standardized. **(B)** The actual E_m values for each city plotted against the predicted values using Model 8 from (A), which captures 66% of the variance in expected job impact from automation across U.S. cities (see section B.4 for additional analysis). **(C)** The distribution of residuals between the actual and predicted values from Model 8, and the rank of some example cities.

3.4.3 How Occupations and Workplace Skills Change with City Size

How do different types of occupations change with city size [178], and how do these changes contribute to the differential impact of automation across cities? While it is tempting to look only for the largest changes in employment share, more subtle differences for very automatable, or very not automatable, occupations can also produce big changes in expected job impact. We capture this confounding effect by decomposing the difference in expected job impact of cities m and n according to

$$\begin{aligned} E_m - E_n &= \sum_{j \in Jobs} p_{auto}(j) \cdot \left(share_m(j) - share_n(j)\right) \\ &= \sum_{j \in Jobs} \left(p_{auto}(j) - E_n\right) \cdot \left(share_m(j) - share_n(j)\right), \end{aligned} \tag{3.8}$$

where we have profited from $\sum E_n \cdot (share_m(j) - share_n(j)) = 0$. We consider the percentage of the difference explained by occupation j according to

$$\delta_{m,n}(j) = 100 \cdot \frac{\left(p_{auto}(j) - E_n\right) \cdot \left(share_m(j) - share_n(j)\right)}{E_m - E_n}. \tag{3.9}$$

Occupation j can increase or decrease the overall difference in expected job impact depending on the sign of the corresponding term in equation (3.8), or, equivalently, the sign of $\delta_{m,n}(j)$. In turn, this sign depends on the relative automatability of the occupation and the relative employment share. More details for this calculation and an example analysis comparing individual cities are provided in the subsection B.3.4.

In Figure 3-4, we employ an "occupation shift" to visualize the contributions of each occupation to the difference in expected job impact in large and small cities. After adding the employment distributions for the 50 largest cities and 50 smallest cities together, respectively, we calculate $\delta(j)$ for each occupation. Each occupation is assigned a quadrant and color based on the sign of $\delta(j)$ and the relative automatability of occupation j. This visualization identifies both occupations that increase the differential impact (i.e. occupations on the right) and occupations that decrease the differential impact (i.e. occupations on the left). For example, increased employment for Cashiers,

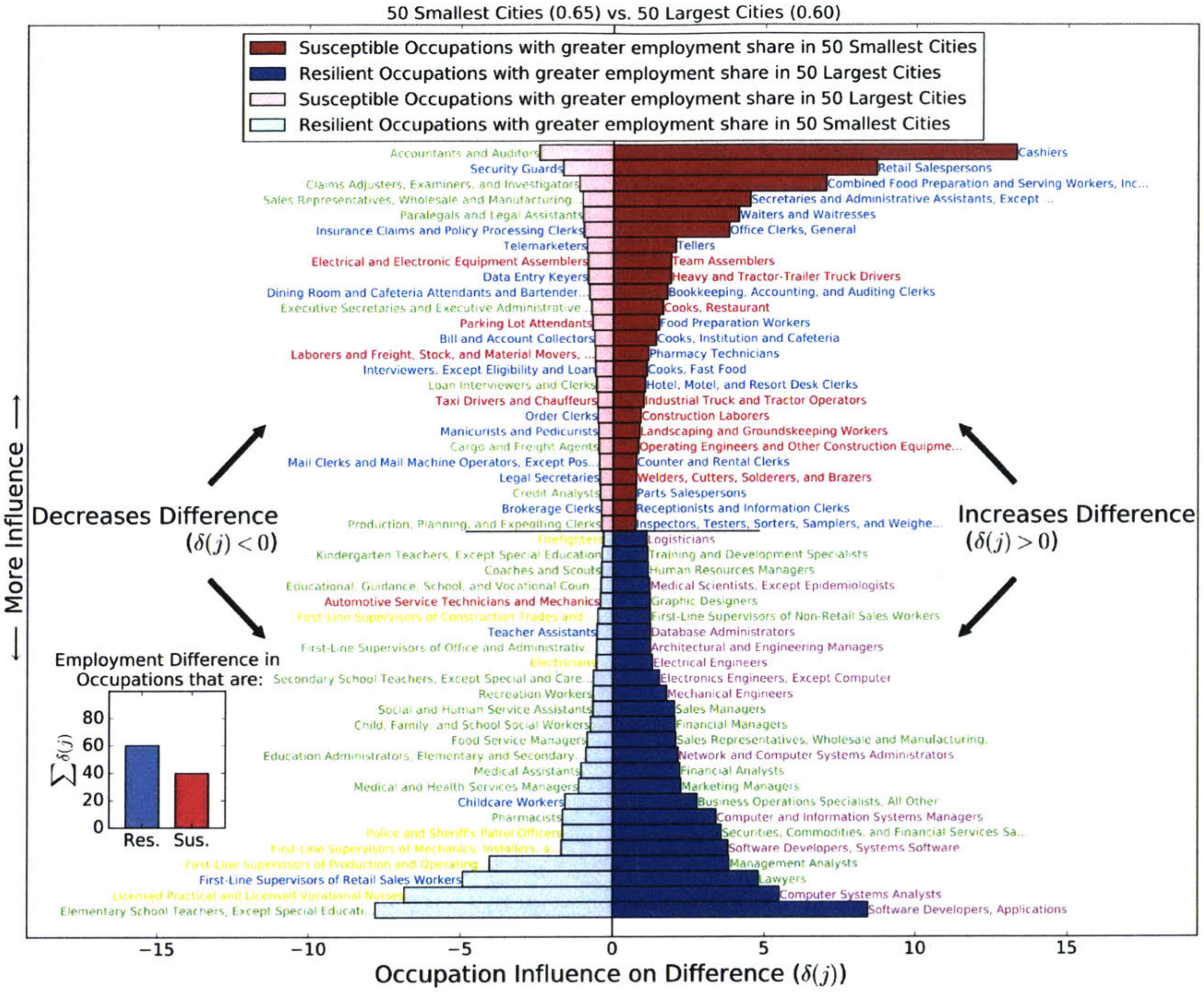


Figure 3-4: An occupation shift explaining the difference in expected job impact for the 50 largest cities (impact: 0.60) compared to the 50 smallest cities (impact: 0.65) using equation (3.8). Each horizontal bar represents $\delta_{(\text{Small Cities},\text{Large Cities})}(j)$. The occupation title is provided next to the corresponding bar and colored according to its job cluster. Red bars represent occupations with higher risk of computerization compared to the expected job impact in large cities. Blue bars represent occupations with lower risk of computerization compared to the expected job impact in large cities. Dark colors represent occupations that increase the difference, while pale colors represent occupations that decrease the difference in expect job impact. Bars in each of the quadrants are vertically ordered according to $|\delta_{(\text{Small Cities},\text{Large Cities})}(j)|$. The inset in the bottom left of the plot summarizes the overall influence of resilient occupations compared to occupations that are at risk of computerization.

which is relatively susceptible to automation, in small cities contributes the most to the overall difference in expected job impact. Likewise, differences in employment for Software Developers, a relatively resilient occupation, also increases the overall difference. On the other hand, increased employment for Elementary School Teachers, which is another relatively resilient occupation, in small cities decreases the difference. On aggregate, differences in employment for occupations that are relatively resilient to automation contribute the most to the differential impact of automation in large and small cities (see Fig. 3-4 inset).

To explore the role of resilient occupations further, we focus on how employment for different occupation types changes with city size. We use K-means clustering algorithm (i.e. occupations are instances and raw O*NET skill importance are features) to identify five clusters of jobs according to skill similarity (see Fig. 3-4 occupation labels and Fig. 3-5A. The complete list of occupations is provided in subsection B.6.3) and examine the scaling relationship between job clusters and city size according to (number of workers) $\propto$ (city size)$^{\beta}$ in Figure 3-5B. Note that the exponent, β, entirely describes the growth rate of these job clusters relative to city size. The job cluster comprised of highly specialized jobs, such as Mathematician and Chemist, exhibits a notably superlinear scaling relationship with city size ($\beta = 1.39$). This scaling exponent is similar to the scaling relationship observed for *Private R&D employment* ($\beta = 1.34$) found in [46] and is in good agreement with similar studies on job growth [44]. Furthermore, our finding of one job cluster exhibiting notably larger scaling than the other job clusters is stable to sub-sampling occupations at various rates (see section B.6.3). Managerial jobs also grow superlinearly, but to a weaker extent ($\beta = 1.08$). The job cluster exhibiting the slowest growth ($\beta = 0.94$) is comprised of entertainment and service jobs. We check the robustness of these scaling relationships using methods from [140] (see section B.6.3).

In Figure 3-5C, we quantify each job cluster's contribution to the differential impact of automation across large and small cities according to

$$\Delta_{\text{Small Cities,Large Cities}}(\text{Job Cluster}) = \sum_{j \in \text{Job Cluster}} \delta_{\text{Small Cities,Large Cities}}(j). \tag{3.10}$$

The low automatability and high difference in employment of highly specialized job cluster (repre-

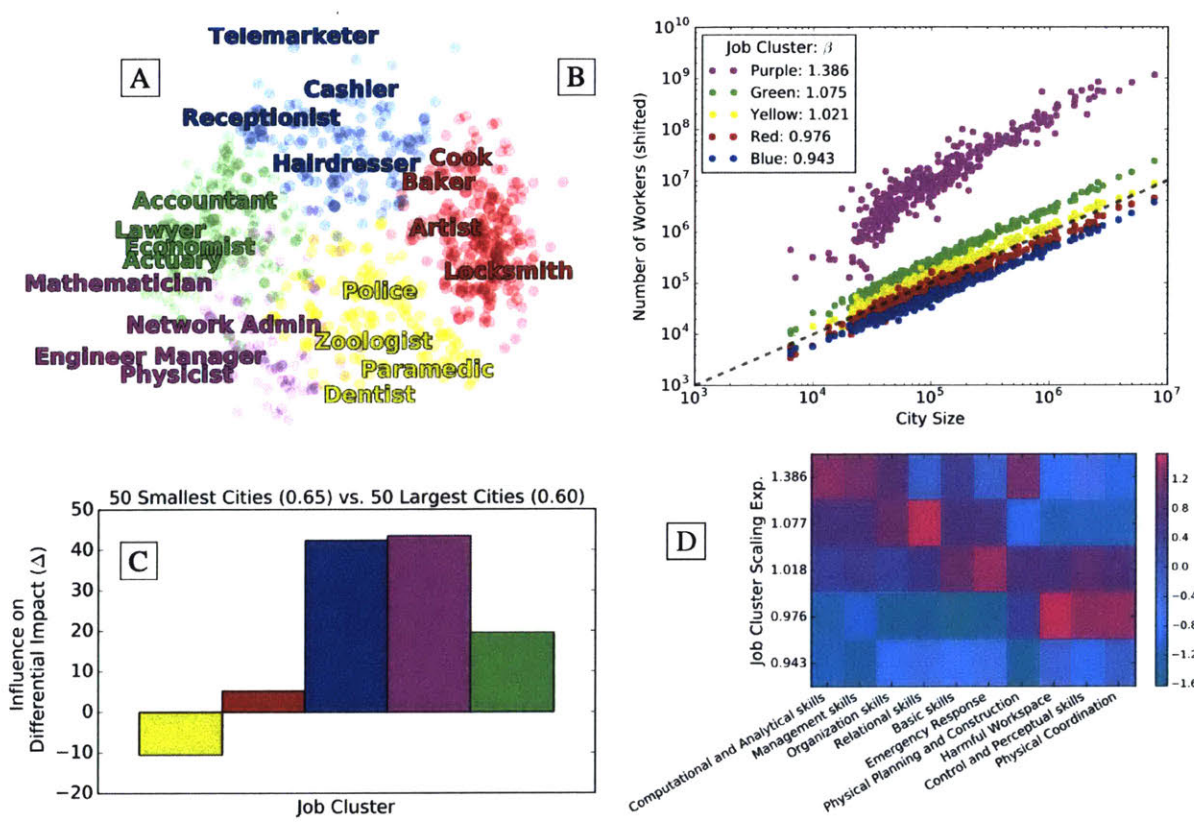
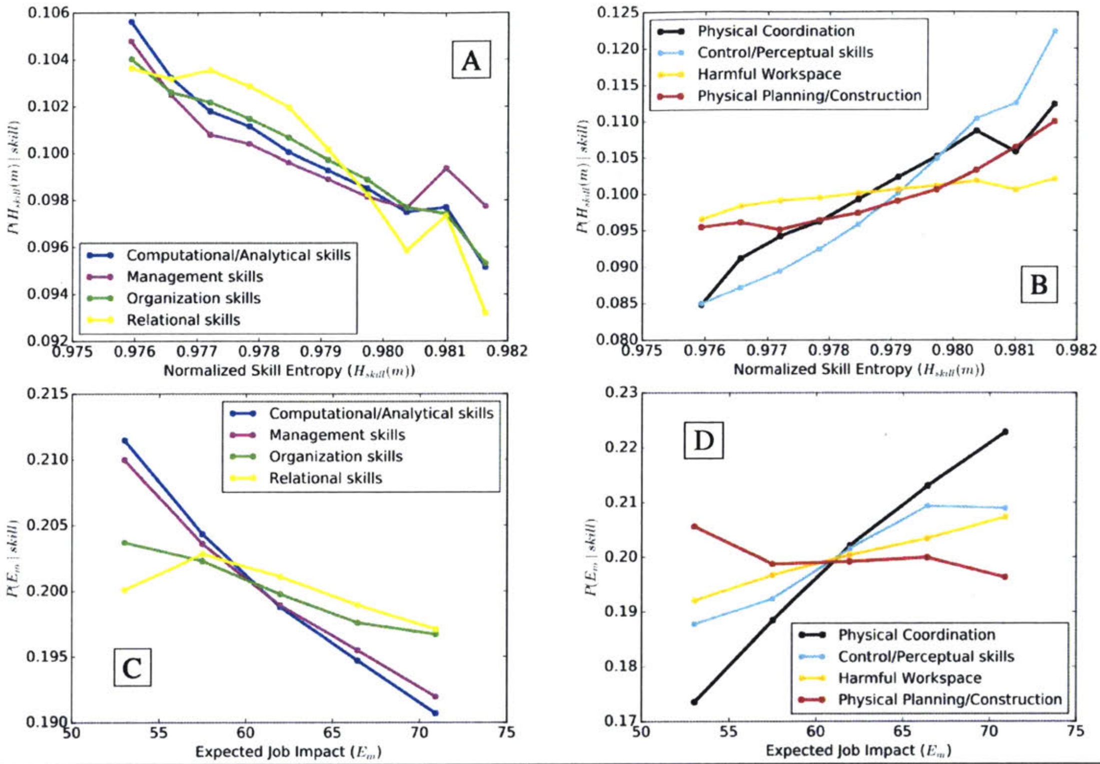


Figure 3-5: Technical occupations grow superlinearly with city size. **(A)** We project jobs onto a 2-D plane using principal component analysis. A few representative jobs from each cluster are highlighted (color). **(B)** We plot the employment (y-axis) in a given job cluster (color) versus the total employment in a city (x-axis), and vertically shift points according to the linear fit in log scale. The black dashed line has a slope of 1 for reference. **(C)** The influence of each job cluster on the difference in expected job impact of the 50 largest cities ($E_{\text{Large Cities}} = 0.60$) compared to the 50 smallest cities ($E_{\text{Small Cities}} = 0.65$) according to equation (3.10). **(D)** After summing the importance of each skill type to each job cluster, we calculate z scores for a skill type according to the distribution of importance across job clusters.

sented by purple) in large and small cities indeed explains a significant amount of the difference in expected job impact. However, we also find that the more susceptible occupations represented by the blue job cluster in Figure 3-5 accounts for a similar proportion of the difference. Interestingly, the differences in occupations from the yellow job cluster serve to decrease the differential impact of automation between large and small cities. These conclusions are supported by the analysis on individual occupations presented in Figure 3-4.

We confirm that the fastest growing job cluster is indeed comprised of "technical" jobs from their constituent workplace skills. We employ K-means clustering (i.e. O*NET skills are instances and the correlation of raw O*NET importance of skills across occupations are features) to simplify the complete space of O*NET skills to ten skill types based on the co-occurrence of skills across jobs (see subsection B.6.5 for complete description of skill clusters). These simplified skill types allow us to intuitively explore which skills indicate specialization or indicate resilience in cities. Computational/Analytical skills and Management skills are more likely in faster growing (i.e. superlinear) jobs, while physical skills, such as Physical Coordination and Control/Perceptual skills, indicate notably slower job growth with city size (Fig. 3-5D). We confirm our findings using alternative definitions for aggregate workplace tasks and skills (see section B.5).

The skills which are relied on by fast-growing technical jobs suggest mechanisms for resilience and growth in cities. Existing work [142] identifies that individual workers can gain skills to compete with automation, gain skills to complement automation, or seek industries removed from the impacts of automation. Similar to individual workers, the division of labor in large cities allows them to specialize in skills removed from the threat of automation. Computational/Analytical, Managerial, Organization, and Relational skills are more likely to be present in specialized and resilient cities (Fig. 3-6A&C), while Physical Coordination and Control/Perceptual skills indicate both decreased specialization and decreased resilience in cities (Fig. 3-6B&D). We confirm our results using alternative groups of workplace tasks [136] provided by O*NET (see subsection B.5.1) and again by examining the routineness of workplace tasks [24] (see subsection B.5.2). Figure 3-6E reflects the same conclusion by comparing the relationship of each skill type to city size (right column) and expected job impact (middle column) (see Figure B-12 in section B.5 for compari-



Skill Type E	**Impact Correlation**	**Log_{10} City Size Correlation**
Computational/ Analytical	-0.88 ($< 10^{-124}$)	0.58 ($< 10^{-34}$)
Management	-0.87 ($< 10^{-120}$)	0.52 ($< 10^{-27}$)
Organization	-0.62 ($< 10^{-41}$)	0.35 ($< 10^{-11}$)
Relationship	-0.26 ($< 10^{-6}$)	-0.06 (0.3)
Physical Planning	-0.07 (0.21)	0.18 (0.0006)
Basic/General	0.13 (0.01)	-0.24 ($< 10^{-5}$)
Control & Perceptual	0.45 ($< 10^{-19}$)	-0.14 (0.005)
Emergency Response	0.46 ($< 10^{-20}$)	-0.32 ($< 10^{-9}$)
Physical Coordination	0.83 ($< 10^{-99}$)	-0.52 ($< 10^{-26}$)
Harmful Workspace	0.90 ($< 10^{-140}$)	-0.54 ($< 10^{-28}$)

Figure 3-6: Workplace skills explain occupational specialization and job impact in cities. **(A) & (B)** Skill types in (A) indicate specialized cities, while skill types in (B) indicate occupational diversity. **(C) & (D)** Skill types in (C) indicate resilient cities, while skill types in (D) indicate increased job impact from automation. **(E)** The Pearson correlation of skill type abundance to the expected job impact and to $\log_{10}$ city size with p-values in parentheses. See subsection B.6.4 for a similar table for raw O*NET skills.

son with raw O*NET skills). Effectively, large cities employ workers whose skills better prepare them to interface with automation technology, while small cities rely more prominently on physical workers, who are more susceptible to automation.

3.4.4 Limitations

Many of the limitations inherent to occupation-level predictions [16,98] apply to our study as well. Specifically, our measure for the expected impact of automation in cities may represent technological unemployment, but also represents the skill re-composition of occupations in response to new technology. This means the expected impact of automation in cities may not relate to changes in net employment in cities. The actual effects of automation on net employment levels depend on several systemic variables including the availability of cheap labor [6, 110], future regulations around technology (e.g. taxing the use of robotics), and market demand with increased worker efficiency [42, 44].

Nevertheless, we expect the impact we are measuring to correspond to costly real-world changes in labor that high impact cities must overcome. For example, cities with high expected impact from automation will need to invest in worker retraining programs. These programs minimize technological unemployment by adapting the existing skills of workers to match the evolving skill demands with changing technology, but these programs are costly. Urban policy makers may also mitigate net employment loss by investing in new industries, but successful investment of this kind requires costly research and capital to attract those companies to a city.

3.5 Discussion

Cities are modern society's hubs for economic productivity and innovation. However, the impact of automation on employment in cities threatens to alter urbanization, which is largely driven by employment opportunity. Fortunately, urbanization itself appears to contain a mitigating solution. It is difficult to concretely identify causal mechanisms at the scale of this investigation, but, despite this difficulty, we highlight evidence for the division of labor in large cities and show its importance as a piece of the automation and urbanization puzzle.

In particular, large cities have more unique occupations and industries [218], but distribute employment less uniformly across those occupations. This juxtaposition of both diversity and specialization in large cities is reconcilable through the division of labor theory [202]. Under the division of labor argument, large firms have better ability to support specialized workers along with the management required to coordinate them [70]. To this end, we find that average number of workers per firm increases logarithmically with city size (see Figure B-1 in section B.1). At the same time, workers possessing specialized skills seek specific employment opportunities which maximize their financial return [51, 191]. The demand for specific specialized jobs increases occupational specialization while also increasing the number of unique job types and industries in a city [184].

What do large cities specialize in and why? The division of labor encourages worker modularity, which has the potential to impact whole groups of workers who are competing with automation technology. Therefore, specialization alone is not enough to explain the resilience to automation impact that we observe across cities. For example, Detroit, which is famous for its specialization in automotive manufacturing, has experienced economic down turn [133], while the San Francisco Bay area, epicenter of the information technology industry, continues to flourish despite the dot-com bubble (perhaps due to its support of a "creative class" of workers [92]). Our analysis highlights specific occupations, such as Mathematician and Chemist, as well as specific types of skills, such as Computational/Analytical skill, that explain the increased resilience of large cities. These occupations and skills may inform policy makers in small cities as they identify new industries and design worker retraining programs to mitigate the negative effects of automation on employment.

By quantifying relative *impact*, we provide an upper bound for *technological unemployment* in cities. Changing labor demands produce systemic effects, which make it difficult to precisely predict employment loss [19]. For example, the introduction of Automated Teller Machines (ATMs) suggested a likely decrease in human bank teller employment. However, contrary to this prediction, ATM technology cut the cost to banks for opening and operating new branches, and, as a result, national bank teller employment *increased* [42, 44]. However, these bank tellers performed different tasks, such as relationship management and investment advice, which required very different skills. Hence, by *impact*, we refer to the *magnitude* of the skill substitution shocks that cities must respond to.

The actual technological unemployment in a city will be shaped both by free market dynamics (e.g. shifts in supply and demand curves) and by economic and educational policy (e.g. worker re-training, or skilled migration). Nevertheless, we observe a strong trend relating city size to automation impact that is robust to errors in the automatability of individual occupations and occupational sub-sampling. For example, the estimates of occupational automation, which we employ in our analysis, would need to be severely flawed (errors over 50%) for the negative dependency on city size to disappear. Recognizing that small cities will experience larger adjustments to automation calls on policy-makers to pay special attention to the pronounced risks we have identified.

Despite being seemingly unrelated societal forces, we uncover a positive interplay between urbanization and automation. Larger cities not only tend to be more innovative [45, 46], but also harbor the workers who are prepared to both use and improve cutting-edge technology. In turn, these workers are more specialized in their workplace skills and less likely to be replaced by automated methods in the foreseeable future. These findings open the door for more controlled investigations with input from policy makers.

Chapter 4

Unpacking the polarization of workplace skills

Note: This chapter summarizes work from [10]. Ahmad Alabdulkareem and I were joint lead authors for this project. Consequently, parts of this study appear in Ahmad's PhD dissertation as well. In general, Ahmad created the skill network in this analysis, and I performed all additional validation analysis.

4.1 Motivation

Economic inequality is one of the biggest challenges facing society today. Inequality has been recently exacerbated by growth in high- and low-wage occupations at the expense of middle-wage occupations leading to a 'hollowing' of the middle class. Yet, our understanding of how workplace skills drive this process is limited. Specifically, how do skill requirements distinguish high- and low-wage occupations and does this distinction constrain the mobility of individuals and urban labor markets? Using unsupervised clustering techniques from network science, we show that skills exhibit a striking polarization into two clusters that highlight the specific social-cognitive skills and sensory-physical skills of high- and low-wage occupations, respectively. The connections between skills explain various dynamics: how workers' transitions between occupations, how cities acquire comparative advantage in new skills, and how individual occupations change their skill requirements. We also show that the polarized skill topology constrains the career mobility of individual workers, with low-skill workers 'stuck' relying on the low-wage skill set. Together, these results provide a new explanation for the persistence of occupational polarization, and inform strategies to mitigate the negative effects of automation and off-shoring of employment. In addition to our analysis, we provide an online tool for the public and policy makers to explore the skill network: skillscape.mit.edu.

4.2 Background: Workplace Skills and Job Polarization in the US

Economic inequality is on the rise, making it one of the central challenges facing U.S. policymakers today [135]. For example, absolute income mobility–the fraction of children who earn more than their parents–has fallen dramatically in the U.S. from 90% for children born in 1940 to 50% for children born in 1980 [68]. Some declared that the diminishing opportunity for prosperity and success marks the fading of the 'American dream' [126, 179], an ideal that is intimately associated with the United States national identity and ethos.

In contemporary political debate, one of the main culprits behind economic inequality has been the the lack of 'good jobs.' Both nationally and in a majority of U.S. metropolitan areas [3], economists have identified *occupational polarization*: an increasing proportion of high- and low-wage employment, accompanied by a relative decrease in employment share in middle-wage occupations [7, 23, 25]. The result is a 'hollowing' of the middle class. Mechanisms driving this trend include the off-shoring of work [83], something that has triggered recent shifts in international trade policy. Another mechanism is the automation of routine work, something that has sparked major concerns about the impact of automation on the future of work [19, 44, 142].

However, while mechanisms like off-shoring and automation ultimately impact people's jobs, they do not typically operate at the level of occupations. Rather, they alter the demand for specific workplace skills, tasks, knowledge, and abilities (hereafter, "skills"). If individual workers–or even entire cities–are unable to adapt their own skills appropriately, their ability to compete in the national and global labor market may be diminished.

Despite the important role of skills in occupational polarization, existing studies have explained the hollowing of the middle class in terms of annual wages [18] and broad subjectively defined occupational categories, such as "cognitive" versus "physical", or "routine" versus "non-routine" [25]. For example, suppose we use wage as a proxy for skill–that is, high-wage occupations are considered high-skilled occupations, etc. Then, if we find that growth in employment in middle-wage occupations is slower than growth in low-wage and high-wage occupations, we may conclude that demand for high-skills and low-skills are driving economic inequality. But this coarse-grained distinction may miss important relationships between skills that impact how workers adapt. This motivates the first set of questions we wish to explore in this paper:

Q1. Can we recover occupational polarization, at the finer-grained level of underlying skills, using an objective (unsupervised) data-driven clustering? How many distinct clusters, if any, does this skill structure contain? And does the skill structure exhibit smooth or abrupt transition between skill clusters?

To answer these questions, we apply data-driven methods to map skill *complementarity* as a network. We then use techniques from network science to identify distinct clusters of skills. Since we employ an unsupervised methodology, we demonstrate the usefulness of the resulting skill network by relating its structure to important real-world labor dynamics. Workers leverage skill complementarity between their existing skills to make career changes [58]. Similarly, cities leverage complementarity between industries to optimize productivity and increase their competitiveness in a global economy [161, 162, 176, 177]. We find that the structure of skill complementarity explains many stylized observations about occupational polarization and the hollowing of the middle class.

Having mapped the structure of skills and identified aggregate structure, the next obvious question to ask is: does the granular structure matter? Studies have identified the aggregate effects of skill complementarity on labor dynamics, such as the redefinition of skills comprising each occupation [44]. We unpack the role of skill complementarity in labor dynamics by exploring the following additional questions:

Q2. Can the skill topology predict changes in the latent skills of different *urban labor markets* (cities)? That is, given the skills used effectively in a given city at time t, can the network structure help us predict which new skills will become competitive in that city at time $t+1$?

Q3. Can the skill topology help us predict changes in the skill requirements of *a given job*–i.e. how the job's requirements changes over time?

Q4. Can the skill topology help us predict changes in the skills of *individual workers* as they transition from one job to another?

Having shown that skill polarization exists and affects some key dynamics, we ask:

Q5. Is the mobility of individual workers between skill sets (as they change jobs) consistent with the polarized structure of skills?

Our analysis suggests that the answer is 'yes.' We provide three types of evidence: i) workers tend to transition between occupations relying on the same skill set, ii) workers are unable to switch away from occupations relying equally on cognitive and physical labor, and iii) this constraining effect is reflected in national employment statistics.

In the next section, we describe our methodology in detail. We then present our analysis and discuss its implications and potential weaknesses before concluding the paper.

4.3 Materials and Methods

The O*NET program by the U.S. Department of Labor annually produces the publicly available O*NET database detailing the importance of 161 workplace skills, knowledge, and abilities to the completion of each of the 672 occupations recognized under the Standard Occupational Classification (SOC) System. The O*NET database is updated regularly allowing for annual snapshots of the relationships between occupations and skills through continual survey of workers from each occupation. We use annual O*NET data from the years 2010 through 2015. We denote the importance of skill $s \in S$ to occupation $j \in J$ using $onet(j,s) \in [0,1]$ where $onet(j,s) = 1$ indicates that

s is essential to j, while $onet(j,s) = 0$ indicates that workers of occupation j need not possess or perform s.

The Bureau of Labor Statistics (BLS) annually produces publicly available data detailing the distribution of SOC occupations in each U.S. metropolitan statistical area (MSA). MSA's represent an entire urban system including areas with large proportions of commuters employed in the city proper. We use the terms MSA and "city" interchangeably. Along with the numbers of workers of each occupation, BLS provides additional details about the annual salary of each occupation in each city.

The U.S. Census Bureau and BLS produce a monthly Current Population Survey (CPS) through a continuous survey process that produces representative samples of the U.S. population. Providing high resolution labor statistics is one of the primary goals of CPS, and, in particular, CPS records changes in occupations of survey participants over a 1.5 year period for which that participant is an active contributor to the survey. For our purpose, we are interested only in participants who reported one occupation when they were first surveyed in 2014 and reported working a different occupation when they were surveyed one year later in 2015. There are several methods for joining different time periods of the CPS data [144], so we employed strict merging criteria including participant ID, gender, sex, state of residency, and age to verify the validity of our occupation transitions. The result is a dataset of 5,400 occupation transitions for individual U.S. workers from 2014 to 2015.

4.4 Results: Exploring Skill Polarization

4.4.1 Constructing the Skill Network

Typically, occupations are the unit of interest in labor dynamics. However, in other situations occupations are broken down even further because the labor requirements that define an occupation are reflected in the skills possessed by workers of that occupation (see Fig. 4-1A). These skill requirements represent key features that uniquely identify occupations, and so we seek a data-driven methodology that maximizes the information about each occupation while minimizing the potential bias that can accompany investigations through ad hoc skill aggregations. However, raw O*NET data do not control for ubiquitous skills, such as "Identifying Objects" and "Communicating with Supervisors and Peers" (see Figure C-1). Therefore, we focus on skills that are over-expressed in an occupation by calculating the revealed comparative advantage [111, 115, 117] (RCA) of each skill in an occupation according to

$$rca(j,s) = \frac{onet(j,s) / \sum_{s' \in S} onet(j,s')}{\sum_{j' \in J} onet(j',s) / \sum_{j' \in J, s' \in S} onet(j',s')}. \tag{4.1}$$

RCA (also known as "location quotient") has been used in a variety of applications, including identifying the key industries in cities [106, 124, 195], identifying the key exports of nations [115,212], and identifying key features in the labor distributions of industries [163]. Similarly, occupations are distinguishable from each other according to their "effective use" of skills; we denote effective use of skills using $e(j,s) = 1$ if $rca(j,s) > 1$, and $e(j,s) = 0$ otherwise. Here, RCA normalization compares the relative importance of a skill to an occupation (i.e. the numerator in eq. (4.1)) to the expected relative importance of a skill on aggregate (i.e. the denominator); $rca(j,s) > 1$ indicates that occupation j relies on skill s more than expected on aggregate. Skill *complementarity* [58, 162] (denoted θ) is then the minimum of the conditional probabilities of a pair of skills being effectively used by the same occupation:

$$\theta(s,s') = \frac{\sum_{j \in J} e(j,s) \cdot e(j,s')}{max\left(\sum_{j \in J} e(j,s), \sum_{j \in J} e(j,s')\right)} \tag{4.2}$$

The distribution of complementarity values is provided in Figure 4-1B. This methodology identifies skill pairs that co-occur across occupations and represent key occupational features. Co-occurrence captures how a pair of skills support each other, either by boosting the productivity of a worker who possesses both skills, or by the ease of acquiring both skills simultaneously. Our definition of complementarity is agnostic to the exact source of the complementarity. We call the resulting network of skill complementarity the "Skillscape" (see Fig. 4-1C. See section C.1 for

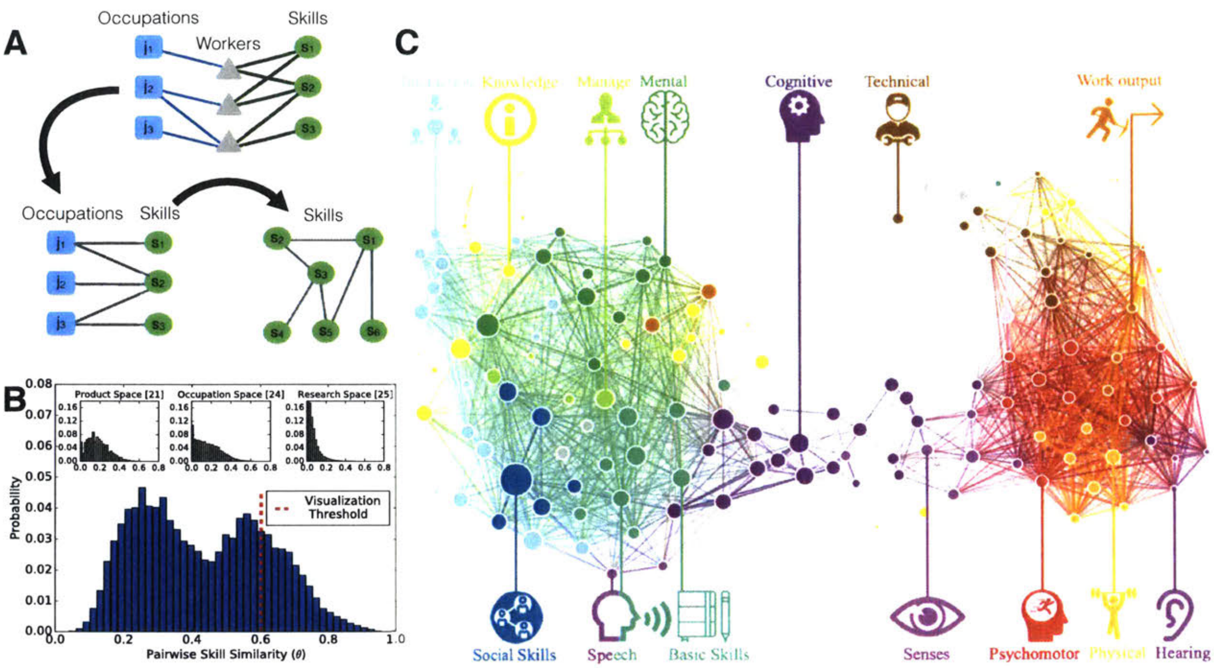


Figure 4-1: Constructing the Skillscape. **(A)** An occupation is identified through the skills of workers of that occupation. The bipartite network connecting occupations to required skills is a result of an underlying tripartite network containing workers as a conduit between occupations and skills. Relationships between skills are determined from their co-occuring importance across occupations. **(B)** Unlike previous applications of RCA (insets), the Skillscape contains a bimodal distribution of pairwise skill complementarity. **(C)** The Skillscape thresholded according to a minimum skill similarity (i.e. $\theta > 0.6$) visibly reveals two communities of complementary skills and respects expertly-derived O*NET categories (colors). Node sizes reflect the total skill similarity shared between that skill and all other skills.

visualizations of this methodology and a visualization of the Skillscape as a skill-to-skill complementarity matrix).

Ideally, the aggregate structure in the skill network should correspond to meaningful labor dynamics. For example, node communities in the skill network represent clusters of complementary skills that define important types of labor. To this end, we identify skill types using Louvain community detection [50]. This method greedily identifies node communities by comparing the density of connections within a community to the density of connections between communities. This method requires no assumptions about the number of communities to be found. This community detection method has been widely used in a variety of fields, including neuroscience [187, 205], transportation research [32], social science [39], business/management research [79], climatology [86], and cybersecurity [71].

4.4.2 Identifying Skill Polarization from the Bottom-Up

Existing studies have explained the hollowing of the middle class in terms of annual wages [18] and broad subjectively defined occupational categories, such as "cognitive" versus "physical", or "routine" versus "non-routine" [25]. For example, it has been shown that some decades are marked by a relative increase in the share of employment in high-wage and low-wage jobs, at the expense of workers in middle-wage jobs. While these results identify the *outcome* of labor polarization, they do not relate this polarization to the underlying topology of skills. The limitations discussed above have led researchers to call for new high-resolution models that more accurately account for raw workplace tasks and skills [7].

On aggregate, our cluster analysis reveals that the skill network is highly polarized into a socio-cognitive cluster of skills and a sensory-physical cluster (see Fig. 4-1C). This polarization is not an artifact of the methods we employ (see Fig. 4-1B), and is significantly different from comparisons to a null model (see section C.4). This divide between traditionally "technical" and "non-technical" skills largely supports previous findings characterizing U.S. occupational polarization. For example, let *SocioCog* denote the set of socio-cognitive skills according to the community detection algorithm (see Fig. 4-2A). We measure the *cognitive skill fraction* of job j according to

$$cognitive_j = \frac{\sum_{s \in SocioCog} onet(j,s)}{\sum_{s \in S} onet(j,s)}. \tag{4.3}$$

Jobs with higher $cognitive_j$ tend to yield higher annual wages (see Fig. 4-2B, Pearson correlation $\rho = 0.42$, $p_{val} < 10^{-26}$). This result demonstrates the direct link between the skill polarization we have identified, and occupational polarization, which is characterized by growing employment share for high- and low-wage occupations [18].

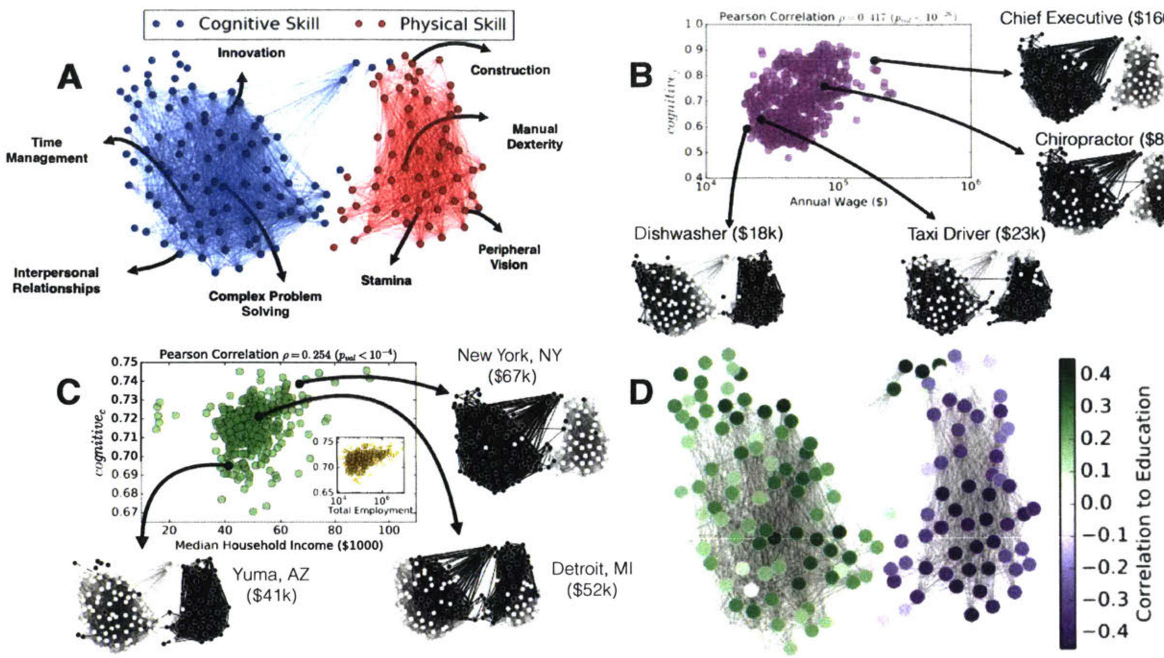


Figure 4-2: The polarized Skillscape explains occupational wage polarization and economic well-being of urban workforces. **(A)** Community detection on the complete Skillscape network (i.e. no minimum θ) reveals two communities of complementary skills: socio-cognitive skills (blue) and sensory-physical skills (red). The displayed network is filtered ($\theta > 0.6$) for visualization purposes. **(B)** Occupations relying on socio-cognitive skills tend to make higher annual salaries. **(C)** Larger cities rely more strongly on socio-cognitive skills (inset) yielding higher median household income by comparison to smaller cities. In (B) & (C), example occupations (cities), along with their annual wages (median household income), are projected onto the Skillscape using black nodes for effectively used skills. **(D)** The skill network colored by correlation between $onet(j,s)$ and the average educational degree requirement across occupations.

4.4.3 Comparison with Top-Down Categorization

One might wonder whether our approach to skill polarization captures factors beyond those well-known in the literature. Previous work has leveraged ad hoc distinctions between occupations based on their reliance on routine versus non-routine skills to study occupational polarization [7, 24]. Does our approach to *skill* polarization add further predictive power?

In agreement with existing work, our investigation of skills should incorporate known worker-related variables, such as education. Education level is a key factor in determining wages [18, 129] as educational institutions act as a social "sorting machine" [129] when students begin their careers. The skill polarization we observe respects the educational requirements of occupations. Indeed, if we correlate $onet(j,s)$ and the average degree requirement for each occupation, we find that skills in the socio-cognitive cluster indicate higher education requirements across occupations. Conversely, occupations with more lenient degree requirements tend to rely on sensory-physical skills (see Fig. 4-2D).

Although the aggregate polarization of skills captures known features that determine worker wages, it remains to show the added predictive power gained from the granularity of our model. In particular, do the existing ad-hoc distinction between routine vs. non-routine skills, and the level of education, completely explain the differences in wages? Or, does the polarized structure of the skill network we have identified play an independent role? We investigate this question by comparing different regression models in Figure 4-3.

In Model 1, we consider the relative importance of routine labor by combining O*NET data with the routine O*NET variables defined in [24] (i.e. $\sum_{s \in R} onet(j,s) / \sum_{s \in S} onet(j,s)$, where R are routine O*NET variables, $R^2 = 0.12$). Model 2 demonstrates the superior performance of $cognitive_j$ ($R^2 = 0.15$). Additionally, we consider the total skill content required by each occupation (i.e. $\sum_{s \in S} onet(j,s)$) in Model 3 ($R^2 = 0.30$). Models 4, 5, & 6 demonstrate total skill content and cognitive skill fraction outperform models using the variable for routine labor (Model 6 has $R^2 = 0.46$) and that total skill content is largely orthogonal to reliance on cognitive skills. In Model 5, we consider variables for each occupation's total employment whose highest educational attainment was a high school diploma, a bachelor's degree, etc. Modelling with these educational variables alone performs worse than using $cognitive_j$ ($R^2 = 0.12$). Finally, Model 8 demonstrates the improved performance from including the variable for routine labor and total skill content ($R^2 = 0.42$), but maximum performance is achieved when including $cognitive_j$ as well (Model 9 has $R^2 = 0.49$). We provide out-of-sample testing to demonstrate the robustness of our models' performance; we find that the inclusion of skill related variables in Models 8 and 9 reduce the variance in model performance. Additionally, the standard error and statistical significance of coefficient estimates are reported in the regression table.

In summary, we find that cognitive skill fraction ($cognitive_j$) explains the annual wages of occupations better than models using routine labor or educational variables alone. Additional regression analyses detailing occupation wages and the median household income of cities are provided in the section C.6.

4.4.4 The Skills of Urban Workforces

We combine the O*NET database with employment distributions in U.S. cities according to the BLS to approximate the importance of each workplace skill to each urban workforce. Denoting the number of workers in city c with occupation j using $bls(c,j)$, we combine the two data sets according to

$$CS(c,s) = \sum_{j \in J} bls(c,j) \cdot onet(j,s), \tag{4.4}$$

where $CS(c,s)$ denotes city c's reliance on workplace skill s (see section C.5). As with the raw O*NET data, certain jobs and certain skills are ubiquitous across many cities. We again apply RCA on $CS(c,s)$ to calculate $rca(c,s)$ (as in eq. (4.1)) and identify which skills are effectively used in each city. Similar to occupations, $rca(c,s) > 1$ indicates the effective use of s in c. Additional explanatory visualizations in section C.5.

By considering $onet(c,s)$ in place of $onet(j,s)$ in equation (4.3), we can compute the same cognitive skill fraction (denoted $cognitive_c$) for entire cities. Analogously, Figure 4-2C shows that cities with higher median household incomes ($\rho = 0.25$, $p_{val} < 10^{-4}$) also tend to rely on socio-cognitive skills. We also find significant correlation between city size and the degree to which the city's local labor market relies on socio-cognitive skills: larger cities are more socio-cognitive (see inset in Figure 4-2C). Together, these results suggest that inequality *between* cities may be driven by processes that operate at the level of skill supply, and the ability of cities to effectively exploit skill complementary within the socio-cognitive niche.

4.4.5 Skillscape Proximity & Skill Acquisition

Does skill complementarity (i.e. θ) correspond to "nearby" skills in practice? We capture this using a measure for the network "proximity" between each pair of skills based on the network topology and an empirical measure for skill acquisition. Let $E_t^{\lambda}(j)$ represent the set of skills that

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
	Dependent Variable: Annual Wage of Occupations								
$cognitive_j$		0.387^{***}		0.352^{***}		0.567^{***}			0.591^{***}
Routine Labor	-0.352^{***}			-0.040	-0.288^{***}			-0.277^{***}	0.244^{***}
$\sum_s onet(j,s)$			0.548^{***}		0.512^{***}	0.448^{***}		0.484^{***}	0.553^{***}
No GED							-0.113^{*}	-0.019	0.002
H.S. Diploma							-0.139^{**}	-0.064^{*}	-0.059^{*}
Associate's Degree							0.021	-0.007	-0.045
Master's Degree							0.102^{**}	-0.027	-0.045
Doctoral Degree							0.234^{***}	0.186^{***}	0.157^{***}
R^2	0.124	0.150	0.301	0.151	0.382	0.456	0.119	0.421	0.491
adj. R^2	0.122	0.149	0.299	0.148	0.380	0.453	0.112	0.415	0.484
	$p_{val} < 0.1^{*}$, $p_{val} < 0.01^{**}$, $p_{val} < 0.001^{***}$								

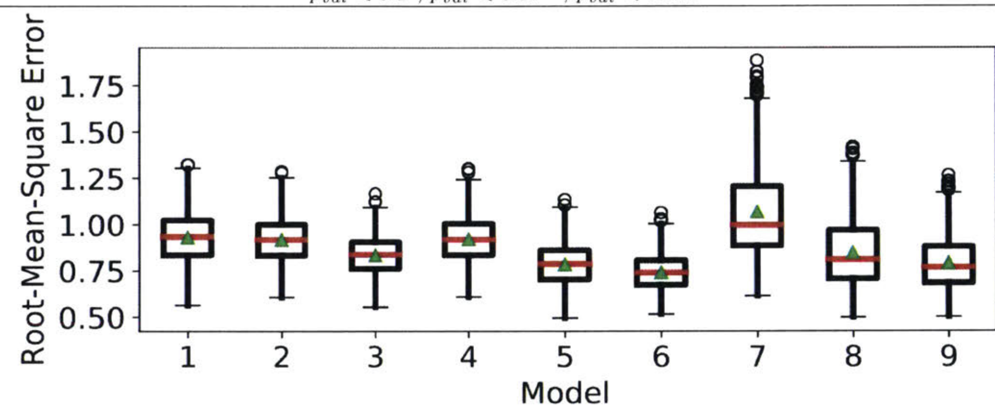


Figure 4-3: Reliance on cognitive skills predicts increased annual wages according to OLS regression. As a baseline, we consider the relative importance of routine labor using routine O*NET variables from [24]. In addition to cognitive skill fraction ($cognitive_j$), we calculate the total skill content ($\sum_s onet(j,s)$) of each occupation. Each educational variable represents the total employment in that occupation whose highest educational degree is a high school diploma, bachelor's degree, etc. All variables were standardized before regression. Standard errors are reported in parentheses and asterisks indicate the statstical significance of coefficient approximations. We perform out-of-sample testing for each model through 1,000 trails of randomly selecting 75% of the occupations as training data and measuring the root-mean squared error of the resulting model applied to the remaining 25% of occupations. We represent the resulting model performance as box plots. Median error represented by a red line, while the mean error is represented by the triangles.

job j effectively uses at time t according to some threshold $\lambda \geq 0$; that is,

$$E_t^{\lambda}(j) = \{s \in S \mid rca_t(j,s) > \lambda\}. \tag{4.5}$$

We say a skill is "acquired" if it was not effectively used at time t_1 and becomes effectively used at t_2. Specifically, we denote the set of occupation j's acquired skills using

$$Acquired_{t_1,t_2}^{\lambda_1,\lambda_2}(j) = \{s \in S \mid s \notin E_{t_1}^{\lambda_1}(j), s \in E_{t_2}^{\lambda_2}(j)\}. \tag{4.6}$$

According to this definition, two different thresholds, λ_1 and λ_2, are selected for time step t_1 and t_2, respectively. This allows us to vary the magnitude of skill change we are interested in; that is, $\lambda_2 - \lambda_1$ determines the severity of the skill change in order for a skill to be acquired for $\lambda_2 > \lambda_1$. Notice that if $\lambda_1 > \lambda_2$, then this would be skill loss instead of acquisition. For the analysis in the main text, we consider discrete choices of λ according to each percentile of empirical RCA values (i.e. $\lambda_1, \lambda_2 = 0\%, 1\%, \ldots, 99\%, 100\%$ such that $\lambda_1 < \lambda_2$).

For a measure to be predictive of skill acquisition, skills with high scores (e.g. in onet) should have higher probability of being acquired for each choice of λ_1 and λ_2. For example, if we consider raw O*NET values (i.e. $onet(j,s)$) as a proxy for skill acquisition, then skills that are not effectively used by an occupation (i.e. $s \notin E_{t_1}^{\lambda_1}$) but have a high score (i.e. $onet(j,s) \to 1$) should have higher probability of being acquired. We capture this by ordering pairs of occupations and skills by their O*NET value such that the skill is not effectively used by that occupation (i.e. $s \notin E_{t_1}^{\lambda_1}(j)$) and binning these pairs into 30 quantiles according to associated O*NET value (i.e. $onet(j,s)$). For each pair, we calculate the probability that the skill is acquired in t_2 (i.e. $s \in Acquired_{t_1,t_2}^{\lambda_1,\lambda_2}$) across all choices of λ_1 and λ_2. This produces several points for each quantile; we use the average and the 95% confidence interval for each quantile to simplify the data for visualization. This method is similar to previous studies using network topology to predict the regional acquisition of new industries [162]. In the main text, we consider a LOWESS interpolation through the averages of each quantile. In addition to raw O*NET as a proxy for skill acquisition, we also consider RCA values and a measure of network skill proximity (described below). In addition to the interpolated plots of the main text, we provide bar plots with the associated error bars in the Figure C-15.

For non-effectively used skills (i.e. $s \notin E_{t_1}^{\lambda_1}(j)$), we say a skill is nearby to occupation j if that skill has strong average complementarity with the effectively used skills of j (i.e. $E_{t_1}^{\lambda_1}$). We capture this by introducing a topological measure for proximity according to

$$proximity(j,s) = \frac{\sum\limits_{s' \in E_{t_1}^{\lambda_1}(j)} \theta(s,s')}{\sum\limits_{s' \in S} \theta(s,s')}. \tag{4.7}$$

This proximity measure only utilizes information at t_1 to evaluate the status of all skills. Note that analogous calculations can determine Skillscape proximity from urban workforces by considering $rca(c,s)$ instead of $rca(j,s)$, and similarly for individual workers. section C.7 provides an alternative analysis using receiver operating characteristic curves (ROC).

4.4.6 Dynamics: Skill Polarization and Transitions Between Jobs

Skill acquisition through explicit education can be costly and time consuming, so, more commonly, workers transition between occupations based on the similarity of their skill set and the skill requirements of each occupation [102]. Ideally, the granular network topology of the skillscape should capture this dynamic. In combination with the aggregate polarization of skills, we also expect that worker mobility between skill categories should be constrained. This hypothesis is not directly testable because we do not understand the precise mechanisms for worker adaptation, nor do we understand that mechanism's interplay with other market equilibrium dynamics [7,44].

However, the hypothesis reveals three labor trends that the skill network should relate to. First, the topological proximity of skills on the network should relate to skill-related trends, including the changing skill requirements of individual workers, the dynamic skill requirements of occupations, and changes in the latent skill sets of urban labor markets. Second, if the connections between skills represents skill complementarity, then workers are more likely to transition to occupations relying on skills in the same skill cluster. Third, skill polarization represents a bottleneck in workers' upward mobility towards high-wage occupations. This should lead to disproportionately high employment below a certain $cognitive_j$ threshold, rather than a smooth distribution of employment across the range of $cognitive_j$ values. In the remainder, we demonstrate how indeed the Skillscape relates to these important features of the U.S. labor market.

We validate our first prediction in Figure 4-4 using a topological measure for skill proximity (i.e. $proximity(j,s)$, see Fig. 4-4A for an example of Skillscape proximity). A worker's skill set can be approximated from the skill requirements of their occupation, and we suppose skills that are nearby to these skill sets in terms of network topology are more attainable by that worker. Analogously, nearby skills to a city's local labor market are more likely to be obtained by workers in that city. We empirically validate our proximity measure by comparison to the probability that a skill is acquired (i.e. $s \in Acquired_{t_1,t_2}^{\lambda_1,\lambda_2}$) by a city (see Fig. 4-4B), an occupation (see Fig. 4-4C), or by an individual worker (see Fig. 4-4D). In each case, network proximity most strongly indicates newly acquired skills, thus demonstrating the highly granular relationship between the skill network topology and labor dynamics. We provide an alternative analysis in section C.7, and bar plots including 95% confidence intervals in subsection C.7.4.

For our second prediction, since occupational transitions represent local changes in workers' skill requirements, the polarized network of skills should constrain mobility between low-wage sensory-physical occupations and high-wage socio-cognitive occupations. We capture this explicitly by binning occupation transitions into quantiles (each representing 780 transitions) according to the cognitive skill fraction of the workers' starting occupation ($cognitive_{j_A}$) and examining the average cognitive change (i.e. $\Delta cognitive = cognitive_{j_B} - cognitive_{j_A}$, see Fig. 4-5A) and average magnitude of cognitive change (Fig. 4-5B) for each bin. We consider workers selecting there new occupations at random as a null model for comparison (See subsection C.7.2 for a discussion of alternative null models, including randomizing the selection of "cognitive skills"). Workers transitioning from sensory-physical occupations tend towards new occupations with higher socio-cognitive skill fraction, but the magnitude of change is less than would be expected under random occupation selection (and vice versa for the other end of the spectrum). By contrast, workers transitioning from mid-quantile occupations, which represent starting occupations which effectively use cognitive and physical skills evenly, exhibit larger magnitudes of change in $cognitive_j$ compared to the null model. In conclusion, workers of occupations relying strongly on one skill community tend towards other occupations within the same skill community, thus validating the second prediction.

For our third prediction, first note that the definition of skill complementarity [58] indicates increasing returns to combining skills within each skill community. Therefore, skill communities may be explained by the easy acquisition of related skills or by production efficiencies offered by workers who have complementary skills. However, this also means that workers relying on sensory-physical skills will face difficulty acquiring socio-cognitive occupations because they are unprepared to exploit large proportions of the socio-cognitive skills. Until they have a sufficient proportion of socio-cognitive skills, sensory-physical workers are bottle-necked by the polarized structure of skill complementarity. If true, then we expect disproportionately high employment in occupations under some threshold of $cognitive_j$.

Indeed, binning national employment according to $cognitive_j$ yields a trimodal distribution (see Fig. 4-5C. Additional years and binning, and city employment distributions, are provided in subsection C.7.4). The upper and lower modes of the distribution correspond to workers who are effectively exploiting the skill complementarity *within* each of their respective skill communities. The presence of a third mode in the middle suggests that skill polarization constrains workers from obtaining attractive socio-cognitive skills. Thus demonstrating the third prediction, adding more evidence towards our hypothesis that the network of skill complementarity *constrains* labor mobility.

Finally, Figure 4-5D quantifies the average complementarity score of each skill as an approximation for that skill's network embeddedness. Considering our hypothesis and the strong relationship between skill proximity and skill acquisition, network embeddedness should correlate with increased labor mobility (individual skills are shown in section C.2.

The Skillscape maps the structure of workplace skill complementarity, and connects urban workforces and occupations to their constituent skills. While our analysis identifies the specific skill requirements of low- and high-skill occupations that characterize occupational polarization, our analysis does not reveal whether occupational polarization is a result of skill polarization, or

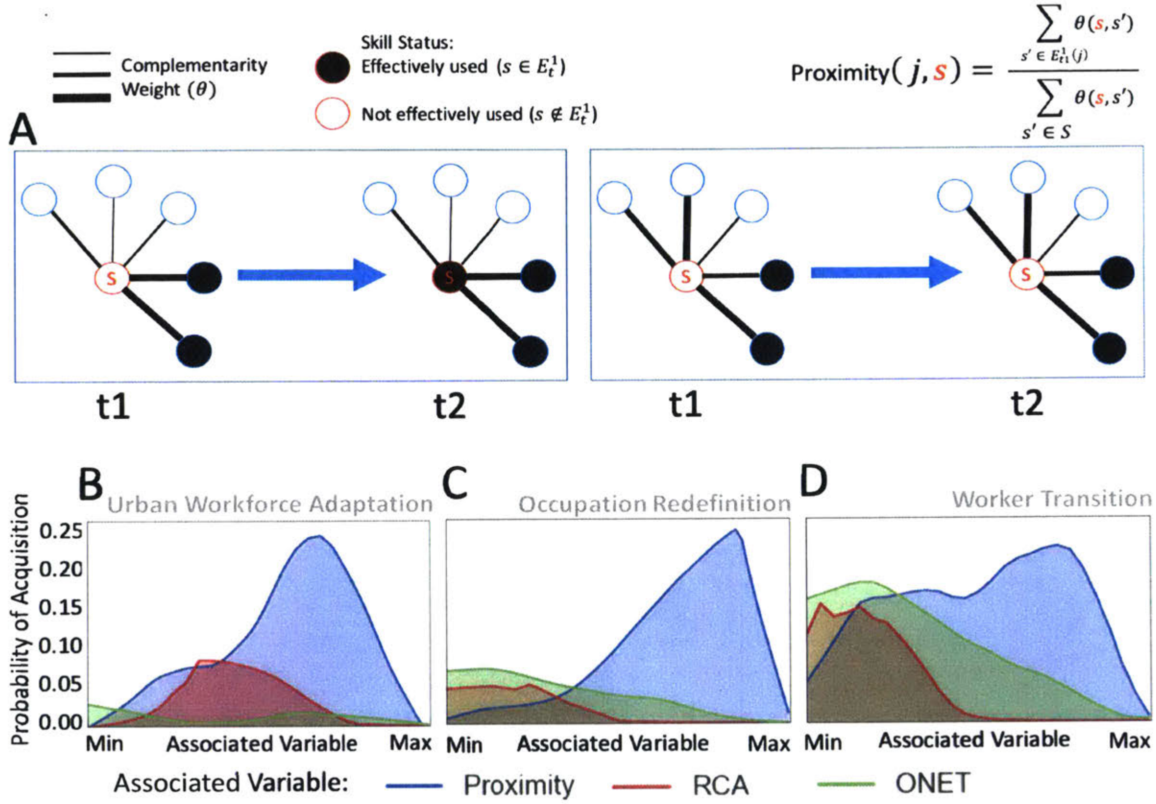


Figure 4-4: Skill proximity predicts worker transitions between occupations, skill redefinition of occupations, and skill acquisition in cities. **(A)** An example demonstrating Skillscape proximity (i.e. $proximity(j,s)$) as a proxy for the connections between effectively used skills and other skills. **(B)** Skills with high proximity to the effectively used skills of an urban labor market in 2010 are more likely to be effectively used by that work force in 2015. **(C)** Skills with high proximity to the effectively used skills of an occupation in 2010 are more likely to be effectively used by that occupation in 2015. **(D)** The effectively used skills of a workers occupation in 2015 are more likely to be effectively used by the workers next occupation in 2016. We provide bar plots including 95% confidence intervals for these probability in subsection C.7.4, and we consider an alternative Receiver Operator Curve analysis in section C.7.

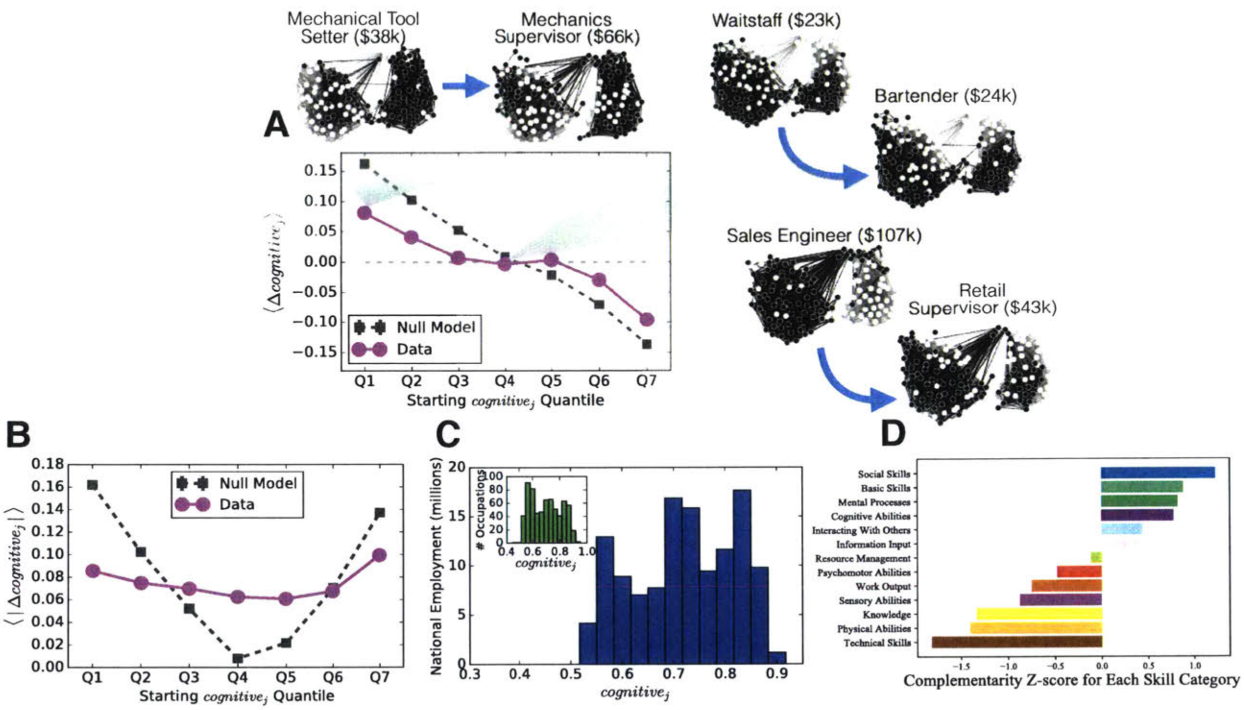


Figure 4-5: The polarized skill network constrains worker mobility. Binning by the $cognitive_j$ of the worker's occupation in 2014 reveals the **(A)** expected cognitive change and **(B)** the expected magnitude of cognitive change when workers change occupations. Random occupation selection is considered as a null model (grey). Standard error bars are provided, but are small. Actual occupation transitions are provided as examples in (A). **(C)** The national distribution of employment by $cognitive_j$ with the distribution of individual occupations as an inset. **(D)** The average complementarity strength that skills possess in each skill category; this measure corresponds to worker mobility because skill proximity is indicative of skill acquisition.

vice versa. In fact, many external factors, such as automation [44, 142] and off-shoring, likely contribute to both effects. Nevertheless, the Skillscape comprehensively explains the polarization of high- and low-skill occupations as a separation between workers with socio-cognitive skills and sensory-physical skills. This high resolution framework for understanding workplace skill requirements provides policy makers with a new explanation for stymied career mobility, while also providing a tool to workers and urban planners trying to traverse the space of workplace skills.

4.5 Discussion

We can summarize the paper's argument as follows: Occupational polarization has been studied using broad subjective occupation categories (i.e. "cognitive" or "physical", and "routine" or "non-routine") that fail to capture the dynamics of workplace skills and decreased labor mobility between low- and high-wage occupations. Rather than subjective occupation categories determined entirely by annual wages, we propose a purely data-driven methodology to map the space of workplace skills based on skill complementarity. The resulting network of skills is polarized in a way that respects stylized facts about occupational polarization; in particular, skill communities distinguish between occupations of different annual wages, thus demonstrating the direct connection between skill polarization and the 'hollowing of the middle class' (see Fig 4-2 A&B and Fig. 4-3).

Beyond the aggregate structure of the skill network (i.e. node communities), we demonstrate that the raw topology of the network corresponds to pathways along which labor dynamics can occur; specifically, we find that the network proximity between skills predicts i) skill adaptation in cities, ii) skill redefinition of occupations, and iii) the changing skill requirements of individual workers as they transition between occupations (see Fig. 4-4). Finally, combining our observations of skill polarization with the labor dynamics determined by the network topology, we hypothesize that worker mobility between physical and cognitive occupations will be constrained, and we provide three types of supporting evidence: i) workers tend to transition between occupations relying on the same skill set, ii) workers are unable to switch away from occupations relying equally on cognitive and physical labor, and iii) this constraining effect is reflected in national employment statistics (see Fig. 4-5). Interesting future work might utilize older sources for skills data, such as the Dictionary of Titles, in combination with our methodology to examine the larger temporal dynamics of skill polarization and their consequences on labor.

While our methods provide more texture to changing labor demands, they have some limitations. Firstly, while the O*NET database facilitates the improved resolution of our model, the taxonomy of O*NET skills may not capture the real-time dynamics of skill categories. For example, consider that a job listing for a software developer in the 1990's may only require "programming" skill while modern listings might require specific types of programming skill, including proficiency in Hadoop, Java, or Python as examples. The O*NET database may miss this change in skill specificity until the taxonomy of skill categories are explicitly updated. External data sources, such as LinkedIn, provide user-defined skills that may allow the future study of skill category dynamics–though they suffer from being non-representative.

Secondly, our analysis provides evidence that cities, occupations, and individual workers leverage the complementarity between skills to navigate changing labor demands and to facilitate career mobility. While our methods provide a data-driven view of the structure underlying these dynamics, they do not account for general market equilibrium dynamics that accompany changing skill demands, and our results demonstrate the need for refined theoretical work that incorporates the granularity of specific workplace tasks and skills. For example, how would the advent of new technology that performs a specific workplace skill change the skill network? And how does the relative cost of capital equipment play into decisions to retrain workers or purchase software or hardware? Answering these types of questions requires knowledge of other mechanisms, such as demand elasticity or capital availability, in addition to knowledge about the skill's location in the skill network. Nevertheless, we hope our framework inspires further investigation into how skill structure dynamics interact with economic equilibrium dynamics studied in traditional models.

Chapter 5

The hidden constraints on worker mobility: how workplace skills determine a worker's next move

5.1 Motivation

Rising economic inequality challenges today's workers and policy makers to find new pathways to career success. However, descriptions of job polarization as a divide between low-skill and high-skill labor are empirically vague. Rather, specific skill requirements determine a worker's employability and mobility between labor markets. Therefore, new methodology is required to advance labor theory to empirical insights that capture the complexity of workplace skills. Here, we operationalize the granular skill requirements of U.S. occupations to predict the career mobility and spatial mobility of workers. Using unsupervised clustering techniques from network science, we reveal the structure of job polarization from shared skill requirements between occupations. Shared skills predict the flow of workers between job titles indicating that skill matching is essential to a worker's employability. Since employment opportunities greatly influence worker relocation, we combine the job network with urban employment distributions to find that densely connected economies exhibit greater flows of enplaned passengers and census migration. Combined, our results empirically connect specific workplace skills to the career and spatial mobility of U.S. workers.

5.2 Background: Skills and the Rungs of the Career Ladder

How do workers advance up the career ladder, and how can policy makers maximize employment opportunities in their communities? As automation [16, 98, 142], off-shoring [83], and globalization [22] reshape U.S. labor, researchers and policy makers must find new empirical methods to forecast changing labor demands. Improved forecasts will allow local officials to enhance career mobility for local workers through better training programs and increased connectedness with regional and national economies [77, 90, 177].

How do worker's obtain employment and navigate their careers? Education determines a worker's entry into the workforce and workers with a head-start may advance further in their careers [129]. However, increasing wealth disparity and decreased absolute income mobility [68] limit general access to higher education [126] and decrease the education's financial returns [5, 18]. Workers instead navigate their careers through social connections [103, 108] and their skills, knowledge, and abilities [10, 12, 53, 156].

Accordingly, existing studies use *skills* to explain U.S. job polarization [2,5,11,23,69,135,179] as a divide between high-skill and low-skill workers [7, 24, 25, 135]. However, these broad labor categories—or even refinements to cognitive and physical or routine and non-routine labor [7]—may not be specific enough to resolve job seeker dynamics. For example, civil engineers and medical doctors are both highly-educated, well-paid, cognitive, non-routine occupations, but the skills required by each occupation are largely non-transferable. As another example, U.S. bank tellers saw increasing national employment with the increased adoption of automated teller machines (ATMs) in part because of the occupation's shift towards reliance on social skills [44] which remain difficult to automate [53, 78, 142]. The differences between occupations and the different forces with which they compete (e.g. different technologies) are most easily understood when we consider occupations as abstract bundles of skills [24, 53].

Just as skills differentiate occupations, urban labor markets may too be characterized according to their employment of different workplace skills [136]. Granular skills data combined with urban employment statistics may explain career mobility within a city and spatial mobility between cities. Specifically, a city that supports many skill sets tends to offer greater wealth and more diverse employment opportunities [47,51,105,136,180,218]. Densely (or "tightly") connected urban labor markets tend to be more productive, innovative [95, 171], and resilient to economic shocks [158, 192, 193, 193–196].

Similarly, city pairs with many shared employment opportunities may have greater potential for worker migration. However, while employment opportunities are the leading factor in urban migration [1, 48, 206], they remain absent from traditional migration models [35, 36]. Since well connected labor markets are more resilient and workers' skills determine their career mobility, the connections between urban workforces may inform models for the spatial mobility of workers as well. In turn, policy makers could use this information to identify synergistic or competitive labor markets as they seek to maintain or grow employment opportunities.

In the remainder of this chapter, we employ data-driven techniques to identify the characteristic skill requirements of occupations from a highly granular taxonomy of workplace tasks and skills. We show that occupations that share characteristic skills exhibit greater flows of workers between them. Since shared skill sets undergird career mobility, we construct a job network from skill similarity scores that is highly polarized and respects stylized facts of job polarization according to unsupervised clustering techniques. In combination with employment distributions in U.S. cities, we demonstrate how workplace skills can inform models for workers' spatial mobility from flight data and migration rates.

5.3 Results: How Skills Underpin Aggregate Labor Trends

5.3.1 Skill Similarity Determines Job Polarization

Can the skill requirements of occupations explain job polarization in the U.S.? To answer this question, we leverage the O*NET skills database from the U.S. Bureau of Labor Statistics to perform an unsupervised data-driven analysis of occupations and workplace skills. The O*NET skills data results from nationally representative worker surveys in each of over 700 different job titles detailing the importance of over 230 examples of relevant workplace knowledge, abilities, and skills (henceforth, "skills"). We denote the importance of skill $s \in S$ to occupation $j \in J$ using $onet_{j,s} \in [0,1]$ where $onet_{j,s} = 1$ identifies an essential skill requirement for that occupation and $onet_{j,s} = 0$ indicates an irrelevant skill.

Ubiquitous workplace skills, such as Communication with Peers, can dampen occupations' distinguishing features. Following previous work [10], we address this issue using revealed comparative advantage [111, 115, 117, 149] (i.e. location quotient [124]) to normalize skill data according to

$$rca_{j,s} = \frac{onet_{j,s} / \sum_{s' \in S} onet_{j,s'}}{\sum_{j' \in J} onet_{j',s} / \sum_{j' \in J} \sum_{s' \in S} onet_{j',s'}}. \tag{5.1}$$

Skills with $rca_{j,s} > 1$ are relatively more important to the average worker in occupation j than would be expected on aggregate. We say these skills are *characteristic* of that occupation and use $I(j,s) = 1$ if $rca_{j,s} > 1$ and $I(j,s) = 0$ otherwise as an indicator function for the characteristic skills of j. We provide visualizations of the raw O*NET data and transformed skill values in section D.1.

How do skill requirements relate occupations? To investigate, we measure the abundance of shared characteristic skills of occupation j and j' according to their Jaccard similarity given by

$$skillsim(j,j') = \frac{\sum_{s \in S} I(j,s) \cdot I(j',s)}{\sum_{s \in S} \big(I(j,s) + I(j',s) - I(j,s) \cdot I(j',s)\big)} \tag{5.2}$$

Jaccard similarity has been used in a variety of applications including natural language processing [165], marine biology [145], and genetics [146].

We compare every pair of occupations to create a map of U.S. occupations as a job network (see Fig. 5-1A). Occupations are represented as nodes which are connected by weighted links according to the skill similarity score for that occupation pair. Nodes are colored according to their node community as identified by Louvain community detection [50]. This unsupervised clustering technique has been used in various problem areas including neuroscience [187, 205], transportation research [32], social science [39], business/management research [79], climatology [86], and cybersecurity [71].

The job network is highly polarized and separates occupations requiring cognitive skills from occupations requiring physical skills according to occupations' cognitive skill fraction [10] (see Fig. 5-1C). These occupation types further distinguish occupations according to annual wages (see Fig. 5-1D) and susceptibility to automation from computers using estimates from two separate studies [16, 98] (see Fig. 5-1E&F). The comparisons in Figure 5-1C-F are each statistically significant (i.e. $p_{value} < 10^{-5}$) according to the t-test and the Kolmogorov-Smirnov statistic. This polarization is not a product of randomness as the network's modularity is significantly larger than randomly rewired networks (see section D.2).

The distinction between cognitive and physical workers is spatially embedded in the U.S. as well. The relative abundance of cognitive workers is not homogeneous across U.S. cities—even within the same regional economy (see Fig. 5-1B). While there is statistically significant spatial autocorrelation ($p_{value} < 0.01$), the effect is modest (Moran's $I = 0.14$). Therefore—while regional economies may be a factor—further investigation is required to understand the differences in cognitive employment across cities and how changing workforce composition (e.g. from worker migration) contribute to these results.

5.3.2 The Job Network Predicts Career Mobility

The 2010 Nobel Prize in Economics was awarded for research highlighting the importance of workers' skills when applying for employment opportunities [156]. Accordingly, we expect workers will have greater transition rates between occupations pairs with shared characteristic skills. To test this hypothesis, we compare $skillsim(j, j')$ to worker transition rates between occupation pairs using data from the U.S. Community Population Survey (see section D.3 for more details).

Job polarization studies suggest that all skills do not equally facilitate career mobility [5, 7] and skill matching [156] suggests that job seekers must satisfy specific skill requirements to obtain employment. Therefore, each individual O*NET skill may contribute differently to workers'

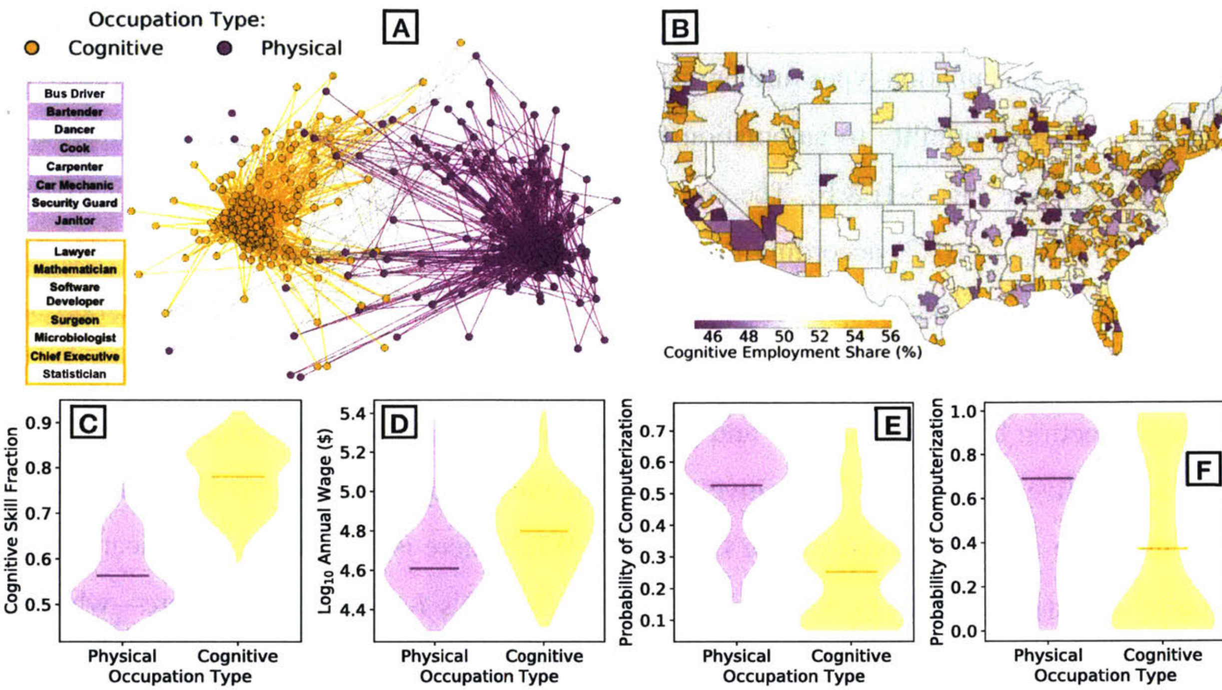


Figure 5-1: The polarized U.S. job network distinguishes cognitive and physical labor. **(A)** Nodes represent U.S. occupations which are connected by weighted edges determined by their skill similarity (i.e. $skillsim(j, j')$). Colors indicate node communities, or *occupation types*, as determined by Louvain community detection [50]. Example job titles from each occupation type are provided in tables. This visualization has filtered edges to include the occupation pairs with the strongest skill similarity scores ($skillsim(j, j') > 0.5$). All analysis, including community detection, was performed using the unfiltered network. **(B)** A map of U.S. cities colored by their cognitive employment share. **(C)** Using socio-cognitive and sensory-physical skill categories from [10], we find that one occupation type (identified by yellow in (A)) relies more strongly on social and cognitive skills. Hence, we refer to these occupations as *cognitive* occupations and refer to the other occupation type (identified by purple in (A)) as *physical* occupations. **(D)** Cognitive occupations tend to earn higher average annual wages. Using occupational computerization estimates from **(E)** [98] and from **(F)** [16], cognitive occupations are less susceptible to automation in the foreseeable future. Each comparison in (C)-(F) is significantly different according to a two-tailed t-test and the KS-statistic ($p_{value} < 10^{-5}$).

transition rates. So, we employ a LASSO regression model of the form

$$\log_{10}(\#\text{ worker transitions}_{j,j'}) \sim \sum_{s \in S} \beta_s \cdot \gamma(j, j', s), \tag{5.3}$$

where β_s signifies a regression coefficient and $\gamma(j, j', s)$ are variables representing the summand for each skill in equation (5.2). This skill similarity model weights the contribution of each individual O*NET skill in combination with all other skills, but is insufficient for causal inference due to skill pair co-linearity. All variables are centered and standarized prior to analysis. Although our sample size is large (i.e. we observe transitions for 2,000 job pairs), we measure model performance from the mean variance explained of 100 independent trials of 10-fold cross-validation (i.e. 1,000 R^2 values). This treatment both reduces the each model's degrees of freedom and provides a rigorous measure for performance that controls for model overfitting.

Figure 5-2A explores the performance of this model as we iteratively include O*NET skills. Skills are included in order of their individual ability to predict worker transition rates. The model's performance increases sharply with the inclusion of the first 50 skills. Including additional skills yields diminishing returns to performance. The complete model including all skills explains around 15% of the variance in job transitions on average.

How do these results compare to standard labor models and, in particular, does the inclusion of raw O*NET skills outperform aggregated skill categories, such as cognitive and physical skill? As a baseline, we consider a random mixing model based on national employment by occupation according to $\log_{10}(employment_j \cdot employment_{j'})$. Following previous job polarization studies [7], we consider an additional baseline model that simplifies skill categories as cognitive, physical, routine, and non-routine labor in combination with national employment, and yet another baseline model combining national employment with O*NET Task Groups [136] (see section D.4 for details). Finally, we consider national employment in combination with the O*NET skills model described above. Interestingly, aggregated skill categories do not add to the performance of the national employment model alone—which performs comparably well to the model using only raw O*NET skills at around 15% variance explained. However, combining the O*NET skills model with national employment statistics significantly improves model performance according to the t-

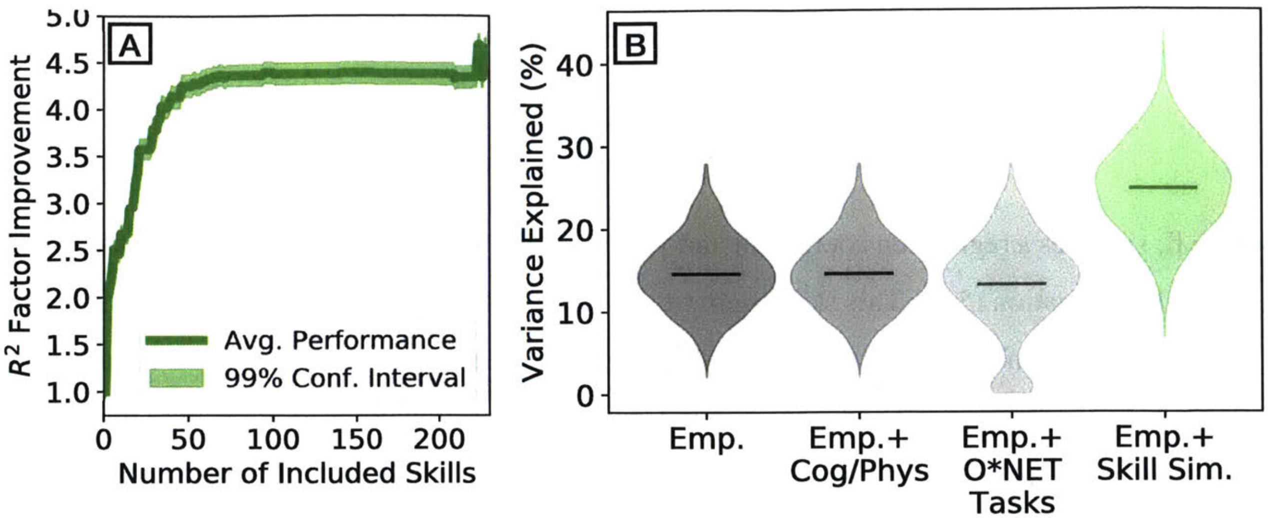


Figure 5-2: Skill similarity predicts worker flows between occupation pairs. **(A)** Incrementally including O*NET skills in the skill similarity model (x-axis) yields diminishing improvements in the model's ability to capture variance in worker flows between job titles according the average R^2 values after 100 trials of 10-fold cross-validation. **(B)** For each model (color), we provide the distribution of R^2 values after 100 trials of 10-fold cross-validation. Solid lines indicate average model performance. Probability density functions are estimated using a Gaussian kernel density estimator (bandwidth of 0.25%). The model combining national employment and weighted skill similarity scores significantly outperforms baseline models.

test and the Kolmogorov-Smirnov statistic, and accounts for 25% of the variance in job transitions on average.

5.3.3 Tightly Connected Urban Economies Support Greater Spatial Mobility

Although employment opportunities are the leading factor in urban migration [1], traditional models for spatial mobility ignore synergies between labor markets [35, 36]. On the other hand, skills play a crucial role in the the job matching process [156] and granular skills data improves predictions of career mobility (see Fig. 5-2). Therefore, cities that support similar workers and provide opportunities for career advancement through skill similarity may experience greater flows of workers moving between them. To test this idea, we adapt a measure for "tightness" within urban labor markets [158, 192, 193, 193–196] to instead measure the tightness of connections *between* urban labor markets.

First, we identify the characteristic occupations of each U.S. city from urban employment distributions using revealed comparative advantage (similar to eq. (5.1)), and we use $I(c,j)$ as an indicator function for the characteristic occupations of city c. In Figure 5-3A, we project example cities onto the job network by coloring characteristic occupations and the connections between them. A city pair's number of shared characteristic occupations (i.e. $overlap(c,c') = \sum_{j \in J} I(c,j) \cdot I(c',j)$) offers a simple measure for the similarity of their workforces, but does not account for the "tightness" between shared opportunities. This overlap can also be projected onto the job network by coloring shared characteristic occupations and the connections between them (see Fig. 5-3B).

Following previous work, the density of connections within a labor market explain more of its economic benefits than the size of the workforce alone. Analogously, the density of connections between occupations that characterize two cities may indicate a stronger relationship with greater career mobility for workers in either city. As an illustration, two cities may share many characteristic occupations (i.e. $overlap(c,c')$ is large) simply because large cities support greater occupational diversity [218]. But, if these cities support many workers within the same industry of a shared occupation, then that is a stronger indication that those workers in one city might find viable employment opportunities in the other city. We approximate this effect with a measure for the *job tightness* between two cities according to

$$tightness(c,c') = \sum_{s \in S} \sum_{j,j' \in J^2} \frac{\gamma(j,j',s) \cdot \big(I(c,j) + I(c',j)\big)}{2 \cdot \sum_{i,i' \in J^2} skillsim(i,i')}. \tag{5.4}$$

Does job tightness predict spatial mobility? We address this question using 2015 Census Bureau migration statistics (31,500 city pairs) and the number of enplaned passengers that flew between city pairs in 2017 according to the U.S. Bureau of Transportation Statistics (8,700 city pairs). Migration represents a relatively permanent relocation, while flight travel additionally captures temporary relocation.

Similar to worker transitions between occupations, specific skills may differentially contribute to a worker's spatial mobility. For example, physical workers may suffer hindered mobility by

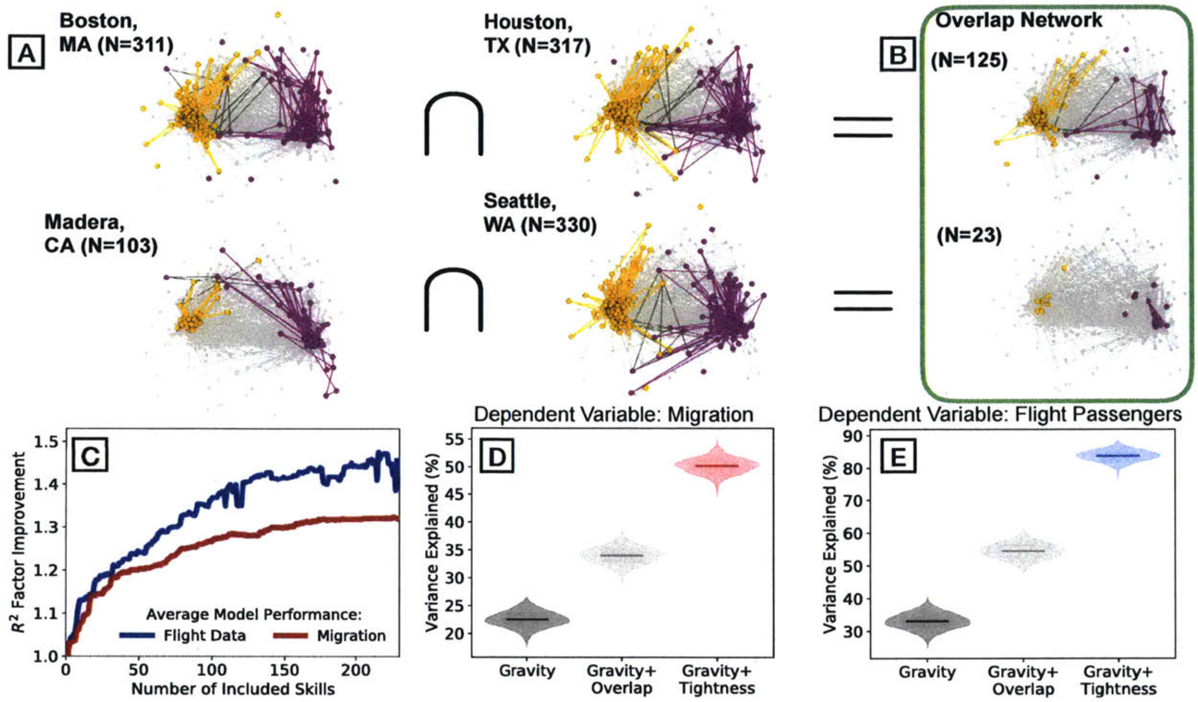


Figure 5-3: The density of connections between urban workforces predicts spatial mobility between cities. **(A)** Example U.S. cities projected on the job network according to their characteristic occupations. The complete job network is provided in grey. The characteristic occupations of each city are colored along with the edges connecting those occupations. In the title of each network, N represents the number of highlighted nodes. **(B)** Similar to (A), overlap networks are created from the occupations that are characteristic in both cities. **(C)** Similar to Fig. 5-2(A), job tightness models yield diminishing improvements with the incremental inclusion of additional O*NET skills. The 99% confidence intervals are provided, but are small. **(D)** Similar to Fig. 5-2(B), we compare each model's ability to predict migration flows between city pairs. Distributions represent R^2 values across 100 trials of 10-fold cross-validation (i.e. each distribution represents 1000 R^2 values). The size of the overlap (i.e. the number of shared characteristic occupations) between city pairs yields a significant improvement in performance over the baseline gravity model alone, but job tightness model, which incorporates the density of connections between shared characteristic occupations, yields the best performance. Probability density functions are estimated using a Gaussian kernel density estimator (bandwidth of 0.25%). **(E)** Similar to (D), the gravity model plus the job tightness model yields the best predictor of the number of enplaned passengers between city pairs.

comparison to in-demand cognitive workers. Or, alternatively, highly skilled workers may have difficulty finding employment opportunities that maximize returns on their training while respecting their social obligations (e.g. the "two body problem" in faculty hiring [183]). Therefore, we again employ a LASSO regression model of the form

$$\log_{10}(flow_{c,c'}) \sim \sum_{s \in S} \beta_s \cdot \Gamma(c,c',s) \tag{5.5}$$

where $\Gamma(c,c',s)$ are variables representing the summand for each skill in equation (5.4). We use the mean variance explained of 100 independent trials of 10-fold cross-validation to assess model performance. All variables are centered and standardized prior to analysis. Similar to before, the iterative inclusion of O*NET skills yields diminishing returns when predicting either type of spatial mobility (see Fig. 5-3C) but, nevertheless, demonstrates the predictive power of granular skill categories.

Do skill requirements and shared characteristic occupations improve spatial mobility models? We investigate by comparison to the gravity model [28,35,73,101] (see subsection D.5.2 for further details). Total employment and the spatial distance between cities in the traditional gravity model captures 23% of the variance in migration rates and 33% of the variance in passenger flight data on average (see Fig. 5-3D&E). Augmenting the gravity model with the overlap of characteristic occupations improves performance to 35% variance explained for migration and 55% variance explained for flight data on average. However, the best performance is achieved when we instead augment the gravity model with the job tightness calculation (i.e. eq. (5.5)) which achieves 50% variance explained for migration rates and 85% variance explained for passenger flight rates on average (see section D.5 for more model details).

How can job tightness inform urban policy makers on the future of work? The rise of "superstar" cities that will develop and support the work of the future, and he suggested that geographic mobility rates are changing to reflect this new reality [20]. Therefore, policy makers can benefit from improved predictions of workers' spatial mobility using job tightness to understand their city's place in the developing future of work.

In addition to improved predictions, the job tightness model highlights the occupations and

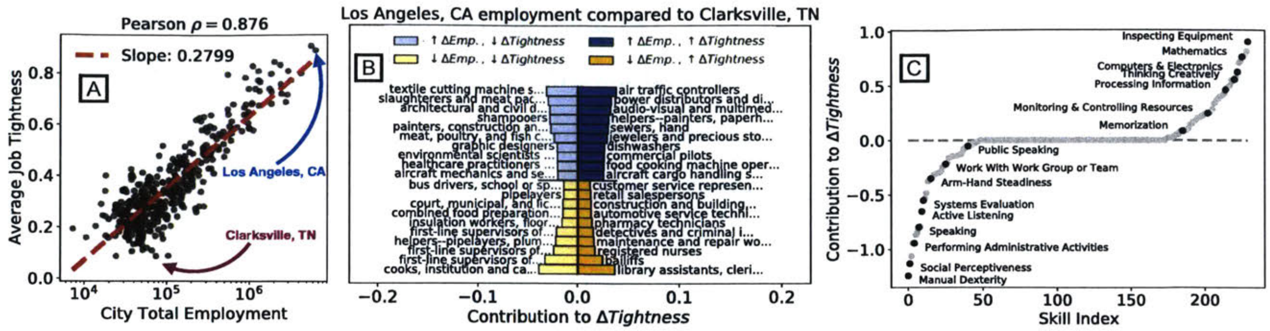


Figure 5-4: The job tightness calculation exposes the occupations and skills that connect cities to the larger economy. **(A)** Larger cities tend to have greater average job tightness with other U.S. cities. **(B)** How differences in employment for specific occupations contribute to the overall difference in average job tightness for Los Angeles, CA and Clarksville, TN. An occupation may be characteristic of Los Angeles and not characteristic of Clarksville ($\uparrow \Delta Emp.$), or vice versa ($\downarrow \Delta Emp.$), and contribute positively ($\uparrow \Delta Tightness$) or negatively ($\downarrow \Delta Tightness$) to the difference in average job tightness. Employment differences for occupations listed on the left increase $\Delta Tightness$, while occupations on the right decrease the difference. **(C)** How differences in employment for workers with specific O*NET skills impact $\Delta Tightness$ for Los Angeles, CA and Clarksville, TN.

skills that increase a labor market's strength of connection to the larger economy. For example, the urban labor markets with the on-average tightest connection to other cities tend to be larger cities (see Fig. 5-4A). We can further decompose the difference in the average job tightness of two cities (denoted $\Delta Tightness(c,c')$) to quantify the impact of employment by individual occupations and by individual skill (see section D.6 for calculation). In particular, differences in characteristic occupations (given by $\Delta Emp(c,c',j) = I(c,j) - I(c',j)$) can be positive or negative in sign and either increase or decrease the difference in average job tightness. These contributions to the difference depend on the occupation's connection to other occupations on the job network and the inferred spatial mobility of workers with the skills required by that occupation (i.e. β_s learned from the spatial mobility regressions). Here, we report results by setting β_s equal to the coefficients from the job tightness model regression for passenger flight data.

As an example, we compare the the average job tightness of Los Angeles, CA, which is a large tightly-connected city, with Clarksville, TN, which is a small weakly-connected city (see Fig. 5-4B&C). Figure 5-4B displays a *job shift* quantifying how differences in employment for individual occupations contribute to the overall difference in average job tightness scores. Differences in employment for occupations on the left increase $\Delta Tightness$, while occupations on the right actually go against the aggregate trend and decrease the difference in average job tightness. Similarly, Figure 5-4C quantifies how employment differences for workers with specific O*NET skills contribute to $\Delta Tightness$. Additional comparisons of city pairs are provided in section D.6.

5.4 Discussion

Policy makers are grappling with job polarization [18] to improve employment opportunities for their constituents. Armed with only broad subjective labor categories, like cognitive/physical or routine/non-routine [7], it remains difficult to forecast the future of work because several drivers of labor trends—including automation [16, 98, 142], offshoring [83], and globalization [22]—impact employment by altering demand for specific workplace skills [24, 37, 41, 53]. Rather than coarse labor categories, our analysis incorporates a granular taxonomy of workplace tasks and skills that may capture these microscopic dynamics while also describing macroscopic employment trends.

For instance, our unsupervised data-driven analysis of occupational skill requirements produces a polarized job network that accurately reflects U.S. job polarization (see Fig. 5-1) as a divide between cognitive and physical occupations. This divided job network further distinguishes automatable occupations from non-automatable occupations and distinguishes high-wage occupations from low-wage occupations. Combined, these results directly connect the job network's occupation types to the "high-" and "low-skill" occupations used to describe the "hollowing of the middle class" [7, 24].

While the aggregate network structure describes job polarization, the specific connections between occupations on the job network capture refined labor dynamics as well. Following job matching theory [156], our results suggest that specific skills play a critical role when workers transition between job titles. Specifically, models that leverage occupational skill similarity based on granular workplace skills are better predictors of worker flows between occupations than models based on national employment and aggregated labor categories, including cognitive/physical and routine/non-routine employment [7] (see Fig. 5-2). The macroscopic job network polarization combined with the microscopic connections between occupations (i.e. skill similarity scores) highlight potential bottlenecks to career mobility based on workers' skill sets. These findings may inform policy makers as they consider worker retraining programs and identify industries for investment that will create viable employment opportunities for workers in their local labor market.

These considerations focus on dynamics within a single labor market, but a local economy's connection to the larger economy can bolster economic resilience as well [90,116,176,177]. Therefore, policy makers must consider how their local employment connects to other labor markets if they seek to grow their workforce [1,48,206]. Motivated by the critical role of skill matching in workers' job transitions, our analysis reveals how skills data and employment distributions can accurately predict migration rates and airline passenger flows between U.S. city pairs based on the *tightness* of their shared characteristic occupations embedded on the job network (see Fig. 5-3). This approach to modelling spatial mobility offers the added benefit of transparency. Policy makers can quantify how differences in employment impact their city's connection to the larger economy and compare their city's employment to that of another city (see Fig. 5-4) to test how differences in policy may be impacting overall connectivity. Specifically, differences in characteristic occupations may strengthen or weaken a city's average job tightness compared to another city, and these observations show where a local labor market has a competitive edge and where employment changes could bolster their economic standing. Differences in employment by occupation can be further decomposed into differences in worker skill sets. This added resolution allows policy makers to refine their strategy in consideration of major labor dynamics that impact demand for skills, such as technological change and off-shoring.

This chapter demonstrates how granular skills data enables improved empirical insights to important labor dynamics including career mobility and the spatial mobility of workers. However, our unsupervised methodology offers only weak controls for the colinearity between pairs of workplace skills which makes causal inference extremely difficult. These examples may represent redundancies in the O*NET skill taxonomy or describe complementary skill pairs. Nevertheless, the use of granular skills data has the potential to improve policy makers' forecasts of changing labor demands in order to bolster the economic resilience and employment opportunities in their local labor market.

Chapter 6

The evolution of citation graphs in artificial intelligence research

Note: This chapter summarizes research from [96].

6.1 Motivation

As artificial intelligence (AI) applications see wider deployment, it becomes increasingly important to study the social and societal implications of AI adoption. Therefore, are AI research and the fields that study social and societal trends keeping pace with each other? Here, we use the Microsoft Academic Graph to study the bibliometric evolution of AI research and its related fields from 1950 to today. Although early AI researchers exhibited strong referencing behavior towards Philosophy, Geography, and Art, modern AI research references Mathematics and Computer Science most strongly. Conversely, other fields, including the social sciences, do not reference AI research in proportion to its growing paper production. Our evidence suggests that the growing preference of AI researchers to publish in topic-specific conferences over academic journals and the increasing presence of industry research poses a challenge to external researchers, as such research is particularly absent from references made by social scientists.

6.2 Background: Artificial Intelligence and Society

Today's artificial intelligence (AI) has implications for the future of work [59], the stock market [55, 131], medicine [118, 211], transportation [4, 52], the future of warfare [188], and the governance of society [72, 181, 210]. On one hand, AI adoption has the positive potential to reduce human error and human bias [151]. As examples, AI systems have balanced judges towards more equitable bail decisions [132], AI systems can assess the safety of neighborhoods from images [159], and AI systems can improve hiring decisions for board directors while reducing gender bias [85]. On the other hand, recent examples suggest that AI technologies can be deployed without understanding the social biases they possess or the social questions they raise. Consider the recent reports of racial bias in facial recognition software [63, 64], the ethical dilemmas of autonomous vehicles [52], and income inequality from computer-driven automation [8, 95, 98].

These examples highlight the diversity of today's AI technology and the breadth of its application; an observation leading some to characterize AI as a general purpose technology [59, 134]. As AI becomes increasingly widespread, researchers and policy makers must balance the positive and negative implications of AI adoption. Therefore, we ask: how tightly connected are the social sciences and cutting-edge machine intelligence research?

Here, we employ the Microsoft Academic Graph (MAG) to explore the research connections between AI research and other academic fields through citation patterns. The MAG data offers coverage for both conference proceedings, where AI papers are often published, and academic journals, where other fields prefer to publish. Although early AI research was inspired by the several other fields, including some social sciences, modern AI research is increasingly focused on engineering applications—perhaps due to the increasingly central role of the technology industry. Further, the most central research institutions within the AI research community are increasingly based in industry rather than academia.

6.3 Modern Artificial Intelligence Research

The effort to create human-like intelligence has dramatically advanced in recent decades thanks to improvements in algorithms and computers. However, engineering the entirety of human intelligence has proven difficult. Instead, progress has come from engineering specific human capabilities. While we often use *artificial intelligence* today in reference to machine learning, the meaning of *AI* has fluctuated in the last 60 years to variably emphasize vision, language, speech, and pattern recognition.

To study the nature of AI research, we use the Microsoft Academic Graph (MAG) to identify relevant Computer Science (CS) subfields from the citations of academic publications from 1950 to 2018. The MAG uses natural language processing, including key word analysis, to identify the academic field of each publication according to a hierarchy of academic fields. This data has been particularly useful for studying bibliometric trends in CS [65, 84, 119, 201]. Our analysis relies strongly on the MAG's field of study classifications and, thus, our analysis is potentially limited in its accounting of more specific research areas within CS and within AI-related fields. This data enables us to study the paper production and referencing behavior of different academic fields. For example, CS has risen to the fourth most productive academic field according to annual paper production (see Fig. S1) with AI being the most prominent subfield of CS in recent decades [88] (see also Fig. 6-1d).

To identify the CS subfields that are most relevant to AI research, we construct a citation network using all CS papers published within each decade from 1950 to 2018. We consider CS subfields to represent AI research if they are strongly associated with Artificial Intelligence, which is itself a CS subfield, throughout a significant proportion of the time period under analysis. Examples include Computer Vision, Machine Learning, and Pattern Recognition. Interestingly, Natural Language Processing (NLP), which is colloquially thought of as a specific problem area in AI [189], is strongly associated with AI research prior to the mid 1980s, after which NLP becomes more strongly associated with Information Retrieval and Data Mining for text-based data (see Fig. 6-1a-c,e). In the remainder, we use papers published in Artificial Intelligence, Computer Vision, Machine Learning, Pattern Recognition, and Natural Language Processing to approximate

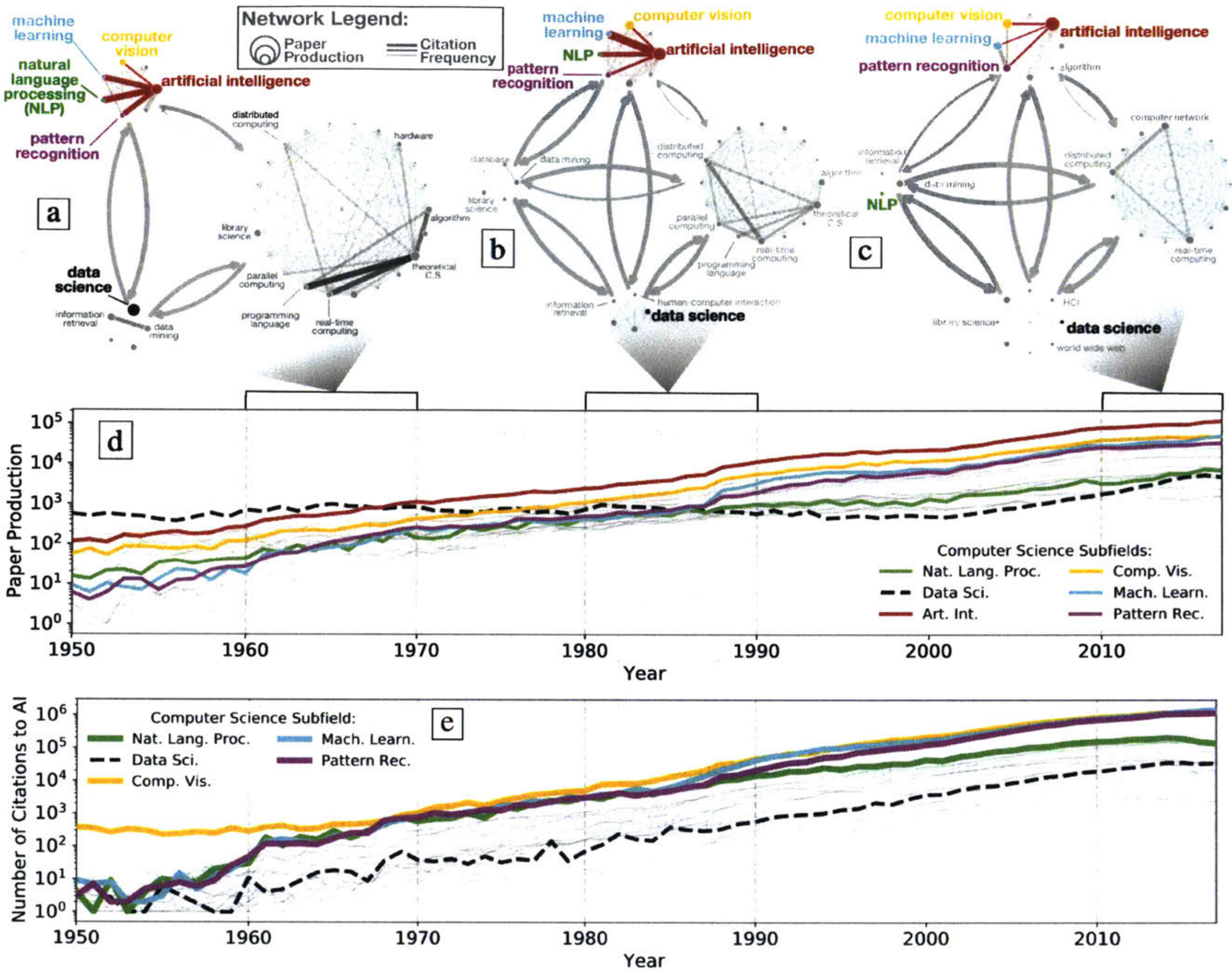


Figure 6-1: Citation patterns among Computer Science subfields identify areas of artificial intelligence-related research. We examine the rate of citations between Computer Science (CS) subfields based on journal and conference publications from three different decades: **(a)** the 1960's, **(b)** the 1980's, and **(c)** 2010's through 2017. For each network, nodes (circles) correspond to CS subfields according to the Microsoft Academic Graph data, and node size corresponds to the number of papers published in each subfield (note: the same paper may belong to multiple subfields). The width of links connecting nodes corresponds to the number of references made between papers published in those subfields. After constructing the complete network, we apply topological clustering [164] and report the number of citations made between these clusters using weighted arrows. Networks with labels for each subfield are provided in section E.2. **(d)** Annual paper production by CS subfield. Subfields related to artificial intelligence are colored, as well as data science (black) because of its notable decline in relative paper production. **(e)** The annual number of references from papers in each CS subfield to papers in the AI subfield, and vice versa (i.e. $(subfield \rightarrow AI) + (subfield \leftarrow AI)$).

AI research from the 1950s to today.

The paper production of CS subfields have varied over the last half century. For example, Data Science has gradually diminished in relative paper production and Theoretical Computer Science has been replaced by increased focus on Real-Time and Distributed Computing. However, AI-related research areas have experienced steadily growing paper production since 1950 and account for the largest share of paper production in CS today (see Fig. 6-1d).

6.4 Shaping the Study of Intelligent Machines

Just as early myths and parables emphasized the social and ethical questions around human-created intelligence [130, 148, 175], today's intelligent machines provide their own interesting social questions. For example, how responsible are the creators, the manufacturers, and the users for the outcomes of an AI system? How should regulators handle distributed agency [93, 210]? How will AI technologies reduce instances of human bias? As AI systems become more widespread [59, 134], it becomes increasingly important to consider these social, ethical, and societal dynamics to completely understand the impact of AI systems [19, 72, 174, 181, 210]. However, the developers of new AI systems are often separate from the scientists who study social questions. Therefore, we might hope to see increasing research interest between these fields of study and AI.

To investigate, we study the association between various academic fields and AI research through the referencing relationship of papers published in each academic field. External fields reference AI research for a number of reasons. Some fields, such as Engineering or Medicine, reference AI research because they use AI methods for optimization or data analysis. Other fields, like Philosophy, reference AI research because they explore its consequences for society (e.g. moral and/or ethical consequences). Similarly, AI researchers reference other fields, such as Mathematics or Psychology, because AI research incorporates methods and models from these areas. AI researchers may also cite other fields because they use them as application domains to benchmark AI techniques.

In Figure 6-2a&c, we examine the share of references made from AI papers to other fields, and from papers published in other fields to AI. The reference share from academic field A to field B

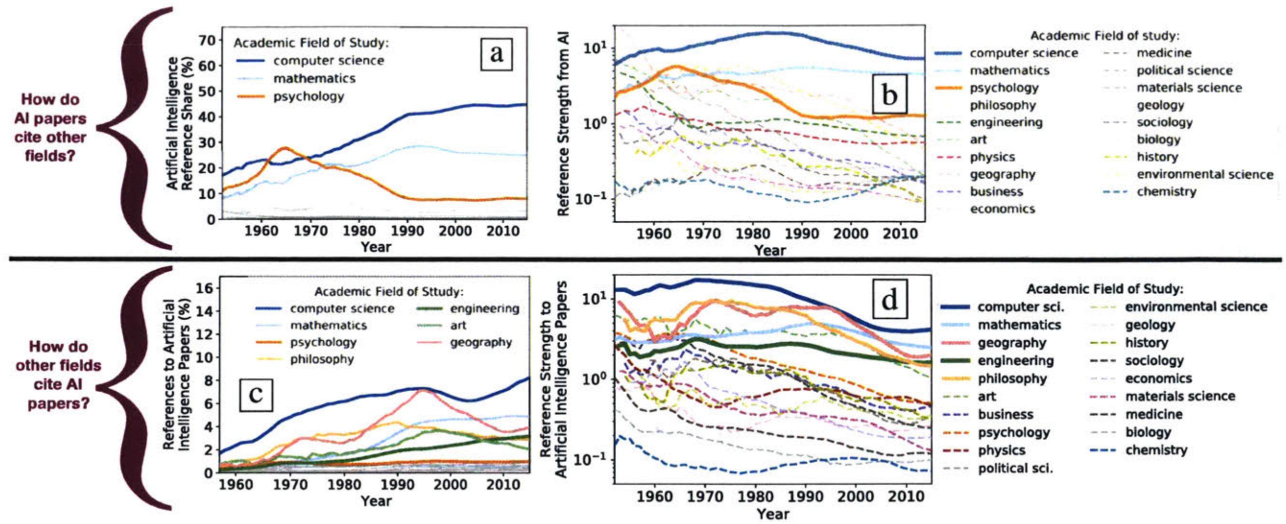


Figure 6-2: The referencing strength between artificial intelligence and other sciences is declining. **(a)** The share of references made by artificial intelligence (AI) papers in each year to papers published in other academic fields. **(b)** The reference strength (see eq. (6.2)) from AI papers to papers published in other academic fields. **(c)** The share of references made by each academic field to AI papers in each year. **(d)** The reference strength from each other academic field to AI papers in each year. All lines are smoothed using a 5-year moving average. In (b) and (d), dashed lines indicate academic fields exhibiting lower reference strength than would be expected under random referencing behavior in 2017.

according to

$$share_{year}(A,B) = \frac{\#\text{ ref. from } A \text{ papers to } B \text{ papers in } year}{\#\text{ ref. made by } A \text{ papers in } year} \tag{6.1}$$

controls for the total paper production of the referencing field over time, and has been used in other bibliometric studies [200]. However, temporal changes in reference share may be explained by paper production in the referenced field; therefore, we consider another measure that also controls for the total paper production in the referenced field as well (see Fig. 6-2b&d). We calculate the *reference strength* from field A to field B according to

$$\begin{aligned} strength_{year}(A,B) &= \frac{\left(\frac{\#\text{ ref. from } A \text{ papers to } B \text{ papers in } year}{\#\text{ ref. made by } A \text{ papers in } year}\right)}{\left(\frac{\#\text{ of } B \text{ papers from 1950 to } year}{\#\text{ of papers from 1950 to } year}\right)} \\ &= \frac{\left(\text{Reference share from } A \text{ to } B \text{ in } year\right)}{\left(B\text{'s share of all papers from 1950 to } year\right)}. \end{aligned} \tag{6.2}$$

A reference strength of $strength_{year}(A,B) > 1$ indicates that the rate of referencing from field A to field B is greater than would be expected by random referencing behavior given the number of published papers in field B. Both reference share and reference strength capture the aggregate referencing behavior between fields of study, but these calculations may obfuscate other dynamics from sub-communities within larger academic fields.

Prior to 1980, AI research made relatively frequent reference to Psychology in addition to Computer Science and Mathematics (see Fig. 6-2a). Controlling for the paper production of the referenced fields, we find that early AI's reference strength towards Philosophy, Geography, and Art were comparable to the field's strength of association with Mathematics (see Fig. 6-2b) suggesting that early AI research was shaped by a diverse set of fields. However, AI research transitioned to strongly relying on Mathematics and Computer Science soon after 1987 which suggests an increasing focus on computational research.

How important is AI research to other academic fields? Unsurprisingly, Computer Science, which includes all of the AI-related subfields in our analysis, steadily increased its share of references made to AI papers throughout the entire period of analysis (see Fig. 6-2c). Surprisingly, Mathematics experienced a notable increase in reference share to AI only after 1980. Meanwhile, several fields that are not often cited in today's AI research played an important role in the field's development, but may not have reciprocated this interest. For example, Psychology was relatively important to early AI research, but Psychology did not reciprocate as strong of an interest at any point from 1990 on-wards (i.e. $strength(\text{Psychology}, \text{AI}) < 1$ in recent years). Instead, Philosophy, Art, Engineering, and Geography have increased their share of references to AI papers up to 1995. On aggregate, when we control for AI paper production over time, we observe decreasing reference strength towards AI from all external academic fields. This suggests that other fields have difficulty keeping track of increasing AI paper production in recent decades (see Fig. S3). This result may in part be explained by the increased complexity of AI-related research that is not relevant to the study of other scientific disciplines.

6.5 The Consolidation of Artificial Intelligence Research

How do leading research institutions shape AI research? On one hand, the prestige of an academic university can boost the scientific impact of CS publications [154]. On the other hand, although scientific research is often undertaken at universities, major AI advances have emerged from industry research centers as well. For example, the AI start-up DeepMind received recent attention for their AlphaGo project [197] and Google has been acknowledged as a leader in the development of autonomous vehicles [40, 114, 173]. With increased industrial and regulatory involvement, recent work suggests that areas of AI, including deep learning [134], are undergoing a consolidation of research and deployment worldwide. While Computer Science on the whole has become increasingly diverse [172], what can be said about AI research?

If the AI research community is experiencing a consolidation of influence, then what types of citation dynamics might indicate such a phenomenon? We investigate by examining the distribution of AI paper production and the distribution of citations made to AI papers by research institution (see section E.3 for visualization of the distributions by decade). Since 1980, the diversity of AI paper production, authorship, and citations to AI papers across institutions have decreased by 30% according to the Gini coefficient applied to annual distributions (see Fig. 6-3a). Repeating this analysis for other academic fields, we find that this decreasing diversity is not simply a reflection of aggregate academic trends since most other fields of study actually exhibit increasing diversity over time according to these metrics (see section E.5).

This decrease in scientific diversity suggests that notable research "hubs" may be forming (similar to the industry use of deep learning [134]). This type of hierarchical structure can occur when referencing between institutions is well modelled by preferential attachment [29]. If *preferential referencing* explains the citation dynamics within AI research, then the proportion of citations gained by a research institution in each year will be proportional to the institution's total accumulation of citations. Figure 6-3b reports estimates of the slope m for the model

$$\log_{10}(\#\text{ citations}) = m \cdot \log_{10}(\text{cumulative \# of citations}) + b, \tag{6.3}$$

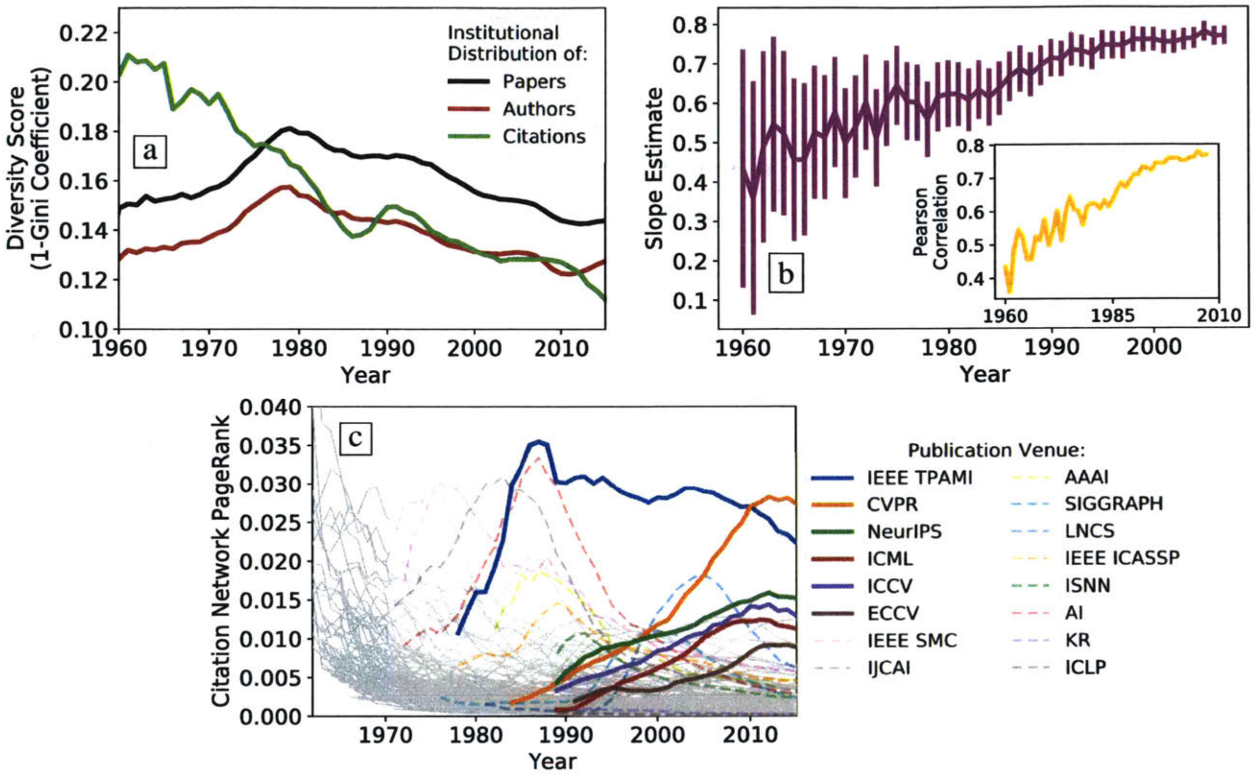


Figure 6-3: Artificial intelligence research is increasingly dominated by only a few research institutions and AI-specific conferences. **(a)** The diversity of the annual distribution of all AI papers (black), AI authors (red), and all citations to AI papers (green) across research institutions according to the Gini coefficient. Example distributions of AI paper production and AI citation share are provided in section E.3. **(b)** To see if preferential attachment explains citation dynamics, we restrict to AI papers with at least one citation and estimate the linear relationship between each research institution's cumulative citation count from 1950 to the institutions citation count in each year (see eq. (6.3)). The model's slope estimation steadily rises throughout the period of analysis to around 0.70 as the model increasingly captures variance in the citation accumulation of institutions according to Pearson correlation (inset). **(c)** The PageRank of each publication venue for AI papers using the number of references from AI papers published in each venue to papers published in each other venue. The lines of notable publication venues are highlighted with color. Dashed lines indicate venues whose PageRank has declined during the period of analysis. In all plots, lines are smoothed using a 5-year moving average. More recent citation results may change as recent publications continue to accumulate citations.

as well as 99% confidence intervals for those slope estimates using linear regression. Both the annual slope estimates and the performance of this model (see inset) rise steadily throughout the period of analysis. Combined, this evidence suggests that preferential referencing may be occurring among AI research institutions.

How have AI publication practices changed over time to enable preferential referencing? To investigate, we calculate the PageRank [169] of each AI publication venue—including both academic journals and conferences—from the references of the AI papers published by each venue in each year (see Fig. 6-3c). Publications venues with larger PageRank are more central to AI research. In the late 1980's, several specific conferences, including the Conference on Computer Vision and Pattern Recognition (CVPR), the Conference on Neural Information Processing Systems (NIPS), and the International Conference on Machine Learning (ICML), rise in prominence, while more general AI conferences, including the National Conference on Artificial Intelligence (AAAI) and the International Joint Conference on Artificial Intelligence (IJCAI), decline in prominence for AI researchers. Meanwhile, very few academic journals maintain high citation PageRank with the exception of the IEEE Transactions on Pattern Analysis and Machine Intelligence (IEEE TPAMI), which remains one of the most central publication venues for AI research.

If preferential referencing is producing research hubs, then which research institutions enjoy a privileged role in the AI research community? To investigate, we calculate the citation PageRank of each institution from the references of the AI papers published by each institution in each year (see Fig. 6-4A). Prior to 1990, the most prominent research institutions were academic, including the Massachusetts Institute of Technology, Stanford University, and Carnegie Mellon University, and included only a few industry-based research institutions, such as Bell Labs and IBM. However, the late 1980's again marks a transition point that reshaped the field. While universities dominate scientific progress across all academic fields [139], industry-based organizations, including Google and Microsoft, are increasingly central to modern AI research, and the PageRank of academic institutions are on the decline. Chinese research institutions at today's forefront of AI research are notably absent from Figure 6-4a because their rise in prominence is recent in the 65 year time span of our analysis. However, the increasing prominence of Chinese research institutions, as well as other non-U.S. based institutions, is apparent when focus on recent years (see section E.8).

While academia has remained the largest source of AI papers throughout the entire period of analysis, the increased presence of industry can be seen from the authorship of AI papers over time (see Fig. 6-4b). Out of the 10% of AI papers with the most citations after 10 years, the relative number of papers with industry-only authorship is on the decline. Meanwhile, collaborations between academia and industry are becoming more abundant.

How are other fields of study responding to the increased presence of industry in AI research? As an example, references from Engineering showed preference for AI papers with industry-only authorship until the late 1980's, which is contrary to the aggregate trend (see Fig. 6-4c, and see section E.4 for similar plots for all academic fields). Similar to reference strength, temporal changes in a field's preference for AI papers with industry authorship (i.e. at least one author has an industry affiliation) may result from the abundance of industry-based AI paper production over time. Therefore, we examine each field's Industry Preference Score, which is given for field A by

$$IPS_{year}(A) = \frac{\left(\text{ref. share from } A \text{ to industry AI papers}\right)}{\left(\text{industry share of AI papers from 1950 to } year\right)}. \tag{6.4}$$

Here, an AI paper has industry authorship if at least one co-author has an affiliation with an industry-based institution. Fields with $IPS(A) > 1$ exhibit stronger preference for industry AI papers than would be expected under random referencing behavior towards AI papers. Academic fields that may be interested in the application of AI technology, such as Materials Science, Engineering, Chemistry, and Physics, tend to have greater preference for industry AI papers. However, many of the social sciences and fields that study social and societal dynamics, such as Sociology, Economics, Philosophy, and Political Science, tend to have lower preference for industry AI papers.

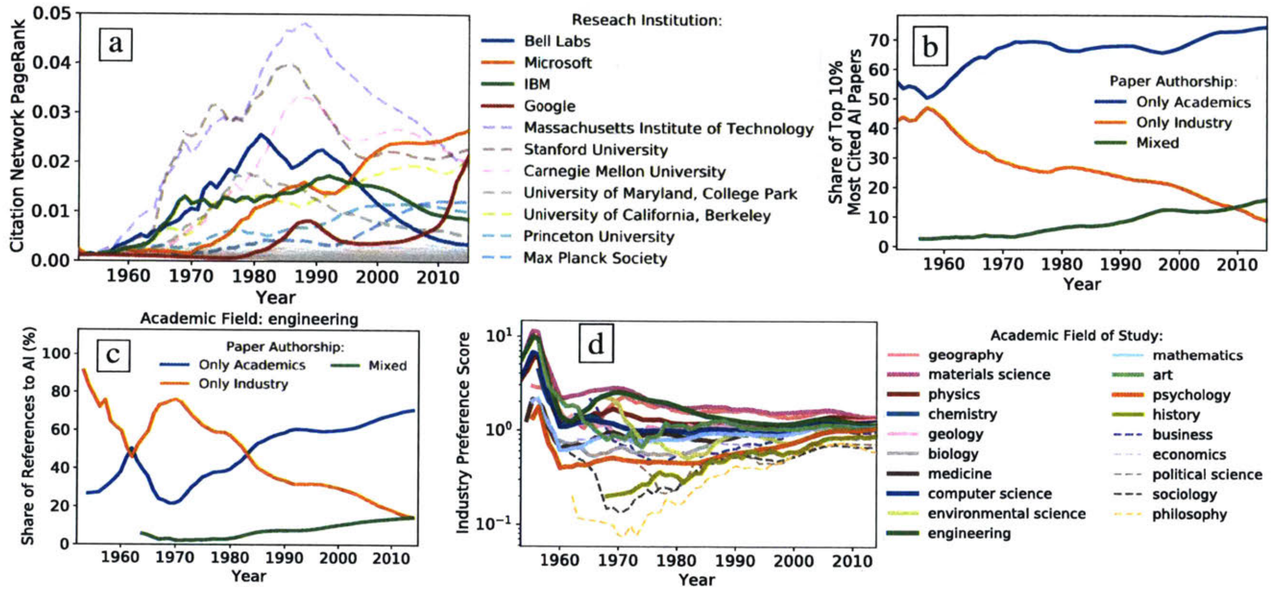


Figure 6-4: Industry is increasingly central to artificial intelligence research, but industry-authored AI papers are referenced less often by other academic fields. **(a)** The PageRank of each research institution using the number of references from AI papers published by each institution to papers published by each other research institution. The lines of notable research institutions are colored for visualization. Dashed lines indicate academic institutions while solid lines indicate industry. **(b)** The share of the top 10% most cited AI papers published in each year with academic-only, industry-only, and mixed authorship. **(c)** Similar to (b), we examine the referencing behavior of Engineering towards AI papers according to the authorship of the AI papers. Analogous plots for each other academic field are provided in section E.4. **(d)** Generalizing on (C), the Industry Preference Score (IPS) calculated from each academic field's referencing behavior towards AI papers (see eq. (6.4)). Solid (dashed) lines indicate fields that reference AI papers with industry-only authorship more (less) than would be expected according to random referencing behavior. In all plots, lines are smoothed using a 5-year moving average.

6.6 Discussion

Humanity's longstanding quest [148] for artificial intelligence (AI) is rapidly advancing in areas like vision, speech, and pattern recognition. However, as we deploy AI systems, their complete impact includes their social, ethical, and societal implications in addition to capabilities and productivity gains. Understanding these implications requires an on-going dialogue between the researchers who develop new AI technology and the researchers who study social and societal dynamics. Therefore, it is concerning to find a gap between AI research and the research conducted in other fields (see Fig. 6-2).

AI paper production has increased quickly and steadily throughout the last half century (see Fig. 6-1) which suggests that the remarkable and seemingly-sudden progress in AI is rooted in decades of research. Although AI research found as much early inspiration in Psychology as Computer Science and Mathematics, it has since transitioned towards computational research. Conversely, several other academic fields are dedicating relatively more references to AI research. For example, Engineering and Mathematics research cite AI papers with increasing relative abundance throughout the period of analysis—making more frequent references to AI papers than would be expected under random referencing behavior (see Fig. 6-2c&d). However, the decreasing reference strength towards AI papers that we observe on aggregate suggests that most researchers are unable to keep up with the explosion of AI paper production (see Fig. 6-2d). These findings may help explain why recent AI technologies have only recently revealed important (and largely unintentional) social consequences, such as racial bias in facial recognition software [63, 64], the ethical dilemmas arisen from autonomous vehicles [52], and income inequality in the age of AI [8,95,98]. If current trends persist, then it may become increasingly difficult for researchers in any academic fields to keep track of cutting-edge AI technology.

The bibliometric gap between AI and other sciences grew with the advent of AI-specific conferences and the increased prominence of industry within AI research. In general, Computer Science conferences can bolster the importance of publications [99] and enable major players to disproportionately influence the entire area of research [172]. Although Computer Science is becoming more diverse on whole [172], the scientific impact of AI research institutions is becoming less diverse (see Fig. 6-3a). In particular, Microsoft and Google have taken away the central role from universities according to citation PageRank (see Fig. 6-4a), perhaps, through preferential referencing of publications within AI (see Fig. 6-3b).

This transition towards industry is challenging for studying the social and societal dynamics of AI technologies. Social science research is less likely to reference AI publications with authors who have industry-based affiliations. Combined with AI's decreasing reference strength towards social sciences, these observations suggest this gap between research areas will continue to grow. The fields that study social bias, ethical concerns, and regulatory challenges may be ignorant of new AI technology—especially when deployed in industry. While our interpretation of these results is somewhat speculative, we believe our observations may highlight an important dynamic within the AI research community that merits further investigation.

The gap between social science and AI research means that researchers and policy makers may be ignorant of the social, ethical, and societal implications of new AI systems. While this gap is concerning from a regulatory viewpoint, it also represents an opportunity for researchers. The academic fields that typically inform policy makers on social issues have the opportunity to fill this gap. While our study is a step towards this goal, further work may explicitly quantify the social and societal benefits and consequences of today's AI technology as well is identifying the mechanisms that limit communication between research domains.

Appendix A

Towards Understanding the Impact of AI on Labor

A.1 Identifying Characteristic Skills

The U.S. Department of Labor regularly surveys workers about the importance of each of 161 workplace skills, knowledge, and abilities (hereafter "skills") to the completion of their daily work. This results in the O*NET database, which details the relationships between each of the 672 occupations in the Standard Occupational Classification (SOC) system and their underlying skill requirements. The O*NET database is updated annually, which allows for temporal snapshots of labor requirements.

Here, we use $onet(j,s) \in [0,1]$ to denote the raw importance of each skill $s \in S$ to a job $j \in J$ according to 2015 O*NET data. $onet(j,s) = 1$ indicates that skill s is essential to workers of job j, while $onet(j,s) = 0$ indicates that the skill is irrelevant to workers of the job. These individual skills represent key features that uniquely identify occupations, and so, we seek an unsupervised data-driven methodology that maximizes our ability to distinguish occupations. To this end, raw skill importance may not control for the dampening effect of ubiquitous skills, such as *Identifying Objects* and *Communicating with Supervisors and Peers*. We control for ubiquitous skills be normalizing raw skill importance across all occupations according to their revealed comparative advantage [111, 115, 117] (also known as location quotient [158, 194, 195]). Specifically, we calculate

$$rca(j,s) = \frac{onet(j,s)/\sum_{s' \in S} onet(j,s')}{\sum_{j' \in J} onet(j',s)/\sum_{j' \in J, s' \in S} onet(j',s')}, \tag{A.1}$$

for each job j and skill s. We interpret this quantity as the ratio between the relative importance of s to j (see the numerator) and the relative importance of that skill across all skills and all occupations (see the denominator). Finally, we say that skill s is *characteristic* of job j if $rca(j,s) > 1$.

A.2 Tracking Programming Language Popularity

Computer science is a rapidly changing and growing field with constant creation of new programming languages and libraries. Therefore, it is difficult to arrive at a complete taxonomy of all existing programming languages at any given time. Here, we restrict our analysis to a finite set of well-known programming languages: Java, Javascript, Scala, Python, PHP, C, C#, Objective-C, R, Ruby, Swift, Visual Basic, Perl, Iua, Delphi, Haskell, Rust, TypeScript, Kotlin, Go, Matlab, and VBA.

We turn to search engine activity through Google.com to assess the monthly popularity of each programming language as function of relevant searches for tutorials. Specifically, we collect data from Google Trends for searches of the form "<programming language> tutorial" (e.g. "Java tutorial"). This allows us to calculate the relative search volume associated with each programming language in each month from 2004 to the beginning of 2018. This methodology represents a conservative estimate of the true search volume associated with each programming language, but controls for other potential issues (e.g. does searching for "Python" refer to the animal or the programming language?).

Appendix B

Small cities face greater impact from automation

B.1 Firm Size Increases with City Size

The U.S. Bureau of Labor Statistics (BLS) uses the annual tax filings of companies to produce a yearly census of those companies. Unfortunately, the data available to the public doesn't include the specific distribution of BLS jobs comprising each firm. Previous work has shown that firm sizes nation wide follow a Zipf distribution [27] indicating that a majority of firms are small, but surprisingly large ones also exist infrequently. Figure B-1 shows that the average number of workers per firm increases logarithmically with city size. Larger firms have more capital with which to hire specialized workers along with organization/managerial staff to coordinate those workers. According to our theory, there exists a positive feedback loop where large firms provide demand for specialized workers and cities provide a richer market of skilled workers to meet that demand.

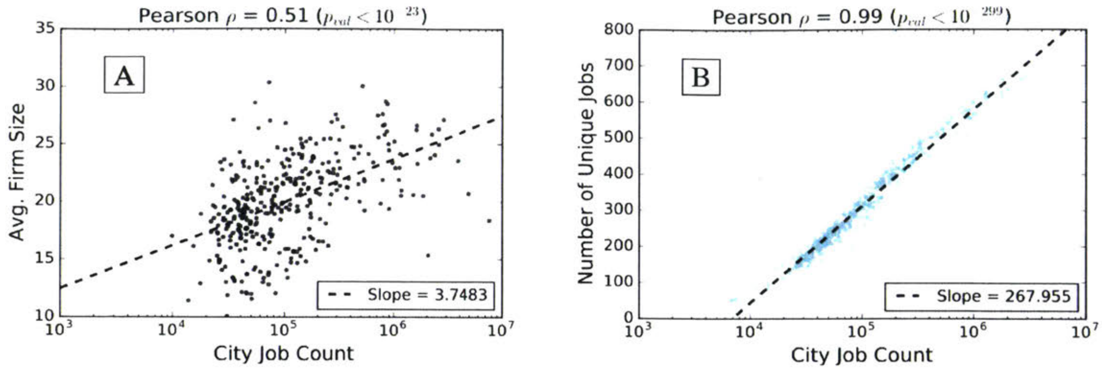


Figure B-1: **(A)** The average number of workers per firm grows logarithmically with city size. **(B)** Consistent with [218], we find the number of unique jobs grows logarithmically with city size.

B.2 Measuring Labor Specialization

B.2.1 Normalized Shannon Entropy

We employ normalized Shannon entropy, as opposed to the standard Shannon entropy definition, to control for size effects on the distributions in cities. For example, it has been shown that the number of different occupations grows with city size (see [218], and SM Fig. B-1B), but this result may be due to randomness as more people (e.g. sampled from a long-tailed distribution) are added to a city, and does not account for how workers are distributed amongst these occupations. Large cities may have only a few workers of otherwise absent occupations in small cities, but, perhaps, this distinction does not mean much qualitatively. This motivates us to consider the distribution of workers amongst different occupations, rather than only considering the number of occupations, and to apply relevant information theoretic methods to measure the diversity/specialization of these distributions. We normalize the standard Shannon entropy calculation to control for the number of different occupations in a city, or, equivalently, we normalize Shannon entropy by the maximum possible Shannon entropy given the number of different occupations in the city (i.e. given a number of occupations, Shannon entropy is maximized for the uniform distribution).

This normalization is a standard practice for comparing the diversity or information of systems of different sizes. For a summary of normalized entropy, see [138]. In particular, normalized Shannon entropy has been used in a variety of fields, including virology [66], climatology [215], and city science [82]. To understand this normalization, consider that a sufficient number of roles of a fair 6-sided dice and, separately, of a fair 20-sided dice should each produce uniform distributions with maximized Shannon entropy. However, the Shannon entropy of the distribution for the 6-sided dice is $-\sum \frac{1}{6} \cdot \log(\frac{1}{6}) = 1.79$ and the entropy of the distribution for the 20-sided dice is $-\sum \frac{1}{20} \cdot \log(\frac{1}{20}) = 3.00$; specifically, they are not equivalent despite both being discrete uniform distributions because the distributions have a different number of bins. This is analogous to cities having a different number of unique occupations due, potentially, to randomness that occurs with increased city size. We control for this effect by normalizing Shannon entropy according to the maximum possible Shannon entropy given the number of bins in the discrete distribution. Specifically, given a discrete system with N bins (i.e. N-sided dice, or a city with N unique occupations), Shannon entropy is maximized when the distribution is uniform, and the maximum value is given by

$$-\sum_{i=1}^{N} \frac{1}{N} \cdot \log(\frac{1}{N}) = -N \cdot \frac{1}{N} \cdot \log(1/N) = -\log(1/N) = \log(N). \tag{B.1}$$

Therefore, to normalize Shannon entropy according to the maximum possible Shannon entropy, we divide the standard Shannon entropy calculation by $\log(N)$ to obtain

$$-\sum_{i=1}^{N} p_i \cdot \frac{\log(p_i)}{\log(N)},$$

where p_i is the probability of bin i. This normalized Shannon entropy produces a value of 1 for discrete uniform distributions regardless of the number of bins (i.e. regardless of N). In particular, this normalization allows us to control for the number of unique occupations across cities of different sizes to determine the uniformity of job and skill distributions in cities.

B.2.2 The Labor Specialization of Individual Jobs

We present BLS jobs ordered by decreasing skill specialization in Table B.2.2. We also provide the scaling exponent of each BLS job, along with the Pearson correlation of the relative abundance of each job to the expected job impact from automation (discussed below) across cities. Figure B-2A shows that specialized jobs tend to have larger scaling exponents. Figure B-2B shows the

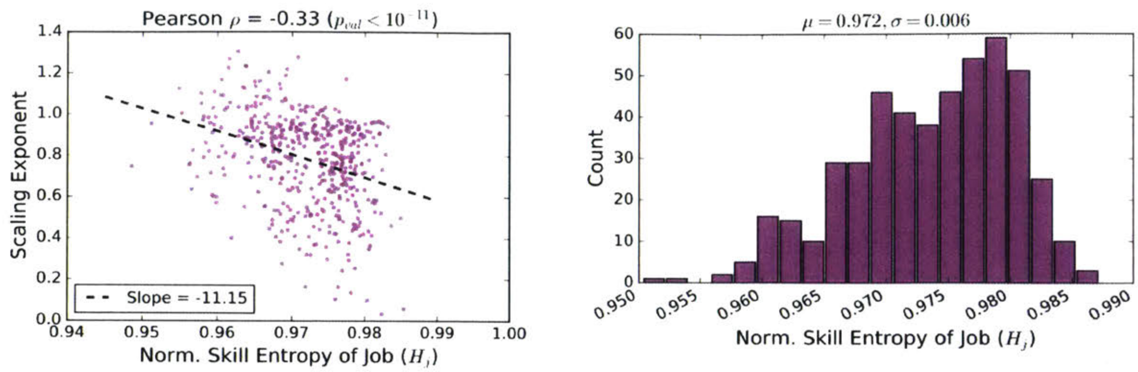


Figure B-2: Characterizing the skill specialization of individual jobs. **(A)** Skill specialization indicates larger scaling exponents with city size for individual jobs. **(B)** The distribution of skill specialization across BLS jobs.

distribution of job specialization.

B.2.3 Characterizing Specialization through O*NET Skills

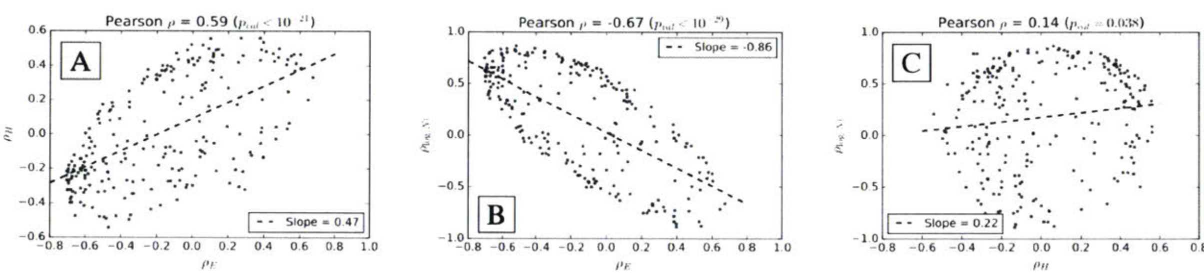


Figure B-3: Comparing the relationships of O*NET skills to city size, expected job impact, and labor specialization. **(A)** We plot Pearson correlation of raw skill importance to expected job impact (ρ_E) on the x-axis versus the Pearson correlation of raw skill importance to city skill entropy (ρ_H) on the y-axis. We see that which indicate job impact from automation also indicate decreased specialization in cities. **(B)** We plot Pearson correlation of raw skill importance to expected job impact (ρ_E) on the x-axis versus the Pearson correlation of raw skill importance to city size ($\rho_{\log(N)}$) on the y-axis. We see that which indicate job impact from automation also indicate smaller city sizes. **(C)** We plot Pearson correlation of raw skill importance to city skill entropy (ρ_H) on the x-axis versus the Pearson correlation of raw skill importance to city size ($\rho_{\log(N)}$) on the y-axis. The correlation between these two variables is not significant

We want to understand how each O*NET skill contributes to the relationships we observe. We present our findings in Table B.6.4. First, we compare the raw importance of a skill in each city by summing the raw importance of the skill across each job. We then measure the Pearson correlation of the sum of a given skill compared to the expected job impact of each city (denoted ρ_E, second column of table), the skill entropy each city (denoted ρ_H, third column of table), and the size of each city (denoted $\rho_{\log(N)}$, right-most column of table). The skills in the Table B.6.4 are ordered according to their correlation with expected job impact in cities. For each column, the p-value for the correlation is presented in parentheses.

Figure B-3 allows us to understand how related each correlation is by taking the Pearson correlation of each ρ we described above. Figure B-3A demonstrates that skills which indicate lower expected job impact in cities also indicate greater skills specializations in cities. Figure B-3B demonstrates that skills which indicate lower expected job impact in cities also indicate larger cities. Interestingly, Figure B-3C demonstrates that skills which indicate skill specialization in cities are not significantly related to the skills which indicate city size. This finding is surprising given the other panels of the figure, and motivates us to consider the relationship between occupational specialization and city size through the jobs in each city (see main analysis).

B.3 Estimating the Affects of Automation

Automation and its impact on labor are increasingly important topics to researchers [57, 67, 167]. Examples throughout history, such as the industrial revolution and the advent of computers, demonstrate how technological advancement can lead to both job loss and job creation [15, 147]. However, it is extremely difficult to predict how quickly a seemingly imminent technology will reach maturity and what the impact of that technology will be. For example, it's currently topical to discuss self-driving cars, but, while autonomous-capable cars are available for purchase, no self-driving cars are currently operated on the mass market. On the other hand, early leaders in computer hardware famously offered pessimistic predictions on the impact of computing:

- *"There is no reason anyone would want a computer in their home"* - Ken Olsen, founder of Digital Equipment Corporation (1977)
- *"I think there is a world market for maybe five computers"* - Thomas Watson, former president of IBM (1943)

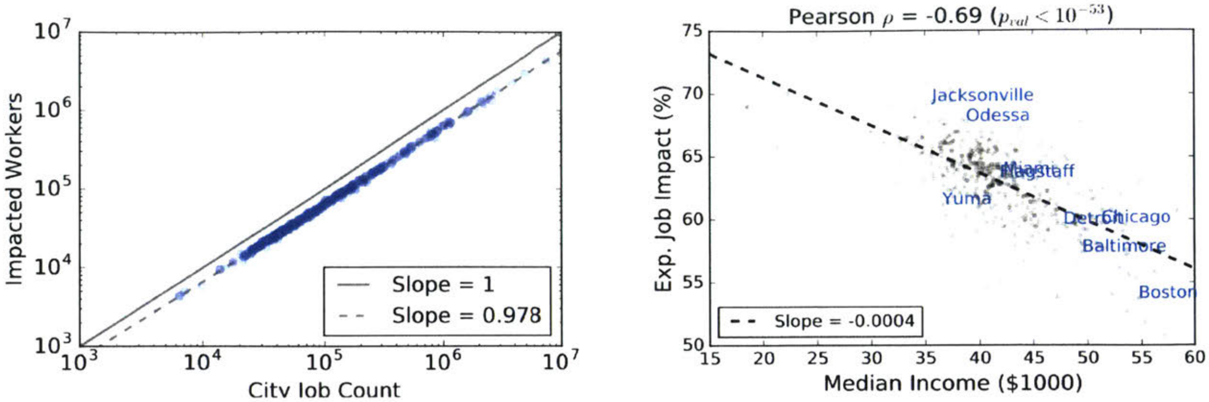


Figure B-4: **(A)** The expected number of displaced workers grows slightly sublinearly ($\beta = 0.978$) with city size. **(B)** Expected job impact is anti-correlated with median income of cities according to U.S. Census.

B.3.1 Estimating Automation Impact using Frey/Osborne Data

Frey and Osborne [98] produced probabilities of computerization for each BLS job. They convened a workshop of leaders in automation to identify which of the BLS jobs were certainly automatable and certainly not-automatable. They used the O*NET skills dataset to identify the raw importance of nine workplace skills to each BLS job: Manual Dexterity, Finger Dexterity, Cramped Workspace/Awkward Positions, Originality, Fine Arts, Social Perceptiveness, Negotiation, Persuasion, and Assisting & Caring for Others. These O*NET skills represent “known bottlenecks to computerization.” Using the importance of these skills to the jobs whose automatability was clear, they used a Gaussian process classifier to produce a probability of computerization for each BLS job.

Frey and Osborne used these probabilities to conclude that 47% of the current U.S. jobs are at “high risk" of computerization. n Several studies [19, 38, 170] utilize these same probabilities to investigate the impacts of automation, which highlights the utility of the probabilities despite the difficulty of the prediction undertaken in [98]. We use the same probabilities in combination with the distribution of BLS jobs across U.S. cities to add spatial resolution to their findings. For a city,

m, the expected job impact from automation is calculated according to

$$E_m = \sum_{j \in Jobs_m} p_{auto}(j) \cdot p_m(j), \tag{B.2}$$

where $p_{auto}(j)$ is the probability of computerization according to [98] and $p_m(j)$ is the proportion of workers in city m with job j. Table B.6.1 demonstrates the ordered list of cities according to expected job impact.

As mentioned above, it's difficult to validate automation predictions. Nonetheless, our calculations for expected job impact represent an aggregate signal for the types of jobs in a city in relation to imminent automation technology. In the Table B.6.1, we present the U.S. cities ordered by their expected job impact from automation. The list produces an ordering that appears to make sense; cities with technology companies and research institutes, such as Boston, M.A., and Boulder, C.O., have the lowest expected job impact, while cities relying on the tourist industry and agriculture, such as Myrtle Beach, S.C., and Napa, C.A., have the highest expected job impact. While the absolute proportions can only be validated with time, we believe the overall trend embodied in expected job impact in cities represents an underlying true signal.

To demonstrate the robustness of our results further, we perform two robustness checks to verify the negative trend between city size and expected job impact from automation (see Fig. 3-1B from the main analysis). The probability of computerization (i.e. $p_{auto}(j)$) from [98] are produced through a machine learning process applied to predictions of the automatability of jobs from experts. Therefore, we expect some errors in the predictions of these experts, and our task is to demonstrate that the error in the resulting $p_{auto}(j)$ would need to be substantial to invalidate our finding. We perform this analysis by artificially adding random noise to each $p_{auto}(j)$ according to

$$p^*_{auto}(j) = p_{auto}(j) + e_j, \tag{B.3}$$

where e_j is chosen uniformly at random from the interval $[-error, +error]$ for each occupation. For each choice of $error$, we perform 500 trials calculating a new $p^*_{auto}(j)$ for each occupation and recalculating the expected job impact from automation in each city according to

$$E_m^* = \sum_{j \in Jobs_m} p_{auto}^*(j) \cdot p_m(j) \tag{B.4}$$

similar to equation B.2. Finally, we measure the Pearson correlation between $\log_{10}$ the total employment in each city and E_m^* so that we can compare to the empirical relationship we observe in Figure 3-1B (Pearson $\rho = -0.53$, $p_{val} < 10^{-28}$). Figure B-5A demonstrates the results of this exercise. We find that substantial error ($error \approx 0.15$) needs to be added to the empirical probabilities of computerization for each occupation before our result from the main analysis no longer represents the observed trend. Even if we make the extremely strong assumption of $error = .5$, we would still observe a strong negative trend, and we would still conclude that small cities face greater impact from automation.

In the second robustness check, we test the robustness of our observed relationship between city size and expected job impact if a randomly selected subset of occupations are removed from the analysis. For each proportion of occupations to be removed, we perform 500 trials of randomly selecting occupations to be ignored and recalculate E_m using the $p_{auto}(j)$ presenting in [98]. We then measure the resulting Pearson correlation between these new E_m and $\log_{10}$ the total employment in each city. Figure B-5B demonstrates that our empirical observation from Figure 3-1B in the main analysis holds even if very large proportions of occupations are ignored. In fact, only when we ignored half of all occupations did we observe any trials demonstrating a trend contrary to the one presented in the main analysis. Therefore, we conclude that small cities face greater impact from automation.

B.3.2 Estimating Automation Impact using OECD Data

The Organization for Economic Co-operation Development (OECD) released alternative estimates for the probability of job automation with a focus on job categories used by OECD countries [16]. Rather than the job-based approach used in [98], assessments on the automatability of workplace skills were derived. These skill assessments can be used in combination with government data

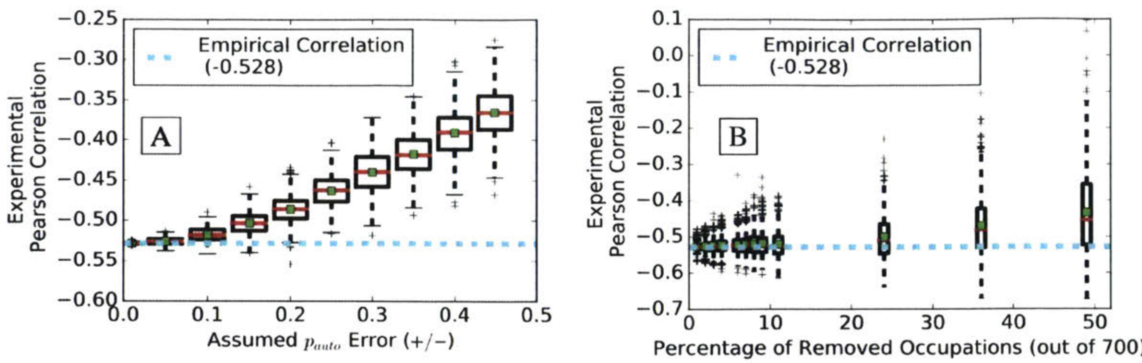


Figure B-5: The relationship between city size and expected job impact from automation is robust. **(A)** For choice of assumed error in the predictions from [98], we perform 500 trials measuring the resulting Pearson correlation between $\log_{10}$ city size and expected job impact from automation after the error has been added to each occupation's probability of computerization (y-axis). **(B)** After selecting a proportion of occupations (x-axis), we perform 500 trials of randomly selecting that many occupations to remove while measuring the resulting Pearson correlation between $\log_{10}$ city size and the expected job impact from automation in cities (y-axis).

relating the importance of skills to jobs to assess the likelihood of computerization for jobs. Contrary to the alarming 47% of jobs at "high risk of computerization" found by Frey and Osborne, these new probabilities produce a more mild conclusion of only 9%. These job probabilities were derived with OECD job definitions in mind, but collaborations between OECD and U.S. BLS have lead to an official mapping between the two job definitions. We utilize this mapping to assess the resilience of labor markets in cities as a function of city size in Figure B-6. Despite the more conservative estimates in [16], our results remain; we again observe significantly decreased expected job impact in large cities (Fig. B-6A).

B.3.3 Expected Job Impact & Labor Specialization in Cities

In Figure B-7, we further characterize the relationship between a city's resilience to job impact from automation and labor specialization. We provide additional figures in the main analysis detailing how workplace skills explain the positive correlation we observe between labor specialization and resilience to job impact in cities. Here, we demonstrate that resilience to job impact is significantly correlated to the number of unique jobs in a city, and more weakly correlated to the Shannon entropy of job distributions. This weaker correlation motivates our investigation into

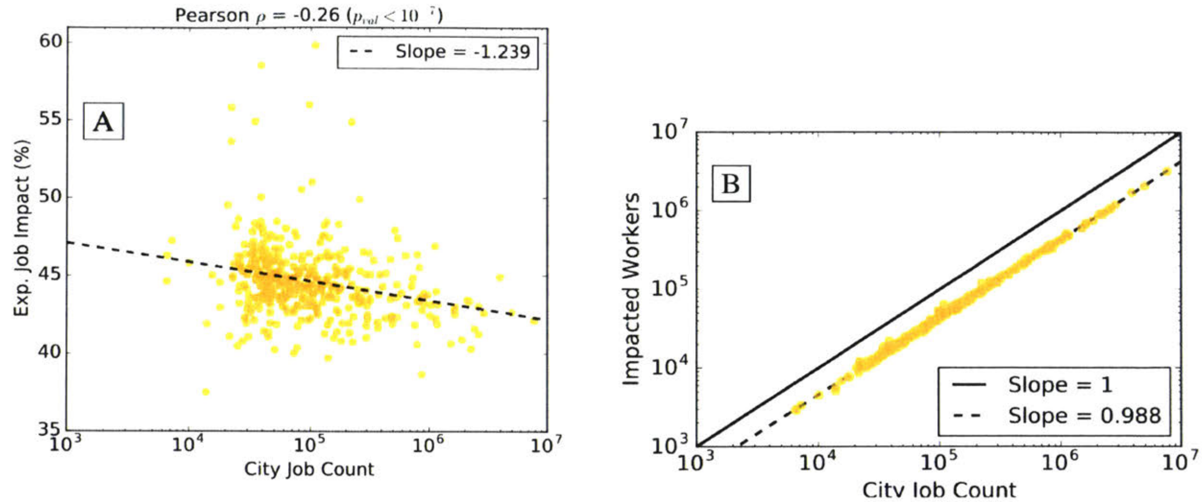


Figure B-6: Expected job impact from automation decreases with city size using conservative estimates of job loss. **(A)** The expected job impact of cities decreases with city size. **(B)** The number of displaced workers per city grows slightly sublinearly with city size ($\beta = 0.988$).

workplace skills, in addition to the distribution of jobs, presented in the main analysis.

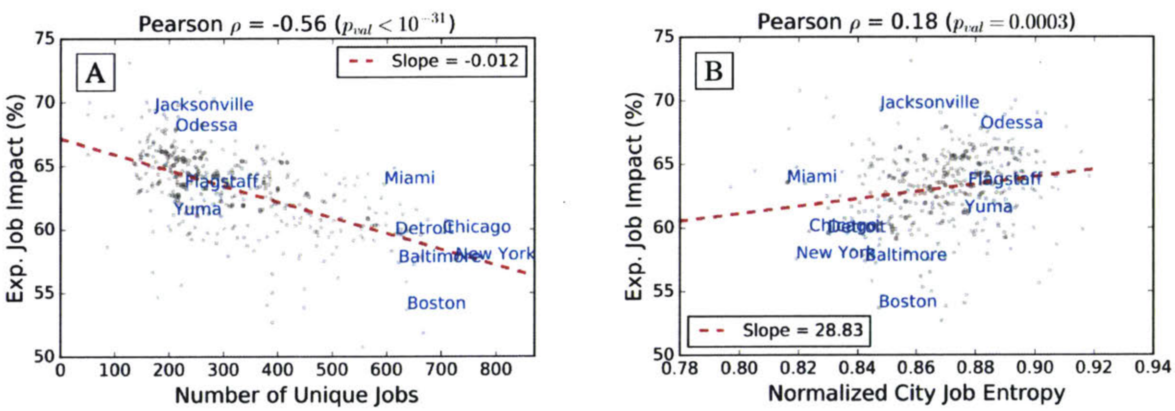


Figure B-7: Characterizing the relationship between labor specialization and job impact from automation in cities. **(A)** Resilience to job impact is correlated with the number of unique jobs in cities. **(B)** Increased labor specialization according to job distributions in cities indicates increased resilience to job impact from automation.

B.3.4 Explaining Differences in Expected Job Impact

From equation B.2, we may observe that both the automatability and the employment share of individual occupations contribute to the expected job impact from automation in a city. Correspondingly, we measure the difference in expected job impact for cities m and n according to

$$\begin{aligned} E_m - E_n &= \sum_{j \in Jobs} p_{auto}(j) \cdot \big(share_m(j) - share_n(j)\big) \\ &= \sum_{j \in Jobs} p_{auto}(j) \cdot \big(share_m(j) - share_n(j)\big) - \sum_{j \in Jobs} E_n \cdot \big(share_m(j) - share_n(j)\big) \\ &= \sum_{j \in Jobs} \big(p_{auto}(j) - E_n\big) \cdot \big(share_m(j) - share_n(j)\big), \end{aligned} \tag{B.5}$$

where we have utilized $\sum E_n \cdot (share_m(j) - share_n(j)) = 0$. Here, we let $Jobs$ denote the set of all occupation types across all cities, $p_{auto}(j)$ denotes the probability of computerization of occupation j according to [98], and $share_m(j)$ denotes the employment share of occupation j in city m. Equation B.5 highlights that occupation j's influence on the difference in expected job impact in cities m and n falls into one of four categories:

1. occupation j is relatively resilient to automation (i.e. $(p_{auto}(j) - E_n) > 0$) and relatively more abundant in city m (i.e. $(share_m(j) - share_n(j)) > 0$),

2. occupation j is relatively susceptible to automation (i.e. $(p_{auto}(j) - E_n) < 0$) and relatively less abundant in city m (i.e. $(share_m(j) - share_n(j)) < 0$),

3. occupation j is relatively resilient to automation (i.e. $(p_{auto}(j) - E_n) > 0$) and relatively less abundant in city m (i.e. $(share_m(j) - share_n(j)) < 0$), or

4. occupation j is relatively susceptible to automation (i.e. $(p_{auto}(j) - E_n) < 0$) and relatively more abundant in city m (i.e. $(share_m(j) - share_n(j)) > 0$).

Occupations in categories 1 and 2 effectively increase $E_m - E_n$, while occupations in categories 3 and 4 effectively decrease the difference.

Let

$$\delta_{(m,n)}(j) = 100 \cdot \frac{\big(p_{auto}(j) - E_n\big) \cdot \big(share_m(j) - share_n(j)\big)}{E_m - E_n} \tag{B.6}$$

denote the percent influence of occupation j on the difference in expected job impact for cities m and n. Figure B-8 demonstrates a visualization of equation B.5 that we call an "occupation shift." Correspondingly, if we add the employment distributions in the 50 largest cities and 50 smallest cities together (respectively), then we can quantify how each occupation contributes to the differential impact of automation on employment in large and small cities. We present this occupation shift in Figure B-9 (also Figure 3-5).

Referring to the job clusters from Figure 3-5, we see that purple occupations and blue occupations contribute the most to the difference in expected job impact, while green and yellow occupation types effectively diminish the difference in both occupation shifts. However, certain occupations, such occupations of the green job cluster, can both increase and decrease the difference between resilient and susceptible cities. The occupation shift allows us to understand which occupations explain the overall trend and which occupations go against the overall trend. If we had only considered occupations that add to the difference (i.e. occupations corresponding to dark colored bars on the right side of the plot), then we may have incorrectly concluded that the differences in relatively susceptible occupations explain the difference we observe in these two examples. This transparency can help urban policy makers determine how labor shifts in different industries may effect their preparedness for the impact of new technology.

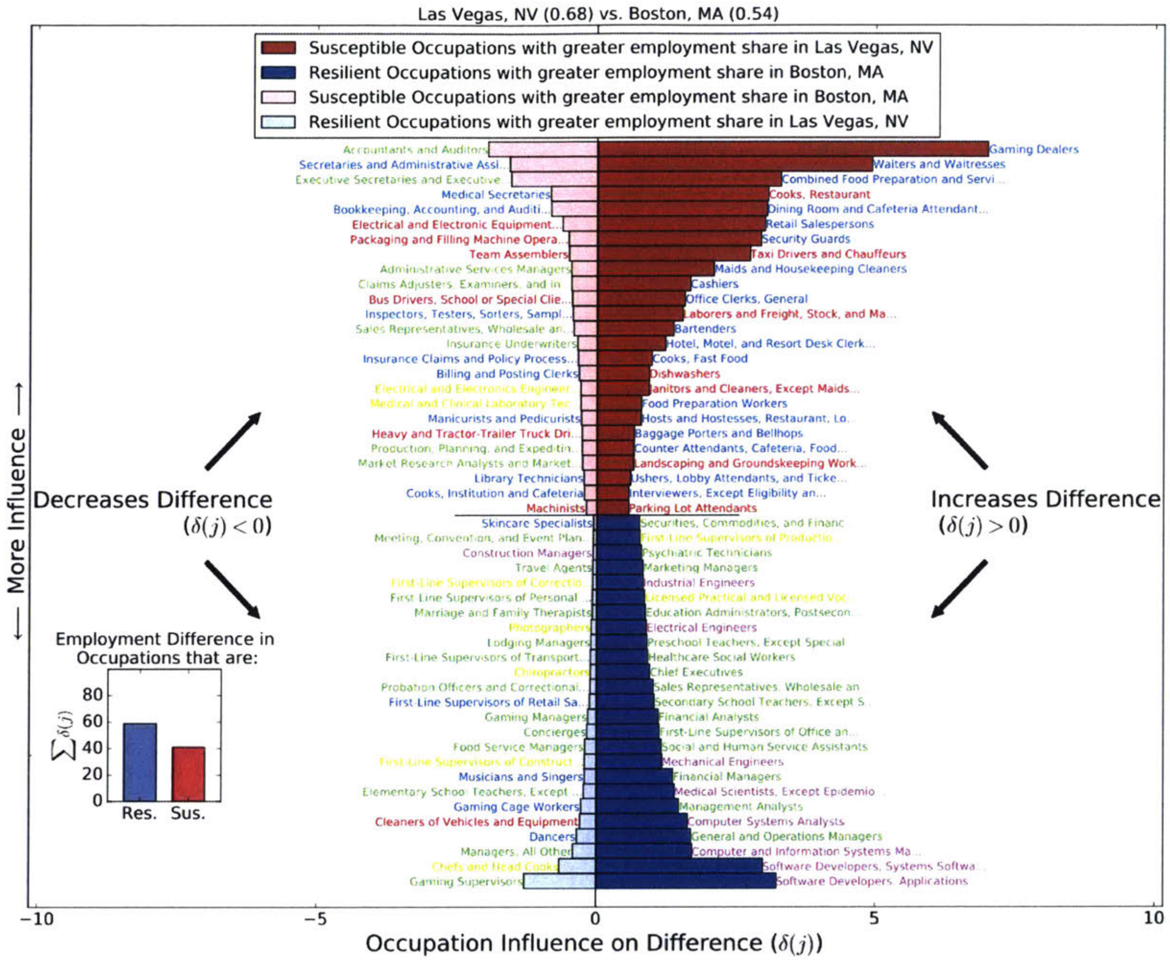


Figure B-8: An occupation shift explaining the difference in expected job impact for Boston, MA ($E_m = 0.54$) compared to Las Vegas, NV ($E_m = 0.68$) using equation B.5. Each horizontal bar represents $\delta_{(\text{Las Vegas,Boston})}(j)$ of occupation j. The occupation title is provided next to the corresponding bar and colored according to its job cluster as identified in Figure 3-4. Red bars represent occupations with higher risk of computerization compared to Boston's expected job impact. Blue bars represent occupations with lower risk of computerization compared to Boston's expected job impact. Dark colors represent occupations that effectively increase the difference, while pale colors represent occupations that effectively decrease the difference in expect job impact. Bars in each of the quadrants are vertically ordered according to $|\delta_{(\text{Las Vegas,Boston})}(j)|$. The inset in the bottom left of the plot summarizes the overall influence of resilient occupations compared to occupations that are at risk of computerization.

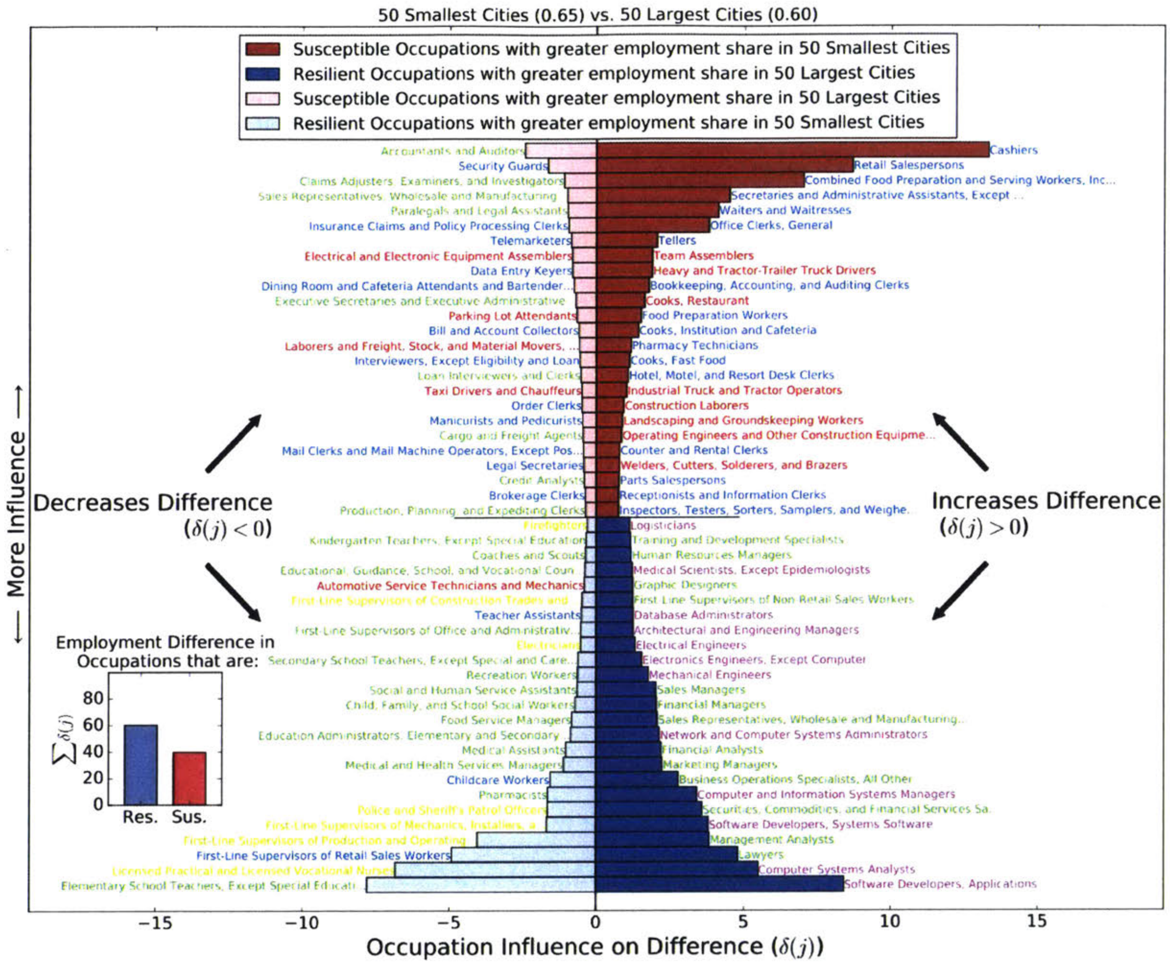


Figure B-9: An occupation shift explaining the difference in expected job impact for the 50 largest cities ($E_m = 0.60$) compared to the 50 smallest cities ($E_m = 0.65$) using equation B.5. Each horizontal bar represents $\delta_{(\text{Small Cities},\text{Large Cities})}(j)$ of occupation j. The occupation title is provided next to the corresponding bar and colored according to its job cluster as identified in Figure 3-5. Red bars represent occupations with higher risk of computerization compared to the expected job impact in large cities. Blue bars represent occupations with lower risk of computerization compared to the expected job impact in large cities. Dark colors represent occupations that effectively increase the difference, while pale colors represent occupations that effectively decrease the difference in expect job impact. Bars in each of the quadrants are vertically ordered according to $|\delta_{(\text{Small Cities},\text{Large Cities})}(j)|$. The inset in the bottom left of the plot summarizes the overall influence of resilient occupations compared to occupations that are at risk of computerization.

B.3.5 Labor Specialization as a Mediator for City Size and Automation Impact

Our main analysis predominantly relies on linear regression to explore the relationship between city size, labor specialization, and the expected impact of automation in cities (see Figures 3-1, 3-2, & 3-3). In particular, we find evidence that the relationship between city size and expected impact may be mediated by the labor specialization in cities. This conceptualization leads us to perform a formal mediation analysis [122] with city size as a treatment variable (i.e. $\log_{10}$ total employment in cities, denoted $size_m$), our various measures for labor specialization (i.e. $H_{job}(m)$, $H_{skill}(m)$, and $1 - T_m$) as independent mediators, and the expected impact from automation (i.e. E_m) as the outcome variable.

We take the generic urban variables from Figure 3-3 as additional control variables; these variables include the median household income ($income_m$), the per capita GDP (GDP_m), the percent of population with a bachelor's degree ($bachelor_m$), and the number of unique occupations in each city ($jobs_m$). The purpose of these control variables are to mitigate omitted variable bias, but the U.S. labor system is a sufficiently complicated system that omitted variable bias can never fully be controlled for. In our opinion, this observation limits the strength of any conclusion about causality [104].

This analysis is subject to further assumptions as well. Firstly, we are assuming that the effect of city size on job impact is constant across cities. Again, this assumption is extremely difficult to prove given the complexity of the U.S. labor system, regional geographies, regional politics, and economic trade. Secondly, the effects of unobserved causes for the mediator and outcome variables (denoted e_1 and e_2, respectively) are uncorrelated. In the analysis below, we measure the Pearson correlation between e_1 and e_2 for each choice of labor specialization measure.

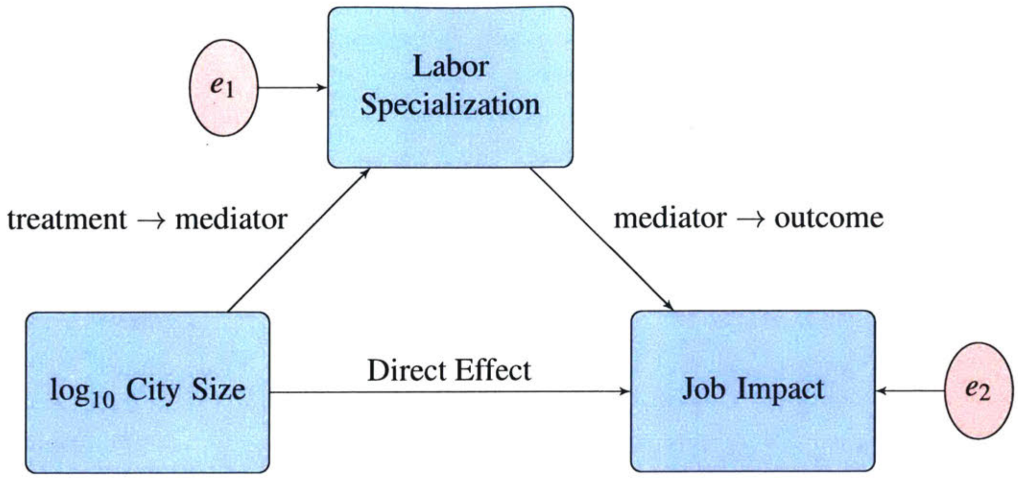


Figure B-10: Schematic for mediation analysis. We consider the city size (i.e. $size_m$) as the treatment variables, as measure for labor specialization (i.e. $H_{job}(m)$, $H_{skill}(m)$, or $1 - T_m$) as the mediator variable, and the expected impact of automation on cities (E_m) as the outcome variable. e_1 represents the unobserved causes of labor specialization, while e_2 represents the unobserved causes of the impact of automation in cities.

Labor Specialization Variable	Avg. Causal Mediated Effect (ACME)	Avg. Direct Effect (ADE)	Total Effect	Pearson(e_1,e_2)
Job Specialization ($H_{job}(m)$	0.793***	−0.516*	0.277	7.38×10^{-16}
Skill Specialization ($H_{skill}(m)$)	−0.078	0.357	0.278	3.85×10^{-16}
Theil Entropy ($1 - T_m$)	−0.107*	0.389	0.282	6.01×10^{-17}
* p-value < .1, ** p-value < .01, *** p-value < .001				

Table B.1: The results of mediation analysis. All variables were standardized prior to analysis. Ignoring whether the necessary assumptions for mediation analysis are met, we find some evidence that job specialization ($H_{job}(m)$) may act as a mediator for the effect of city size on the expected impact of automation.

B.4 Robustness Check of the Linear Regression Model for E_m

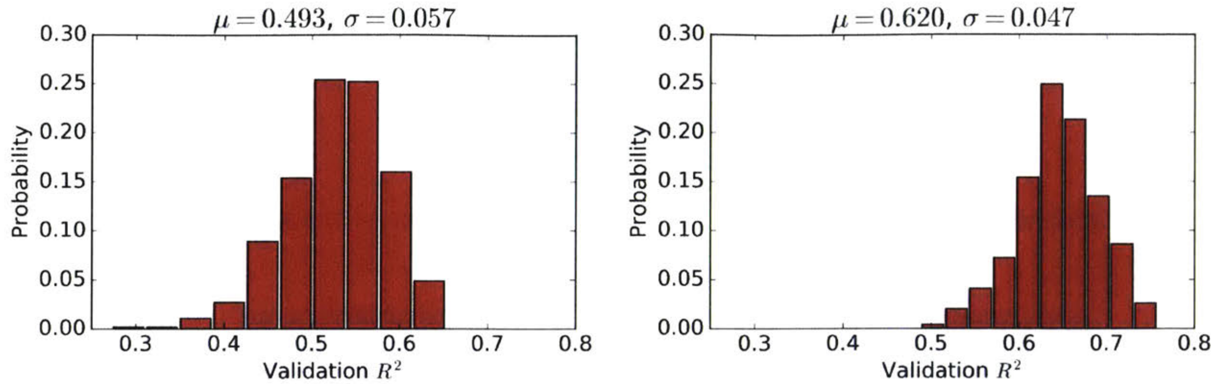


Figure B-11: To confirm the validity of the regression models, we perform 1,000 trials where half of the cities are randomly selected without replacement as training data and the remaining cites are used for validation. We under go this process for the regression model using only generic urban indicators **(A)** and the regression model using all variables **(B)**. The resulting distributions of variance explained (R^2) when the trained models are applied to separate validation data confirms that the full regression model accounts for an additional 10% of variance on average.

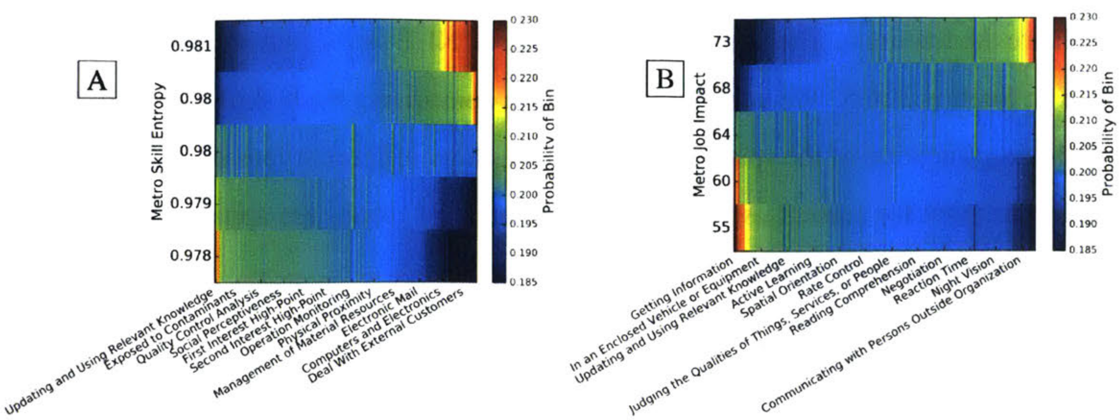


Figure B-12: The shape of skills indicating specialization (left) and resilience to expected job displacement (right) in cities is maintained when observing raw O*NET skills in place of the aggregate skills. The colors in each column indicate the probability of a city having the quality on the y-axis given an observation of a labor skill on the x-axis. We have labelled a few of the raw O*NET skills on the x-axis for reference.

B.5 Simplifying Jobs & Skills

In an effort to clearly identify how jobs contribute to labor specialization in larger cities, we identify aggregate job types based on common workplace skills. Previous studies, such as [218], examined the relationship between industry size and city size for various abstractions for industry according to NAICS. Here, we are seeking an organic representation of the forces in effect, and so we use K-means clustering based on the raw skill values for each job to identify five clusters of similar jobs (i.e. occupations are instances and the raw O*NET importance of each skill are features). These job groups represent collections of jobs which rely on similar skills for completion. The BLS jobs comprising each job type are shown in Section B.6.3. Note that our results and interpretations are consistent for anywhere from three to seven clusters (see B.6.3). This simplification of the space of jobs allows us to clearly understand which job types are disproportionately emphasized in large cities through the scaling behaviors of these job types.

We also seek to explain our results on the basis of workplace skills. To this end, we measure the correlation of raw skill values across all BLS jobs for each pair of O*NET skills and employ K-means clustering to identify ten groups of co-occurring skills (i.e. workplace skills are instances and the Pearson correlation of the raw O*NET importance of that skill to the importance of each other skill are the features). The complete lists of raw O*NET skills comprising each skill type are presented in Section B.6.5. We summarize the skills comprising each skill type with the groups' titles. This simplification of the space of skills clarifies how different types of skills explain our results, and trends that we present using these aggregated skill groups are apparent when reproduced using the raw O*NET skills instead.

B.5.1 O*NET Task Groups

An alternative simplification of the raw O*NET skills is the O*NET Task Groups, which represent collections of similar work activities. We provide the definitions for these task groups in Table B.3. These task groups have been used to investigate the task connectivity of urban labor markets in relation to employment growth [136]. In Figure B-13, we use these groups as alternative skill aggregations and assess which tasks indicate resilience to job displacement from automation in cities (Fig. B-13A), which tasks indicate occupational specialization in cities (Fig. B-13B), and which tasks indicate superlinear scaling of job types (Fig. B-13C). We find that Mental Process tasks are indicative of increased specialization in cities, increased resilience to job displacement in cities, and superlinear scaling of job types. On the other hand, Work Output tasks, which focuses on physical skills, indicate less specialization in cities, less resilience to job displacement in cities, and linear or sublinear scaling of job types. These findings are in agreement with our other results.

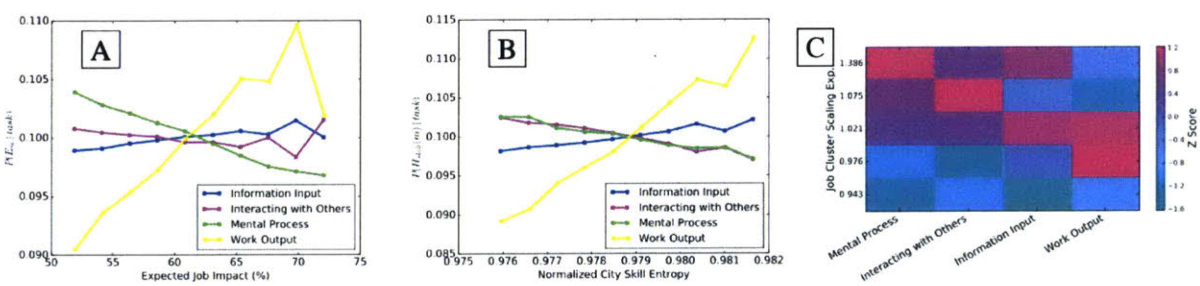


Figure B-13: The relationships between O*NET tasks expected job impact from automation, labor specialization, and the scaling of job types. **(A)** We bin cities according to their expected job impact from automation (x-axis). For each task (legend), we normalize the importance of that task across bins to a probability $P(E_m \mid task)$ representing how strongly that task indicates each level of job displacement (y-axis). **(B)** We bin cities according to their skill specialization (x- axis) and sum the importance of each task for each bin. For each task (leg- end), we normalize the importance of that task across bins to a probability $P(H_{skill}(m) \mid task)$ representing how strongly that task indicates each level of specialization (y-axis). **(C)** By summing the importance of each task to each job type, we assess how strongly a task indicates a scaling relationship according to its z score. For a given task, z scores are calculated according to the distribution of importance across job clusters.

O*NET Task Group	Job Impact Corr.	Log_{10} City Size Corr.
Mental Process	-0.86 ($< 10^{-113}$)	0.67 ($< 10^{-49}$)
Interacting with Others	-0.46 ($< 10^{-20}$)	0.13 (0.01)
Information Input	-0.082 (0.11)	0.34 ($< 10^{-11}$)
Work Output	0.69 ($< 10^{-53}$)	-0.37 ($< 10^{-12}$)

Table B.2: Summarizing the relationship between tasks, job impact from automation, and city size. In the middle (right) column, we present the Pearson correlation of the proportion of each task to the expected job impact (log 10 city size). We provide the associated p-values in parentheses

B.5.2 The Routineness of Tasks

Autor et al. [24, 25] identify workplace tasks according to their type and how routine the task is. They find that non-routine tasks are becoming increasingly important to workers relative to routine tasks. We provide the definitions for these task groups in Table B.5. In Figure B-14, we use these groups as alternative skill aggregations and assess which tasks indicate resilience to job impact from automation in cities (Fig. B-14A), which tasks indicate occupational specialization in cities (Fig. B-14B), and which tasks indicate superlinear scaling of job types (Fig. B-14C). We find that all non-routine tasks are indicative of increased specialization in cities and increased resilience to job impact in cities. Non-routine analytic tasks and non-routine interactive tasks are indicative of superlinear scaling of job types, while non-routine manual tasks indicate linear or sublinear scaling of job types. Routine tasks indicate less specialization in cities, less resilience to job impact in cities, and linear or sublinear scaling of job types. These findings are in agreement with our other results.

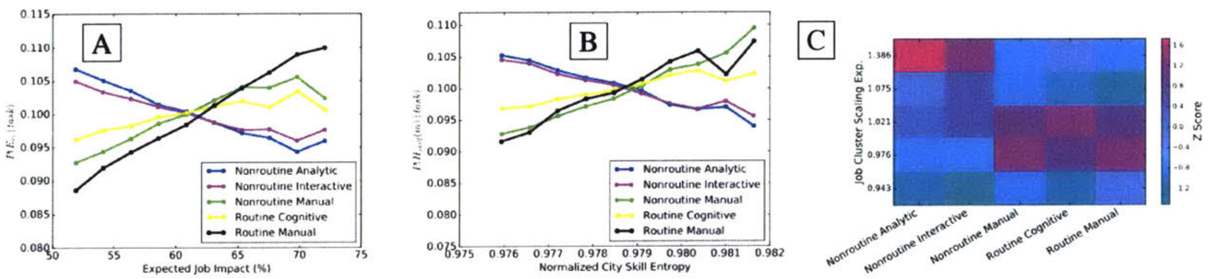


Figure B-14: The relationships between O*NET tasks expected job impact from automation, labor specialization, and the scaling of job types. **(A)** We bin cities according to their expected job impact from automation (x-axis). For each task (legend), we normalize the importance of that task across bins to a probability $P(E_m \mid task)$ representing how strongly that task indicates each level of job displacement (y-axis). **(B)** We bin cities according to their skill specialization (x- axis) and sum the importance of each task for each bin. For each task (leg- end), we normalize the importance of that task across bins to a probability $P(H_{skill}(m) \mid task)$ representing how strongly that task indicates each level of specialization (y-axis). **(C)** By summing the importance of each task to each job type, we assess how strongly a task indicates a scaling relationship according to its z score. For a given task, z scores are calculated according to the distribution of importance across job clusters.

Task Group	O*NET Skills
Information Input	Getting Information, Monitor Processes, Materials, or Surroundings, Identifying Objects, Actions, and Events, Inspecting Equipment, Structures, or Material, Estimating the Quantifiable Characteristics of Products, Events, or Information
Mental Process	Judging the Qualities of Things, Services, or People, Processing Information, Evaluating Information to Determine Compliance with Standards, Analyzing Data or Information, Making Decisions and Solving Problems, Thinking Creatively, Updating and Using Relevant Knowledge, Developing Objectives and Strategies, Scheduling Work and Activities, Organizing, Planning, and Prioritizing Work
Work Output	Performing General Physical Activities, Handling and Moving Objects, Controlling Machines and Processes, Operating Vehicles, Mechanized Devices, or Equipment, Interacting With Computers, Drafting, Laying Out, and Specifying Technical Devices. Parts. and Equipment, Repairing and Maintaining Mechanical Equipment, Repairing and Maintaining Electronic Equipment, Documenting or Recording Information
Interacting with Others	Interpreting the Meaning of Information for Others, Communicating with Supervisors, Peers, or Subordinates, Communicating with Persons Outside Organization, Establishing and Maintaining Interpersonal Relationships, Assisting and Caring for Others, Selling or Influencing Others, Resolving Conflicts and Negotiating with Others, Performing for or Working Directly with the Public, Coordinating the Work and Activities of Others, Developing and Building Teams, Training and Teaching Others, Guiding, Directing, and Motivating Subordinates, Coaching and Developing Others, Provide Consultation and Advice to Others, Performing Administrative Activities, Staffing Organizational Units, Monitoring and Controlling Resources

Table B.3: The O*NET skills comprising each Task Group.

O*NET Task Type	Job Impact Corr.	Log_{10} City Size Corr.
Non-routine Analytic	-0.79 ($< 10^{-80}$)	0.50 ($< 10^{-24}$)
Non-routine Interactive	-0.73 ($< 10^{-62}$)	0.52 ($< 10^{26}$)
Routine Cognitive	0.47 ($< 10^{-21}$)	-0.14 (0.005)
Non-routine Manual	0.64 ($< 10^{-43}$)	-0.30 ($< 10^{-8}$)
Routine Manual	0.83 ($< 10^{-97}$)	-0.49 ($< 10^{-23}$)

Table B.4: Summarizing the relationship between tasks, job impact, and city size. In the middle (right) column, we present the Pearson correlation of the proportion of each task to the expected job impact (log 10 city size). We provide the associated p-values in parentheses

Task Type	**O*NET Skills**
Non-routine Analytic	Mathematical Reasoning, Mathematics, Deductive Reasoning, Number Facility, Physics, Programming
Non-routine Interactive	Design, Administration and Management, Economics and Accounting, Equipment Selection, Estimating the Quantifiable Characteristics of Products, Events, or Information, Importance of Being Exact or Accurate, Management of Financial Resources, Management of Material Resources, Management of Personnel Resources, Organizing, Planning, and Prioritizing Work, Personnel and Human Resources, Quality Control Analysis, Sales and Marketing, Scheduling Work and Activities, Technology Design, Visualization
Routine Cognitive	Consequence of Error, Control Precision, Controlling Machines and Processes, Documenting/Recording Information, Evaluating Information to Determine Compliance with Standards, Inspecting Equipment, Structures, or Material, Operation and Control, Quality Control Analysis
Routine Manual	Finger Dexterity, Manual Dexterity, Arm-Hand Steadiness, Wrist-Finger Speed
Non-routine Manual	Reaction Time, Response Orientation, Cramped Work Space, Awkward Positions, Dynamic Flexibility, Spatial Orientation, Transportation, Coordination

Table B.5: The O*NET skills comprising each Task Type.

B.6 Data Tables

B.6.1 Cities Ordered by Expected Job Impact from Automation

Rank	Metro. Area	Exp. Job Impact (%)
1	San Jose-Sunnyvale-Santa Clara, CA	50.79
2	Washington-Arlington-Alexandria, DC-VA-MD-WV	51.85
3	Trenton-Ewing, NJ	52.71
4	Boston-Cambridge-Quincy, MA-NH	53.72
5	Durham-Chapel Hill, NC	53.85
6	Boulder, CO	54.07
7	Warner Robins, GA	54.69
8	Huntsville, AL	55.00
9	Bridgeport-Stamford-Norwalk, CT	55.33
10	Ithaca, NY	55.64
11	San Francisco-Oakland-Fremont, CA	55.84
12	Hartford-West Hartford-East Hartford, CT	56.17
13	Ann Arbor, MI	56.50
14	Corvallis, OR	56.63
15	Seattle-Tacoma-Bellevue, WA	57.23
16	Baltimore-Towson, MD	57.40
17	Madison, WI	57.49
18	New York-Northern New Jersey-Long Island, NY-NJ-PA	57.59
19	New Haven, CT	57.70
20	Minneapolis-St. Paul-Bloomington, MN-WI	57.85
21	Charlottesville, VA	57.93
22	Worcester, MA-CT	58.04
23	Albany-Schenectady-Troy, NY	58.09
24	Colorado Springs, CO	58.12
25	Denver-Aurora-Broomfield, CO	58.30
26	Burlington-South Burlington, VT	58.31
27	Raleigh-Cary, NC	58.62
28	Hinesville-Fort Stewart, GA	58.64
29	Springfield, MA-CT	58.71
30	Cedar Rapids, IA	58.73
31	Rochester, MN	58.93
32	Sacramento–Arden-Arcade–Roseville, CA	59.11
33	Richmond, VA	59.11
34	Tallahassee, FL	59.14
35	Bremerton-Silverdale, WA	59.18
36	Austin-Round Rock-San Marcos, TX	59.22

37	Portsmouth, NH-ME	59.24
38	Manchester, NH	59.36
39	Peoria, IL	59.38
40	Dayton, OH	59.39
41	Provo-Orem, UT	59.41
42	Tucson, AZ	59.47
43	Columbus, OH	59.51
44	Atlanta-Sandy Springs-Marietta, GA	59.54
45	Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	59.57
46	San Diego-Carlsbad-San Marcos, CA	59.61
47	Waterbury, CT	59.62
48	Detroit-Warren-Livonia, MI	59.62
49	Springfield, IL	59.69
50	Chicago-Joliet-Naperville, IL-IN-WI	59.73
51	Olympia, WA	59.73
52	Danbury, CT	59.77
53	Little Rock-North Little Rock-Conway, AR	59.77
54	Palm Bay-Melbourne-Titusville, FL	59.80
55	Albuquerque, NM	59.80
56	Des Moines-West Des Moines, IA	59.83
57	Portland-Vancouver-Hillsboro, OR-WA	59.86
58	Pittsfield, MA	59.88
59	Rochester, NY	59.90
60	Leominster-Fitchburg-Gardner, MA	59.97
61	Phoenix-Mesa-Glendale, AZ	59.99
62	Providence-Fall River-Warwick, RI-MA	60.05
63	Topeka, KS	60.08
64	Norwich-New London, CT-RI	60.10
65	Columbia, SC	60.19
66	Jefferson City, MO	60.19
67	Baton Rouge, LA	60.20
68	Charlotte-Gastonia-Rock Hill, NC-SC	60.28
69	Bakersfield-Delano, CA	60.29
70	Salt Lake City, UT	60.29
71	Pascagoula, MS	60.31
72	Virginia Beach-Norfolk-Newport News, VA-NC	60.32
73	Kennewick-Pasco-Richland, WA	60.32
74	Milwaukee-Waukesha-West Allis, WI	60.34
75	Fort Collins-Loveland, CO	60.35
76	Cincinnati-Middletown, OH-KY-IN	60.36
77	Lansing-East Lansing, MI	60.38

78	Bloomington-Normal, IL	60.44
79	Portland-South Portland-Biddeford, ME	60.51
80	Anchorage, AK	60.65
81	Rochester-Dover, NH-ME	60.68
82	Honolulu, HI	60.70
83	Los Angeles-Long Beach-Santa Ana, CA	60.72
84	Syracuse, NY	60.74
85	Houston-Sugar Land-Baytown, TX	60.76
86	Cleveland-Elyria-Mentor, OH	60.82
87	Omaha-Council Bluffs, NE-IA	60.90
88	Oklahoma City, OK	60.91
89	Charleston-North Charleston-Summerville, SC	60.92
90	Jackson, MS	60.93
91	Champaign-Urbana, IL	61.06
92	Iowa City, IA	61.06
93	Akron, OH	61.14
94	Santa Fe, NM	61.15
95	Ogden-Clearfield, UT	61.15
96	Yuma, AZ	61.18
97	St. Louis, MO-IL	61.23
98	Bismarck, ND	61.27
99	Las Cruces, NM	61.31
100	Kansas City, MO-KS	61.31
101	Boise City-Nampa, ID	61.32
102	Indianapolis-Carmel, IN	61.33
103	Salem, OR	61.39
104	Cumberland, MD-WV	61.52
105	Dallas-Fort Worth-Arlington, TX	61.55
106	Duluth, MN-WI	61.57
107	Harrisburg-Carlisle, PA	61.58
108	Santa Barbara-Santa Maria-Goleta, CA	61.60
109	Poughkeepsie-Newburgh-Middletown, NY	61.66
110	Athens-Clarke County, GA	61.66
111	Nashville-Davidson–Murfreesboro–Franklin, TN	61.68
112	Battle Creek, MI	61.69
113	New Bedford, MA	61.75
114	Bangor, ME	61.76
115	Lincoln, NE	61.82
116	Columbus, GA-AL	61.82
117	Gainesville, FL	61.83
118	Binghamton, NY	61.87

119	Oxnard-Thousand Oaks-Ventura, CA	61.89
120	Utica-Rome, NY	61.90
121	Vallejo-Fairfield, CA	61.95
122	Chattanooga, TN-GA	61.96
123	Kalamazoo-Portage, MI	61.96
124	Lexington-Fayette, KY	61.97
125	Pittsburgh, PA	61.99
126	Killeen-Temple-Fort Hood, TX	61.99
127	Barnstable Town, MA	62.01
128	Fayetteville-Springdale-Rogers, AR-MO	62.02
129	Augusta-Richmond County, GA-SC	62.08
130	Johnstown, PA	62.10
131	Knoxville, TN	62.13
132	Pueblo, CO	62.20
133	Fayetteville, NC	62.20
134	Allentown-Bethlehem-Easton, PA-NJ	62.25
135	Buffalo-Niagara Falls, NY	62.26
136	Santa Cruz-Watsonville, CA	62.26
137	Cheyenne, WY	62.34
138	Florence, SC	62.34
139	Huntington-Ashland, WV-KY-OH	62.39
140	Santa Rosa-Petaluma, CA	62.42
141	Saginaw-Saginaw Township North, MI	62.45
142	Tampa-St. Petersburg-Clearwater, FL	62.49
143	San Antonio-New Braunfels, TX	62.52
144	Fresno, CA	62.56
145	Alexandria, LA	62.56
146	Tulsa, OK	62.58
147	Lynchburg, VA	62.59
148	Eugene-Springfield, OR	62.60
149	Morgantown, WV	62.60
150	Hagerstown-Martinsburg, MD-WV	62.64
151	Wichita, KS	62.67
152	Merced, CA	62.68
153	Vineland-Millville-Bridgeton, NJ	62.68
154	Rockford, IL	62.69
155	Albany, GA	62.73
156	Jackson, MI	62.75
157	Birmingham-Hoover, AL	62.77
158	Mankato-North Mankato, MN	62.78
159	Grand Rapids-Wyoming, MI	62.79

160	Crestview-Fort Walton Beach-Destin, FL	62.81
161	Salinas, CA	62.84
162	Oshkosh-Neenah, WI	62.84
163	Pine Bluff, AR	62.89
164	Toledo, OH	62.90
165	Green Bay, WI	62.91
166	College Station-Bryan, TX	62.92
167	Fairbanks, AK	62.96
168	Lewiston-Auburn, ME	63.03
169	Winston-Salem, NC	63.04
170	Pocatello, ID	63.05
171	Madera-Chowchilla, CA	63.05
172	Rome, GA	63.13
173	State College, PA	63.18
174	Evansville, IN-KY	63.21
175	Johnson City, TN	63.24
176	McAllen-Edinburg-Mission, TX	63.25
177	Beaumont-Port Arthur, TX	63.25
178	Clarksville, TN-KY	63.26
179	Yakima, WA	63.27
180	Davenport-Moline-Rock Island, IA-IL	63.28
181	Fargo, ND-MN	63.29
182	San Luis Obispo-Paso Robles, CA	63.32
183	Flagstaff, AZ	63.33
184	Visalia-Porterville, CA	63.34
185	St. Cloud, MN	63.36
186	Reading, PA	63.37
187	Salisbury, MD	63.39
188	Springfield, OH	63.41
189	New Orleans-Metairie-Kenner, LA	63.44
190	Memphis, TN-MS-AR	63.48
191	Kingston, NY	63.52
192	Canton-Massillon, OH	63.54
193	Spokane, WA	63.55
194	Miami-Fort Lauderdale-Pompano Beach, FL	63.56
195	Monroe, LA	63.57
196	Niles-Benton Harbor, MI	63.57
197	Jackson, TN	63.59
198	San Juan-Caguas-Guaynabo, PR	63.60
199	Lawton, OK	63.63
200	Flint, MI	63.63

201	Charleston, WV	63.63
202	Sherman-Denison, TX	63.64
203	Blacksburg-Christiansburg-Radford, VA	63.66
204	Yuba City, CA	63.67
205	Roanoke, VA	63.69
206	Manhattan, KS	63.76
207	Amarillo, TX	63.81
208	Steubenville-Weirton, OH-WV	63.86
209	Wilmington, NC	63.87
210	Greenville-Mauldin-Easley, SC	63.87
211	Pensacola-Ferry Pass-Brent, FL	63.92
212	Corpus Christi, TX	63.96
213	Tyler, TX	64.00
214	Kingsport-Bristol-Bristol, TN-VA	64.00
215	Redding, CA	64.00
216	Ames, IA	64.01
217	Carson City, NV	64.02
218	Jacksonville, FL	64.02
219	Appleton, WI	64.05
220	Decatur, IL	64.09
221	Wheeling, WV-OH	64.09
222	Scranton–Wilkes-Barre, PA	64.09
223	Chico, CA	64.10
224	Louisville-Jefferson County, KY-IN	64.11
225	Macon, GA	64.11
226	Eau Claire, WI	64.13
227	Decatur, AL	64.15
228	Idaho Falls, ID	64.15
229	Shreveport-Bossier City, LA	64.18
230	Springfield, MO	64.20
231	Medford, OR	64.20
232	Bloomington, IN	64.22
233	La Crosse, WI-MN	64.22
234	Billings, MT	64.22
235	Bellingham, WA	64.23
236	Glens Falls, NY	64.25
237	Missoula, MT	64.30
238	Coeur d'Alene, ID	64.32
239	Orlando-Kissimmee-Sanford, FL	64.33
240	South Bend-Mishawaka, IN-MI	64.34
241	Ponce, PR	64.36

242	Holland-Grand Haven, MI	64.37
243	Greensboro-High Point, NC	64.37
244	Lewiston, ID-WA	64.37
245	Mansfield, OH	64.38
246	Racine, WI	64.38
247	El Paso, TX	64.39
248	Bowling Green, KY	64.39
249	York-Hanover, PA	64.40
250	Waterloo-Cedar Falls, IA	64.42
251	Cleveland, TN	64.43
252	Gulfport-Biloxi, MS	64.45
253	Port St. Lucie, FL	64.46
254	Elizabethtown, KY	64.48
255	Brownsville-Harlingen, TX	64.48
256	Abilene, TX	64.49
257	Stockton, CA	64.54
258	Greeley, CO	64.57
259	Rocky Mount, NC	64.64
260	Longview, TX	64.64
261	Lima, OH	64.66
262	Spartanburg, SC	64.72
263	Parkersburg-Marietta-Vienna, WV-OH	64.76
264	Riverside-San Bernardino-Ontario, CA	64.79
265	Lubbock, TX	64.82
266	Lawrence, KS	64.84
267	Kankakee-Bradley, IL	64.91
268	Wichita Falls, TX	64.96
269	Terre Haute, IN	64.97
270	El Centro, CA	65.00
271	Greenville, NC	65.02
272	Erie, PA	65.03
273	Victoria, TX	65.05
274	Anderson, SC	65.06
275	Atlantic City-Hammonton, NJ	65.06
276	Mount Vernon-Anacortes, WA	65.08
277	Farmington, NM	65.09
278	Mobile, AL	65.10
279	Prescott, AZ	65.14
280	Hanford-Corcoran, CA	65.16
281	Sumter, SC	65.23
282	Lancaster, PA	65.24

283	Asheville, NC	65.24
284	North Port-Bradenton-Sarasota, FL	65.28
285	Fort Wayne, IN	65.31
286	Fort Smith, AR-OK	65.33
287	Janesville, WI	65.35
288	Montgomery, AL	65.35
289	Anderson, IN	65.35
290	Winchester, VA-WV	65.35
291	Lake Havasu City - Kingman, AZ	65.35
292	Savannah, GA	65.36
293	Altoona, PA	65.37
294	Modesto, CA	65.40
295	Youngstown-Warren-Boardman, OH-PA	65.41
296	Sheboygan, WI	65.47
297	Grand Forks, ND-MN	65.51
298	Joplin, MO	65.52
299	Bay City, MI	65.54
300	Columbia, MO	65.54
301	Sioux Falls, SD	65.55
302	Lakeland-Winter Haven, FL	65.58
303	Bend, OR	65.58
304	Muskegon-Norton Shores, MI	65.61
305	St. George, UT	65.64
306	Panama City-Lynn Haven-Panama City Beach, FL	65.66
307	Houma-Bayou Cane-Thibodaux, LA	65.67
308	Midland, TX	65.67
309	Dover, DE	65.69
310	Texarkana-Texarkana, TX-AR	65.72
311	Logan, UT-ID	65.75
312	Aguadilla-Isabela-San Sebastian, PR	65.75
313	Casper, WY	65.76
314	Goldsboro, NC	65.78
315	Sioux City, IA-NE-SD	65.82
316	Brunswick, GA	65.87
317	Longview, WA	65.92
318	Monroe, MI	65.93
319	Guayama, PR	65.93
320	Waco, TX	65.94
321	Hattiesburg, MS	65.95
322	Reno-Sparks, NV	65.98
323	Muncie, IN	66.01

324	Fond du Lac, WI	66.01
325	Lake Charles, LA	66.03
326	Rapid City, SD	66.10
327	Valdosta, GA	66.12
328	Dubuque, IA	66.13
329	Wenatchee-East Wenatchee, WA	66.16
330	Lafayette, IN	66.16
331	St. Joseph, MO-KS	66.16
332	Morristown, TN	66.20
333	Sandusky, OH	66.21
334	Owensboro, KY	66.21
335	Wausau, WI	66.23
336	Elmira, NY	66.25
337	Grand Junction, CO	66.25
338	Tuscaloosa, AL	66.32
339	Hickory-Lenoir-Morganton, NC	66.37
340	Jonesboro, AR	66.48
341	Danville, VA	66.50
342	Florence-Muscle Shoals, AL	66.51
343	Cape Coral-Fort Myers, FL	66.55
344	Cape Girardeau-Jackson, MO-IL	66.56
345	Deltona-Daytona Beach-Ormond Beach, FL	66.57
346	Dothan, AL	66.64
347	Great Falls, MT	66.75
348	Naples-Marco Island, FL	66.93
349	Columbus, IN	66.93
350	Gainesville, GA	66.93
351	Kokomo, IN	66.94
352	Hot Springs, AR	66.95
353	Lafayette, LA	66.97
354	Williamsport, PA	66.97
355	Ocean City, NJ	67.24
356	Anniston-Oxford, AL	67.53
357	Sebastian-Vero Beach, FL	67.55
358	Auburn-Opelika, AL	67.77
359	Odessa, TX	67.78
360	Las Vegas-Paradise, NV	67.79
361	Lebanon, PA	67.89
362	Burlington, NC	67.94
363	Danville, IL	67.94
364	San Angelo, TX	67.96

365	Ocala, FL	68.23
366	Laredo, TX	68.75
367	Gadsden, AL	68.87
368	San German-Cabo Rojo, PR	68.88
369	Napa, CA	68.88
370	Palm Coast, FL	68.98
371	Yauco, PR	69.06
372	Dalton, GA	69.07
373	Jacksonville, NC	69.39
374	Michigan City-La Porte, IN	69.40
375	Harrisonburg, VA	69.84
376	Punta Gorda, FL	70.03
377	Fajardo, PR	70.04
378	Elkhart-Goshen, IN	70.28
379	Myrtle Beach-North Myrtle Beach-Conway, SC	70.80
380	Mayaguez, PR	73.14

B.6.2 Relating City Trends to BLS Jobs

We present BLS jobs ordered by decreasing skill specialization in Table B.2.2. We also provide the scaling exponent of each BLS job, along with the Pearson correlation of the relative abundance of each job to the expected job impact from automation (discussed below) across cities. p-values for the correlations are presented in parentheses.

Rank	Job Title	H_j	β	Corr. to Job Impact
1	Statisticians	0.949	0.748	-0.434 (0)
2	Telemarketers	0.951	0.955	0.233 (0)
3	Securities, Commodities, and Financial Services Sales Agents	0.955	1.128	-0.265 (0)
4	Loan Interviewers and Clerks	0.955	1.005	0.060 ($< 10^{-54}$)
5	Actuaries	0.956	0.756	-0.193 ($< 10^{-103}$)
6	Court, Municipal, and License Clerks	0.956	0.762	0.176 (0)
7	Court Reporters	0.957	0.637	0.140 ($< 10^{-65}$)
8	Credit Counselors	0.957	0.825	0.220 ($< 10^{-230}$)
9	Medical Transcriptionists	0.957	0.696	0.341 (0)
10	Financial Managers	0.958	1.103	-0.493 (0)
11	Training and Development Specialists	0.958	1.054	-0.316 (0)
12	Billing and Posting Clerks	0.958	0.972	0.145 (0)
13	Credit Authorizers, Checkers, and Clerks	0.958	0.792	0.192 ($< 10^{-239}$)
14	Legal Secretaries	0.958	0.990	0.053 ($< 10^{-41}$)
15	Clinical, Counseling, and School Psychologists	0.958	0.861	-0.087 ($< 10^{-103}$)
16	Operations Research Analysts	0.959	0.985	-0.282 (0)
17	Eligibility Interviewers, Government Programs	0.959	0.821	0.119 ($< 10^{-180}$)

18	Bookkeeping, Accounting, and Auditing Clerks	0.959	0.958	0.098 ($< 10^{-183}$)
19	Demonstrators and Product Promoters	0.959	0.718	0.283 (0)
20	Marriage and Family Therapists	0.959	0.610	0.280 (0)
21	Financial Examiners	0.959	0.910	0.006 ($< 10^{0}$)
22	Insurance Claims and Policy Processing Clerks	0.959	1.031	-0.055 ($< 10^{-33}$)
23	Judges, Magistrate Judges, and Magistrates	0.959	0.601	0.327 (0)
24	Payroll and Timekeeping Clerks	0.960	0.949	0.135 (0)
25	Accountants and Auditors	0.960	1.111	-0.458 (0)
26	Cost Estimators	0.960	0.986	0.059 ($< 10^{-61}$)
27	Administrative Law Judges, Adjudicators, and Hearing Officers	0.960	0.572	0.132 ($< 10^{-63}$)
28	Word Processors and Typists	0.960	0.600	0.226 (0)
29	Budget Analysts	0.960	0.795	-0.123 ($< 10^{-128}$)
30	Paralegals and Legal Assistants	0.960	1.093	-0.213 (0)
31	Office Clerks, General	0.960	0.934	0.196 (0)
32	Computer Programmers	0.960	1.168	-0.501 (0)
33	Procurement Clerks	0.961	0.845	0.142 ($< 10^{-261}$)
34	Crossing Guards	0.961	0.738	0.277 (0)
35	Executive Secretaries and Executive Administrative Assistants	0.961	1.093	-0.341 (0)
36	Real Estate Brokers	0.961	0.706	0.249 ($< 10^{-321}$)
37	Insurance Sales Agents	0.961	1.021	0.051 ($< 10^{-47}$)
38	Mental Health Counselors	0.961	0.814	-0.008 ($< 10^{0}$)
39	Loan Officers	0.961	1.033	-0.012 ($< 10^{-2}$)
40	Lawyers	0.961	1.226	-0.473 (0)
41	Financial Analysts	0.962	1.270	-0.529 (0)
42	Travel Agents	0.962	1.056	0.042 ($< 10^{-13}$)
43	Statistical Assistants	0.962	0.402	0.313 ($< 10^{-274}$)
44	Health Educators	0.962	0.793	-0.084 ($< 10^{-73}$)
45	Title Examiners, Abstractors, and Searchers	0.962	0.806	0.313 (0)
46	Software Developers, Applications	0.963	1.304	-0.663 (0)
47	Compensation, Benefits, and Job Analysis Specialists	0.963	1.037	-0.391 (0)
48	Medical Records and Health Information Technicians	0.963	0.893	0.073 ($< 10^{-92}$)
49	Educational, Guidance, School, and Vocational Counselors	0.963	0.919	-0.097 ($< 10^{-170}$)
50	Personal Financial Advisors	0.963	1.132	-0.356 (0)
51	Public Relations and Fundraising Managers	0.963	0.909	-0.405 (0)
52	Healthcare Social Workers	0.964	0.885	-0.053 ($< 10^{-46}$)
53	Psychologists, All Other	0.964	0.552	0.194 ($< 10^{-102}$)
54	Bill and Account Collectors	0.964	1.129	0.028 ($< 10^{-13}$)
55	Market Research Analysts and Marketing Specialists	0.964	1.237	-0.575 (0)
56	Compensation and Benefits Managers	0.964	0.877	-0.449 (0)
57	Software Developers, Systems Software	0.964	1.233	-0.612 (0)
58	Human Resources Assistants, Except Payroll and Timekeeping	0.964	0.937	0.026 ($< 10^{-11}$)

59	Elementary School Teachers, Except Special Education	0.964	0.855	0.226 (0)
60	Human Resources Managers	0.964	1.025	-0.348 (0)
61	Managers, All Other	0.964	1.040	-0.298 (0)
62	Technical Writers	0.964	0.888	-0.382 (0)
63	Library Technicians	0.964	0.765	-0.003 ($< 10^{0}$)
64	Speech-Language Pathologists	0.964	0.884	0.080 ($< 10^{-105}$)
65	Photographic Process Workers and Processing Machine Operators	0.964	0.695	0.148 ($< 10^{-107}$)
66	Chief Executives	0.964	0.977	-0.072 ($< 10^{-88}$)
67	Credit Analysts	0.964	1.039	-0.064 ($< 10^{-31}$)
68	Receptionists and Information Clerks	0.965	0.955	0.136 (0)
69	Tax Examiners and Collectors, and Revenue Agents	0.965	0.843	-0.054 ($< 10^{-24}$)
70	Education Administrators, Postsecondary	0.965	0.792	-0.000 ($< 10^{0}$)
71	Switchboard Operators, Including Answering Service	0.965	0.938	0.247 (0)
72	Financial Specialists, All Other	0.965	0.966	-0.258 (0)
73	Insurance Underwriters	0.965	0.887	-0.053 ($< 10^{-18}$)
74	Secretaries and Administrative Assistants, Except Legal, Medical, and Executive	0.965	0.921	0.221 (0)
75	Advertising Sales Agents	0.965	0.991	0.066 ($< 10^{-64}$)
76	Security Guards	0.965	1.161	0.053 ($< 10^{-51}$)
77	Producers and Directors	0.965	1.057	-0.171 ($< 10^{-267}$)
78	Claims Adjusters, Examiners, and Investigators	0.965	1.137	-0.104 ($< 10^{-134}$)
79	Brokerage Clerks	0.965	1.038	-0.082 ($< 10^{-42}$)
80	First-Line Supervisors of Non-Retail Sales Workers	0.965	1.088	-0.075 ($< 10^{-99}$)
81	Interviewers, Except Eligibility and Loan	0.965	0.934	0.038 ($< 10^{-17}$)
82	Merchandise Displayers and Window Trimmers	0.965	0.870	0.218 (0)
83	Molding, Coremaking, and Casting Machine Setters, Operators, and Tenders, Metal and Plastic	0.965	0.624	0.288 (0)
84	New Accounts Clerks	0.966	0.701	0.396 (0)
85	Transportation, Storage, and Distribution Managers	0.966	0.961	0.140 ($< 10^{-269}$)
86	Marketing Managers	0.966	1.163	-0.549 (0)
87	Public Relations Specialists	0.966	1.052	-0.398 (0)
88	Education Administrators, Elementary and Secondary School	0.966	0.885	0.086 ($< 10^{-135}$)
89	Kindergarten Teachers, Except Special Education	0.966	0.841	0.217 (0)
90	Writers and Authors	0.966	0.911	-0.384 (0)
91	Police, Fire, and Ambulance Dispatchers	0.966	0.781	0.246 (0)
92	Massage Therapists	0.966	0.852	0.163 ($< 10^{-244}$)
93	Sales Representatives, Wholesale and Manufacturing, Except Technical and Scientific Products	0.966	1.082	0.155 (0)
94	Ushers, Lobby Attendants, and Ticket Takers	0.966	1.084	0.114 ($< 10^{-88}$)
95	Music Directors and Composers	0.966	0.555	0.297 (0)
96	Probation Officers and Correctional Treatment Specialists	0.966	0.769	0.187 ($< 10^{-308}$)
97	Property, Real Estate, and Community Association Managers	0.966	0.986	0.173 (0)
98	Radio and Television Announcers	0.966	0.557	0.417 (0)

99	Health Diagnosing and Treating Practitioners, All Other	0.966	0.786	-0.155 ($< 10^{-105}$)
100	Editors	0.966	0.956	-0.285 (0)
101	Database Administrators	0.967	1.087	-0.229 (0)
102	Order Clerks	0.967	1.001	0.104 ($< 10^{-157}$)
103	Business Operations Specialists, All Other	0.967	1.117	-0.468 (0)
104	Management Analysts	0.967	1.205	-0.460 (0)
105	Mechanical Drafters	0.967	0.751	0.347 (0)
106	Mental Health and Substance Abuse Social Workers	0.967	0.749	0.098 ($< 10^{-118}$)
107	Advertising and Promotions Managers	0.967	0.940	-0.155 ($< 10^{-121}$)
108	Cartographers and Photogrammetrists	0.967	0.446	-0.146 ($< 10^{-68}$)
109	Data Entry Keyers	0.967	1.076	-0.090 ($< 10^{-123}$)
110	Medical Secretaries	0.967	0.906	0.114 ($< 10^{-241}$)
111	Computer Hardware Engineers	0.967	0.689	-0.355 (0)
112	Rehabilitation Counselors	0.967	0.770	0.010 ($< 10^{0}$)
113	Instructional Coordinators	0.967	0.884	-0.119 ($< 10^{-209}$)
114	Cargo and Freight Agents	0.967	0.869	0.279 (0)
115	Stock Clerks and Order Fillers	0.967	0.936	0.307 (0)
116	Physical Scientists, All Other	0.967	0.466	-0.231 ($< 10^{-137}$)
117	Directors, Religious Activities and Education	0.967	0.489	0.243 ($< 10^{-206}$)
118	Secondary School Teachers, Except Special and Career/Technical Education	0.967	0.918	0.103 ($< 10^{-175}$)
119	Special Education Teachers, Secondary School	0.967	0.850	0.064 ($< 10^{-52}$)
120	Education Administrators, Preschool and Childcare Center/Program	0.967	0.819	-0.227 (0)
121	Parts Salespersons	0.968	0.847	0.384 (0)
122	Weighers, Measurers, Checkers, and Samplers, Recordkeeping	0.968	0.749	0.350 (0)
123	Musicians and Singers	0.968	0.738	0.077 ($< 10^{-25}$)
124	Social and Community Service Managers	0.968	0.835	-0.153 (0)
125	Training and Development Managers	0.968	0.894	-0.233 ($< 10^{-313}$)
126	Social Scientists and Related Workers, All Other	0.968	0.536	-0.182 ($< 10^{-144}$)
127	Interpreters and Translators	0.968	0.772	0.139 ($< 10^{-138}$)
128	Meeting, Convention, and Event Planners	0.968	0.933	-0.113 ($< 10^{-131}$)
129	Child, Family, and School Social Workers	0.968	0.883	0.011 ($< 10^{-2}$)
130	Bailiffs	0.968	0.409	0.397 (0)
131	Mail Clerks and Mail Machine Operators, Except Postal Service	0.968	0.952	-0.023 ($< 10^{-5}$)
132	Biological Scientists, All Other	0.968	0.534	-0.082 ($< 10^{-27}$)
133	Special Education Teachers, Middle School	0.968	0.820	0.120 ($< 10^{-148}$)
134	Appraisers and Assessors of Real Estate	0.968	0.765	0.361 (0)
135	Tellers	0.968	0.850	0.458 (0)
136	Teacher Assistants	0.968	0.894	-0.076 ($< 10^{-109}$)
137	Network and Computer Systems Administrators	0.968	1.149	-0.564 (0)
138	First-Line Supervisors of Office and Administrative Support Workers	0.968	0.989	-0.028 ($< 10^{-15}$)
139	Manicurists and Pedicurists	0.968	0.942	-0.213 ($< 10^{-168}$)

140	Concierges	0.968	0.861	0.291 ($< 10^{-292}$)
141	Food and Tobacco Roasting, Baking, and Drying Machine Operators and Tenders	0.968	0.384	0.364 (0)
142	Print Binding and Finishing Workers	0.968	0.667	0.306 (0)
143	Sales Managers	0.968	1.126	-0.267 (0)
144	Curators	0.968	0.518	-0.147 ($< 10^{-72}$)
145	Reporters and Correspondents	0.969	0.762	0.061 ($< 10^{-26}$)
146	Waiters and Waitresses	0.969	0.948	0.306 (0)
147	Pharmacists	0.969	0.921	0.179 (0)
148	Bus Drivers, School or Special Client	0.969	0.887	0.151 (0)
149	Locker Room, Coatroom, and Dressing Room Attendants	0.969	0.689	0.403 (0)
150	Social and Human Service Assistants	0.969	0.901	-0.105 ($< 10^{-192}$)
151	Graphic Designers	0.969	1.063	-0.304 (0)
152	Furnace, Kiln, Oven, Drier, and Kettle Operators and Tenders	0.969	0.433	0.368 (0)
153	Furniture Finishers	0.969	0.276	0.416 (0)
154	Clergy	0.969	0.693	0.072 ($< 10^{-45}$)
155	Purchasing Managers	0.969	0.924	-0.119 ($< 10^{-146}$)
156	Adult Basic and Secondary Education and Literacy Teachers and Instructors	0.969	0.569	0.187 ($< 10^{-241}$)
157	Reservation and Transportation Ticket Agents and Travel Clerks	0.969	1.078	0.187 ($< 10^{-178}$)
158	Skincare Specialists	0.969	0.798	0.304 (0)
159	Art Directors	0.969	0.974	-0.186 ($< 10^{-176}$)
160	Customer Service Representatives	0.969	1.069	-0.115 ($< 10^{-249}$)
161	Sales Representatives, Wholesale and Manufacturing, Technical and Scientific Products	0.969	1.189	-0.409 (0)
162	Team Assemblers	0.969	0.882	0.306 (0)
163	Computer and Information Systems Managers	0.969	1.219	-0.718 (0)
164	Hotel, Motel, and Resort Desk Clerks	0.969	0.788	0.313 (0)
165	Aerospace Engineers	0.969	0.478	-0.236 ($< 10^{-164}$)
166	Health and Safety Engineers, Except Mining Safety Engineers and Inspectors	0.969	0.634	0.133 ($< 10^{-94}$)
167	Tax Preparers	0.969	0.845	0.214 (0)
168	Private Detectives and Investigators	0.969	0.555	0.194 ($< 10^{-107}$)
169	Computer and Information Research Scientists	0.969	0.440	-0.343 ($< 10^{-320}$)
170	Civil Engineering Technicians	0.970	0.721	0.148 ($< 10^{-256}$)
171	Printing Press Operators	0.970	0.930	0.126 ($< 10^{-243}$)
172	Biological Technicians	0.970	0.632	-0.081 ($< 10^{-44}$)
173	Milling and Planing Machine Setters, Operators, and Tenders, Metal and Plastic	0.970	0.538	0.315 (0)
174	Electrical and Electronics Drafters	0.970	0.798	0.001 ($< 10^{0}$)
175	Medical Scientists, Except Epidemiologists	0.970	0.937	-0.475 (0)
176	Insurance Appraisers, Auto Damage	0.970	0.490	0.332 ($< 10^{-297}$)
177	Real Estate Sales Agents	0.970	0.929	0.247 (0)
178	Library Assistants, Clerical	0.970	0.777	0.067 ($< 10^{-60}$)
179	Environmental Scientists and Specialists, Including Health	0.970	0.796	-0.131 ($< 10^{-200}$)
180	Counter Attendants, Cafeteria, Food Concession, and Coffee Shop	0.970	0.958	0.099 ($< 10^{-137}$)

181	Civil Engineers	0.970	1.039	-0.233 (0)
182	Librarians	0.970	0.873	-0.031 ($< 10^{-16}$)
183	Construction Managers	0.970	1.027	-0.002 ($< 10^{0}$)
184	Mechanical Engineers	0.970	0.971	-0.166 (0)
185	Commercial and Industrial Designers	0.970	0.644	0.224 ($< 10^{-251}$)
186	Hairdressers, Hairstylists, and Cosmetologists	0.970	0.952	0.013 ($< 10^{-2}$)
187	Recreational Therapists	0.970	0.485	0.222 ($< 10^{-260}$)
188	Couriers and Messengers	0.970	0.834	0.236 (0)
189	Helpers–Production Workers	0.970	0.864	0.307 (0)
190	Lathe and Turning Machine Tool Setters, Operators, and Tenders, Metal and Plastic	0.970	0.503	0.317 (0)
191	Residential Advisors	0.970	0.692	0.209 (0)
192	Psychiatric Technicians	0.970	0.412	0.206 ($< 10^{-182}$)
193	Film and Video Editors	0.971	0.916	0.089 ($< 10^{-22}$)
194	Substance Abuse and Behavioral Disorder Counselors	0.971	0.726	0.125 ($< 10^{-180}$)
195	Logisticians	0.971	0.936	-0.237 (0)
196	Architectural and Civil Drafters	0.971	0.934	0.022 ($< 10^{-5}$)
197	Chemical Technicians	0.971	0.760	0.155 ($< 10^{-221}$)
198	Biomedical Engineers	0.971	0.645	-0.565 (0)
199	Wholesale and Retail Buyers, Except Farm Products	0.971	0.986	-0.048 ($< 10^{-27}$)
200	Computer Systems Analysts	0.971	1.277	-0.584 (0)
201	Engineers, All Other	0.971	0.894	-0.262 (0)
202	Chemical Engineers	0.971	0.524	0.207 ($< 10^{-215}$)
203	Occupational Therapists	0.971	0.879	0.037 ($< 10^{-21}$)
204	Pharmacy Aides	0.971	0.627	0.399 (0)
205	Middle School Teachers, Except Special and Career/Technical Education	0.971	0.926	0.138 ($< 10^{-302}$)
206	Purchasing Agents, Except Wholesale, Retail, and Farm Products	0.971	1.047	-0.296 (0)
207	Computer Operators	0.971	0.858	0.041 ($< 10^{-14}$)
208	Postmasters and Mail Superintendents	0.971	0.348	0.344 (0)
209	Fitness Trainers and Aerobics Instructors	0.971	0.952	-0.204 (0)
210	Administrative Services Managers	0.971	1.020	-0.288 (0)
211	Medical Assistants	0.971	0.951	0.124 ($< 10^{-287}$)
212	Urban and Regional Planners	0.971	0.670	0.031 ($< 10^{-7}$)
213	Medical and Health Services Managers	0.971	0.924	-0.178 (0)
214	Natural Sciences Managers	0.971	0.711	-0.371 (0)
215	Preschool Teachers, Except Special Education	0.972	0.956	-0.145 (0)
216	Dietitians and Nutritionists	0.972	0.844	0.031 ($< 10^{-11}$)
217	Production, Planning, and Expediting Clerks	0.972	1.022	0.002 ($< 10^{0}$)
218	Lifeguards, Ski Patrol, and Other Recreational Protective Service Workers	0.972	0.908	0.151 ($< 10^{-241}$)
219	Floral Designers	0.972	0.751	0.296 (0)
220	Postal Service Mail Carriers	0.972	0.891	0.253 (0)
221	Multimedia Artists and Animators	0.972	1.012	-0.244 ($< 10^{-196}$)

222	Architectural and Engineering Managers	0.972	1.039	-0.390 (0)
223	Chemical Equipment Operators and Tenders	0.972	0.508	0.305 (0)
224	Soil and Plant Scientists	0.972	0.166	0.116 ($< 10^{-38}$)
225	Sewing Machine Operators	0.972	0.743	0.277 (0)
226	Chemists	0.972	0.881	-0.200 (0)
227	Psychiatric Aides	0.972	0.322	0.331 (0)
228	Inspectors, Testers, Sorters, Samplers, and Weighers	0.972	0.906	0.267 (0)
229	Tour Guides and Escorts	0.972	0.551	0.363 (0)
230	Maids and Housekeeping Cleaners	0.972	0.907	0.260 (0)
231	Compliance Officers	0.972	1.039	-0.111 ($< 10^{-217}$)
232	Microbiologists	0.972	0.531	-0.128 ($< 10^{-47}$)
233	Woodworking Machine Setters, Operators, and Tenders, Except Sawing	0.972	0.410	0.492 (0)
234	Funeral Attendants	0.973	0.509	0.452 (0)
235	Sawing Machine Setters, Operators, and Tenders, Wood	0.973	0.320	0.472 (0)
236	Paper Goods Machine Setters, Operators, and Tenders	0.973	0.514	0.251 (0)
237	Cutting and Slicing Machine Setters, Operators, and Tenders	0.973	0.514	0.411 (0)
238	Helpers–Electricians	0.973	0.640	0.441 (0)
239	Electrical and Electronic Equipment Assemblers	0.973	0.914	-0.030 ($< 10^{-8}$)
240	Postal Service Mail Sorters, Processors, and Processing Machine Operators	0.973	0.975	-0.023 ($< 10^{-5}$)
241	Interior Designers	0.973	1.038	0.080 ($< 10^{-40}$)
242	Physical Therapists	0.973	0.934	-0.057 ($< 10^{-56}$)
243	Industrial Engineering Technicians	0.973	0.600	0.136 ($< 10^{-155}$)
244	First-Line Supervisors of Correctional Officers	0.973	0.456	0.195 ($< 10^{-236}$)
245	Hosts and Hostesses, Restaurant, Lounge, and Coffee Shop	0.973	0.946	0.246 (0)
246	Buyers and Purchasing Agents, Farm Products	0.973	0.370	0.335 (0)
247	First-Line Supervisors of Personal Service Workers	0.973	0.913	0.090 ($< 10^{-142}$)
248	Amusement and Recreation Attendants	0.973	0.979	0.232 (0)
249	Coaches and Scouts	0.973	0.876	-0.014 ($< 10^{-2}$)
250	Self-Enrichment Education Teachers	0.973	0.948	-0.213 (0)
251	Medical and Clinical Laboratory Technologists	0.973	0.905	0.237 (0)
252	Pharmacy Technicians	0.973	0.882	0.310 (0)
253	Food Scientists and Technologists	0.973	0.373	0.268 ($< 10^{-216}$)
254	Painting, Coating, and Decorating Workers	0.973	0.570	0.519 (0)
255	Sales Engineers	0.973	1.004	-0.442 (0)
256	Food Servers, Nonrestaurant	0.973	0.912	0.064 ($< 10^{-60}$)
257	Geoscientists, Except Hydrologists and Geographers	0.974	0.515	0.320 (0)
258	Electronics Engineers, Except Computer	0.974	1.002	-0.350 (0)
259	Optometrists	0.974	0.813	0.175 ($< 10^{-216}$)
260	Occupational Therapy Assistants	0.974	0.583	0.323 (0)
261	Personal Care Aides	0.974	0.877	0.100 ($< 10^{-169}$)
262	Postal Service Clerks	0.974	0.814	0.219 (0)

263	Painters, Transportation Equipment	0.974	0.701	0.350 (0)
264	Food Batchmakers	0.974	0.663	0.311 (0)
265	Welding, Soldering, and Brazing Machine Setters, Operators, and Tenders	0.974	0.443	0.240 (0)
266	Dining Room and Cafeteria Attendants and Bartender Helpers	0.974	1.019	0.139 ($< 10^{-312}$)
267	Environmental Engineers	0.974	0.780	-0.154 ($< 10^{-199}$)
268	Industrial Engineers	0.974	0.899	0.080 ($< 10^{-94}$)
269	Electrical Engineers	0.974	1.020	-0.403 (0)
270	Industrial Production Managers	0.974	0.840	0.256 (0)
271	Pressers, Textile, Garment, and Related Materials	0.974	0.749	0.351 (0)
272	Materials Engineers	0.974	0.679	0.037 ($< 10^{-5}$)
273	Extruding, Forming, Pressing, and Compacting Machine Setters, Operators, and Tenders	0.974	0.540	0.329 (0)
274	Plating and Coating Machine Setters, Operators, and Tenders, Metal and Plastic	0.974	0.487	0.382 (0)
275	Cooks, Institution and Cafeteria	0.974	0.776	0.290 (0)
276	Bartenders	0.974	0.932	0.118 ($< 10^{-253}$)
277	General and Operations Managers	0.974	1.024	-0.269 (0)
278	Dispatchers, Except Police, Fire, and Ambulance	0.974	0.991	0.225 (0)
279	Cutting, Punching, and Press Machine Setters, Operators, and Tenders, Metal and Plastic	0.974	0.712	0.355 (0)
280	Licensed Practical and Licensed Vocational Nurses	0.974	0.845	0.293 (0)
281	First-Line Supervisors of Food Preparation and Serving Workers	0.974	0.881	0.363 (0)
282	Tool and Die Makers	0.974	0.531	0.256 (0)
283	Landscape Architects	0.974	0.680	0.087 ($< 10^{-21}$)
284	Lodging Managers	0.975	0.593	0.397 (0)
285	Physician Assistants	0.975	0.816	0.031 ($< 10^{-12}$)
286	Architects, Except Landscape and Naval	0.975	1.034	-0.288 (0)
287	Counter and Rental Clerks	0.975	0.933	0.215 (0)
288	Baggage Porters and Bellhops	0.975	0.783	0.439 (0)
289	Food Cooking Machine Operators and Tenders	0.975	0.359	0.486 (0)
290	Computer, Automated Teller, and Office Machine Repairers	0.975	0.982	0.028 ($< 10^{-9}$)
291	Cashiers	0.975	0.854	0.491 (0)
292	Electrical and Electronics Engineering Technicians	0.975	0.958	-0.310 (0)
293	Zoologists and Wildlife Biologists	0.975	0.299	0.147 ($< 10^{-82}$)
294	Nonfarm Animal Caretakers	0.975	0.960	0.048 ($< 10^{-37}$)
295	First-Line Supervisors of Transportation and Material-Moving Machine and Vehicle Operators	0.975	0.949	0.319 (0)
296	Dental Hygienists	0.975	0.898	0.103 ($< 10^{-184}$)
297	Dental Assistants	0.975	0.920	0.017 ($< 10^{-5}$)
298	Ophthalmic Laboratory Technicians	0.975	0.632	0.152 ($< 10^{-95}$)
299	Retail Salespersons	0.975	0.904	0.456 (0)
300	Chiropractors	0.975	0.727	0.295 (0)
301	Coin, Vending, and Amusement Machine Servicers and Repairers	0.975	0.635	0.400 (0)
302	Upholsterers	0.975	0.459	0.345 (0)

303	Motorcycle Mechanics	0.975	0.429	0.490 (0)
304	Hazardous Materials Removal Workers	0.975	0.712	0.319 (0)
305	Grinding, Lapping, Polishing, and Buffing Machine Tool Setters, Operators, and Tenders, Metal and Plastic	0.975	0.624	0.335 (0)
306	Dentists, General	0.975	0.900	0.074 ($< 10^{-73}$)
307	Photographers	0.975	0.929	0.203 (0)
308	Jewelers and Precious Stone and Metal Workers	0.975	0.743	0.187 ($< 10^{-108}$)
309	Maintenance Workers, Machinery	0.975	0.685	0.412 (0)
310	Stationary Engineers and Boiler Operators	0.975	0.572	0.205 ($< 10^{-253}$)
311	Cabinetmakers and Bench Carpenters	0.976	0.726	0.329 (0)
312	Medical Equipment Preparers	0.976	0.766	0.182 ($< 10^{-291}$)
313	Coating, Painting, and Spraying Machine Setters, Operators, and Tenders	0.976	0.657	0.398 (0)
314	Butchers and Meat Cutters	0.976	0.844	0.307 (0)
315	Driver/Sales Workers	0.976	0.943	0.204 (0)
316	Aircraft Mechanics and Service Technicians	0.976	0.812	-0.034 ($< 10^{-9}$)
317	Glaziers	0.976	0.728	0.249 ($< 10^{-305}$)
318	Home Health Aides	0.976	0.937	0.018 ($< 10^{-4}$)
319	Power Plant Operators	0.976	0.549	0.364 (0)
320	Electromechanical Equipment Assemblers	0.976	0.625	0.103 ($< 10^{-41}$)
321	Broadcast Technicians	0.976	0.713	0.309 (0)
322	Prepress Technicians and Workers	0.976	0.742	0.199 ($< 10^{-280}$)
323	Tile and Marble Setters	0.976	0.646	0.378 (0)
324	Excavating and Loading Machine and Dragline Operators	0.976	0.423	0.335 (0)
325	Veterinarians	0.976	0.813	0.132 ($< 10^{-228}$)
326	Shipping, Receiving, and Traffic Clerks	0.976	1.026	0.181 (0)
327	Cooks, Fast Food	0.976	0.763	0.395 (0)
328	Logging Equipment Operators	0.976	0.104	0.150 ($< 10^{-65}$)
329	Cardiovascular Technologists and Technicians	0.976	0.807	0.350 (0)
330	Packaging and Filling Machine Operators and Tenders	0.976	0.805	0.273 (0)
331	Helpers–Brickmasons, Blockmasons, Stonemasons, and Tile and Marble Setters	0.976	0.510	0.374 (0)
332	Childcare Workers	0.976	0.945	0.003 ($< 10^{0}$)
333	Rolling Machine Setters, Operators, and Tenders, Metal and Plastic	0.976	0.422	0.446 (0)
334	Home Appliance Repairers	0.976	0.766	0.426 (0)
335	Helpers–Pipelayers, Plumbers, Pipefitters, and Steamfitters	0.976	0.640	0.247 (0)
336	Packers and Packagers, Hand	0.976	0.969	0.176 (0)
337	Farm Equipment Mechanics and Service Technicians	0.976	0.164	0.425 (0)
338	Emergency Management Directors	0.976	0.385	0.265 ($< 10^{-182}$)
339	Computer Numerically Controlled Machine Tool Programmers, Metal and Plastic	0.976	0.534	0.236 ($< 10^{-300}$)
340	Opticians, Dispensing	0.976	0.798	0.319 (0)
341	Laborers and Freight, Stock, and Material Movers, Hand	0.976	1.015	0.262 (0)
342	Crane and Tower Operators	0.976	0.509	0.323 (0)

343	Animal Control Workers	0.977	0.491	0.296 (0)
344	Grinding and Polishing Workers, Hand	0.977	0.499	0.392 (0)
345	Food Preparation Workers	0.977	0.914	0.219 (0)
346	Meter Readers, Utilities	0.977	0.605	0.320 (0)
347	Pipelayers	0.977	0.595	0.471 (0)
348	Audio and Video Equipment Technicians	0.977	0.974	0.072 ($< 10^{-36}$)
349	Police and Sheriff's Patrol Officers	0.977	0.918	0.137 (0)
350	First-Line Supervisors of Retail Sales Workers	0.977	0.868	0.479 (0)
351	Helpers–Installation, Maintenance, and Repair Workers	0.977	0.855	0.344 (0)
352	Occupational Health and Safety Specialists	0.977	0.786	0.077 ($< 10^{-71}$)
353	Medical and Clinical Laboratory Technicians	0.977	0.953	0.027 ($< 10^{-9}$)
354	Correctional Officers and Jailers	0.977	0.743	0.102 ($< 10^{-107}$)
355	Drywall and Ceiling Tile Installers	0.977	0.801	0.198 ($< 10^{-309}$)
356	Engineering Technicians, Except Drafters, All Other	0.977	0.724	-0.013 ($< 10^{0}$)
357	Cooks, Short Order	0.977	0.759	0.078 ($< 10^{-59}$)
358	Athletic Trainers	0.977	0.662	0.145 ($< 10^{-120}$)
359	Veterinary Technologists and Technicians	0.977	0.843	0.042 ($< 10^{-21}$)
360	Taxi Drivers and Chauffeurs	0.977	0.958	-0.014 ($< 10^{-2}$)
361	Detectives and Criminal Investigators	0.977	0.832	0.109 ($< 10^{-150}$)
362	Agricultural and Food Science Technicians	0.977	0.245	0.353 (0)
363	Physical Therapist Aides	0.977	0.702	0.294 (0)
364	Life, Physical, and Social Science Technicians, All Other	0.977	0.780	-0.109 ($< 10^{-92}$)
365	Veterinary Assistants and Laboratory Animal Caretakers	0.977	0.764	0.103 ($< 10^{-108}$)
366	First-Line Supervisors of Production and Operating Workers	0.977	0.856	0.350 (0)
367	Structural Metal Fabricators and Fitters	0.977	0.660	0.378 (0)
368	Drilling and Boring Machine Tool Setters, Operators, and Tenders, Metal and Plastic	0.977	0.482	0.576 (0)
369	Construction and Building Inspectors	0.977	0.875	0.081 ($< 10^{-88}$)
370	Nuclear Medicine Technologists	0.977	0.685	0.332 (0)
371	Electrical Power-Line Installers and Repairers	0.977	0.723	0.247 (0)
372	Dental Laboratory Technicians	0.977	0.722	0.298 (0)
373	Medical Equipment Repairers	0.977	0.794	0.207 ($< 10^{-322}$)
374	Physical Therapist Assistants	0.977	0.742	0.361 (0)
375	Locksmiths and Safe Repairers	0.977	0.669	0.406 (0)
376	Respiratory Therapists	0.977	0.865	0.282 (0)
377	Security and Fire Alarm Systems Installers	0.977	0.885	0.178 ($< 10^{-230}$)
378	Machine Feeders and Offbearers	0.977	0.628	0.281 (0)
379	Bakers	0.977	0.924	0.168 (0)
380	Parking Lot Attendants	0.977	1.174	0.063 ($< 10^{-29}$)
381	Recreation Workers	0.978	0.940	0.003 ($< 10^{0}$)
382	Landscaping and Groundskeeping Workers	0.978	0.972	0.163 (0)
383	Career/Technical Education Teachers, Secondary School	0.978	0.657	0.367 (0)

384	Surveyors	0.978	0.690	0.304 (0)
385	Office Machine Operators, Except Computer	0.978	0.944	-0.002 (< 10^{0})
386	Commercial Pilots	0.978	0.586	0.249 (0)
387	Engine and Other Machine Assemblers	0.978	0.322	0.109 (< 10^{-30})
388	Bus Drivers, Transit and Intercity	0.978	0.938	0.034 (< 10^{-7})
389	Industrial Truck and Tractor Operators	0.978	0.917	0.345 (0)
390	Pesticide Handlers, Sprayers, and Applicators, Vegetation	0.978	0.293	0.523 (0)
391	Roofers	0.978	0.795	0.382 (0)
392	Motorboat Mechanics and Service Technicians	0.978	0.380	0.490 (0)
393	Light Truck or Delivery Services Drivers	0.978	0.964	0.174 (0)
394	Operating Engineers and Other Construction Equipment Operators	0.978	0.853	0.221 (0)
395	Mixing and Blending Machine Setters, Operators, and Tenders	0.978	0.747	0.317 (0)
396	Carpet Installers	0.978	0.644	0.244 (< 10^{-162})
397	Industrial Machinery Mechanics	0.978	0.773	0.382 (0)
398	Brickmasons and Blockmasons	0.978	0.775	0.266 (0)
399	Surveying and Mapping Technicians	0.978	0.688	0.262 (0)
400	Agricultural Inspectors	0.978	0.137	0.520 (0)
401	First-Line Supervisors of Housekeeping and Janitorial Workers	0.978	0.938	0.088 (< 10^{-140})
402	Janitors and Cleaners, Except Maids and Housekeeping Cleaners	0.978	0.940	0.114 (< 10^{-247})
403	Welders, Cutters, Solderers, and Brazers	0.978	0.812	0.247 (0)
404	First-Line Supervisors of Police and Detectives	0.978	0.791	0.114 (< 10^{-201})
405	Animal Trainers	0.978	0.437	0.420 (0)
406	Radiation Therapists	0.978	0.610	0.372 (0)
407	First-Line Supervisors of Landscaping, Lawn Service, and Groundskeeping Workers	0.978	0.902	0.226 (0)
408	Computer-Controlled Machine Tool Operators, Metal and Plastic	0.978	0.649	0.281 (0)
409	Fire Inspectors and Investigators	0.979	0.468	0.207 (< 10^{-145})
410	Surgical Technologists	0.979	0.796	0.350 (0)
411	Sheet Metal Workers	0.979	0.870	0.167 (0)
412	Camera Operators, Television, Video, and Motion Picture	0.979	0.860	0.126 (< 10^{-60})
413	Electrical and Electronics Repairers, Powerhouse, Substation, and Relay	0.979	0.448	0.204 (< 10^{-125})
414	Environmental Science and Protection Technicians, Including Health	0.979	0.717	0.206 (< 10^{-315})
415	Pest Control Workers	0.979	0.734	0.435 (0)
416	File Clerks	0.979	0.965	0.087 (< 10^{-114})
417	Combined Food Preparation and Serving Workers, Including Fast Food	0.979	0.909	0.281 (0)
418	Conservation Scientists	0.979	0.255	0.092 (< 10^{-37})
419	Millwrights	0.979	0.466	0.295 (0)
420	Dietetic Technicians	0.979	0.574	0.311 (0)
421	Structural Iron and Steel Workers	0.979	0.713	0.339 (0)
422	Mechanical Engineering Technicians	0.979	0.751	0.037 (< 10^{-9})
423	Molders, Shapers, and Casters, Except Metal and Plastic	0.979	0.558	0.386 (0)
424	Conveyor Operators and Tenders	0.979	0.446	0.476 (0)

425	Transportation Inspectors	0.979	0.634	0.130 ($< 10^{-55}$)
426	Forensic Science Technicians	0.979	0.498	0.156 ($< 10^{-83}$)
427	Diagnostic Medical Sonographers	0.979	0.817	0.198 (0)
428	Cleaners of Vehicles and Equipment	0.979	0.930	0.313 (0)
429	Cement Masons and Concrete Finishers	0.979	0.847	0.277 (0)
430	Machinists	0.979	0.877	0.194 (0)
431	Dishwashers	0.979	0.945	0.133 (0)
432	Multiple Machine Tool Setters, Operators, and Tenders, Metal and Plastic	0.979	0.460	0.317 (0)
433	Chefs and Head Cooks	0.979	0.931	0.072 ($< 10^{-65}$)
434	Outdoor Power Equipment and Other Small Engine Mechanics	0.979	0.532	0.532 (0)
435	Insulation Workers, Mechanical	0.979	0.414	0.326 (0)
436	Heavy and Tractor-Trailer Truck Drivers	0.979	0.901	0.380 (0)
437	Electric Motor, Power Tool, and Related Repairers	0.979	0.479	0.443 (0)
438	Food Service Managers	0.980	0.906	0.048 ($< 10^{-41}$)
439	Automotive and Watercraft Service Attendants	0.980	0.762	0.155 ($< 10^{-280}$)
440	Refuse and Recyclable Material Collectors	0.980	0.763	0.329 (0)
441	Automotive Service Technicians and Mechanics	0.980	0.888	0.324 (0)
442	Electrical and Electronics Repairers, Commercial and Industrial Equipment	0.980	0.715	-0.076 ($< 10^{-57}$)
443	Occupational Health and Safety Technicians	0.980	0.420	0.436 (0)
444	Bus and Truck Mechanics and Diesel Engine Specialists	0.980	0.895	0.286 (0)
445	Paving, Surfacing, and Tamping Equipment Operators	0.980	0.635	0.376 (0)
446	Water and Wastewater Treatment Plant and System Operators	0.980	0.721	0.312 (0)
447	First-Line Supervisors of Construction Trades and Extraction Workers	0.980	0.955	0.155 (0)
448	Telecommunications Line Installers and Repairers	0.980	0.854	0.256 (0)
449	Laundry and Dry-Cleaning Workers	0.980	0.921	0.239 (0)
450	Chemical Plant and System Operators	0.980	0.099	0.412 (0)
451	Environmental Engineering Technicians	0.981	0.589	0.287 (0)
452	Crushing, Grinding, and Polishing Machine Setters, Operators, and Tenders	0.981	0.435	0.425 (0)
453	Mobile Heavy Equipment Mechanics, Except Engines	0.981	0.780	0.203 (0)
454	Cleaning, Washing, and Metal Pickling Equipment Operators and Tenders	0.981	0.349	0.255 ($< 10^{-234}$)
455	Heating, Air Conditioning, and Refrigeration Mechanics and Installers	0.981	0.916	0.208 (0)
456	Separating, Filtering, Clarifying, Precipitating, and Still Machine Setters, Operators, and Tenders	0.981	0.366	0.312 (0)
457	Plumbers, Pipefitters, and Steamfitters	0.981	0.960	0.037 ($< 10^{-24}$)
458	Heat Treating Equipment Setters, Operators, and Tenders, Metal and Plastic	0.981	0.415	0.366 (0)
459	Helpers–Carpenters	0.981	0.576	0.297 (0)
460	Tire Repairers and Changers	0.981	0.698	0.451 (0)
461	Electricians	0.981	0.964	0.062 ($< 10^{-70}$)
462	Automotive Body and Related Repairers	0.981	0.890	0.293 (0)
463	Highway Maintenance Workers	0.981	0.587	0.212 (0)
464	Electronic Home Entertainment Equipment Installers and Repairers	0.981	0.571	0.401 (0)

465	Meat, Poultry, and Fish Cutters and Trimmers	0.981	0.528	0.352 (0)
466	Cooks, Restaurant	0.982	0.964	0.197 (0)
467	Firefighters	0.982	0.851	0.207 (0)
468	Control and Valve Installers and Repairers, Except Mechanical Door	0.982	0.584	0.415 (0)
469	Painters, Construction and Maintenance	0.982	0.969	0.054 ($< 10^{-49}$)
470	Maintenance and Repair Workers, General	0.982	0.908	0.364 (0)
471	First-Line Supervisors of Mechanics, Installers, and Repairers	0.982	0.911	0.181 (0)
472	Telecommunications Equipment Installers and Repairers, Except Line Installers	0.982	0.963	0.062 ($< 10^{-55}$)
473	Septic Tank Servicers and Sewer Pipe Cleaners	0.982	0.522	0.475 (0)
474	First-Line Supervisors of Farming, Fishing, and Forestry Workers	0.982	0.035	0.321 (0)
475	First-Line Supervisors of Fire Fighting and Prevention Workers	0.982	0.734	0.205 (0)
476	Carpenters	0.983	0.957	0.115 ($< 10^{-250}$)
477	Emergency Medical Technicians and Paramedics	0.983	0.807	0.217 (0)
478	Construction Laborers	0.983	0.943	0.230 (0)
479	Extruding and Drawing Machine Setters, Operators, and Tenders, Metal and Plastic	0.984	0.463	0.429 (0)
480	Insulation Workers, Floor, Ceiling, and Wall	0.985	0.499	0.289 ($< 10^{-218}$)
481	Forest and Conservation Technicians	0.985	0.058	0.376 (0)

B.6.3 Job Groups

The O*NET skills database allows us to identify how important each of 230 workplace skills is to completing each of the BLS jobs. We use K-means clustering to group jobs into five groups according to the skills required to perform those jobs. The complete list of BLS jobs comprising each job group is presented in the table below. Our interpretation about the scaling behaviors of jobs, and how aggregate skills indicate those scaling behaviors, is the same if we use anywhere between three and seven job groups instead of five while computing the K-means clustering algorithm.

Group (β)	BLS Jobs

Purple (1.39)	Aerospace Engineers, Agricultural Engineers, Animal Scientists, Anthropologists and Archeologists, Architects, Except Landscape and Naval, Architectural and Civil Drafters, Architectural and Engineering Managers, Astronomers, Atmospheric and Space Scientists, Biochemists and Biophysicists, Biological Scientists, All Other, Biomedical Engineers, Cartographers and Photogrammetrists, Chemical Engineers, Chemists, Civil Engineering Technicians, Civil Engineers, Commercial and Industrial Designers, Computer Hardware Engineers, Computer Programmers, Computer Systems Analysts, Computer and Information Research Scientists, Computer and Information Systems Managers, Construction Managers, Database Administrators, Electrical Engineers, Electrical and Electronics Drafters, Electronics Engineers, Except Computer, Engineers, All Other, Environmental Engineers, Environmental Scientists and Specialists, Including Health, Food Scientists and Technologists, Geographers, Geoscientists, Except Hydrologists and Geographers, Health and Safety Engineers, Except Mining Safety Engineers and Inspectors, Hydrologists, Industrial Engineering Technicians, Industrial Engineers, Landscape Architects, Logisticians, Marine Engineers and Naval Architects, Materials Engineers, Materials Scientists, Mathematical Technicians, Mathematicians, Mechanical Drafters, Mechanical Engineers, Medical Scientists, Except Epidemiologists, Microbiologists, Mining and Geological Engineers, Including Mining Safety Engineers, Multimedia Artists and Animators, Natural Sciences Managers, Network and Computer Systems Administrators, Nuclear Engineers, Occupational Health and Safety Specialists, Operations Research Analysts, Petroleum Engineers, Physical Scientists, All Other, Physicists, Sales Engineers, Set and Exhibit Designers, Social Scientists and Related Workers, All Other, Software Developers, Applications, Software Developers, Systems Software, Soil and Plant Scientists, Statistical Assistants, Statisticians, Technical Writers

Green (1.08)	Accountants and Auditors, Actuaries, Administrative Law Judges, Adjudicators, and Hearing Officers, Administrative Services Managers, Adult Basic and Secondary Education and Literacy Teachers and Instructors, Advertising Sales Agents, Advertising and Promotions Managers, Agents and Business Managers of Artists, Performers, and Athletes, Air Traffic Controllers, Appraisers and Assessors of Real Estate, Arbitrators, Mediators, and Conciliators, Archivists, Art Directors, Audiologists, Broadcast News Analysts, Budget Analysts, Business Operations Specialists, All Other, Buyers and Purchasing Agents, Farm Products, Career/Technical Education Teachers, Middle School, Cargo and Freight Agents, Chief Executives, Child, Family, and School Social Workers, Choreographers, Claims Adjusters, Examiners, and Investigators, Clergy, Clinical, Counseling, and School Psychologists, Coaches and Scouts, Compensation and Benefits Managers, Compensation, Benefits, and Job Analysis Specialists, Compliance Officers, Concierges, Cost Estimators, Credit Analysts, Credit Authorizers, Checkers, and Clerks, Credit Counselors, Curators, Customer Service Representatives, Dietitians and Nutritionists, Directors, Religious Activities and Education, Dispatchers, Except Police, Fire, and Ambulance, Economists, Editors, Education Administrators, Elementary and Secondary School, Education Administrators, Postsecondary, Education Administrators, Preschool and Childcare Center/Program, Educational, Guidance, School, and Vocational Counselors, Elementary School Teachers, Except Special Education, Eligibility Interviewers, Government Programs, Emergency Management Directors, Epidemiologists, Executive Secretaries and Executive Administrative Assistants, Farm and Home Management Advisors, Fashion Designers, Film and Video Editors, Financial Analysts, Financial Examiners, Financial Managers, Financial Specialists, All Other, First-Line Supervisors of Non-Retail Sales Workers, First-Line Supervisors of Office and Administrative Support Workers, First-Line Supervisors of Personal Service Workers, First-Line Supervisors of Transportation and Material-Moving Machine and Vehicle Operators, Food Service Managers, Gaming Managers, Gaming Supervisors, General and Operations Managers, Graphic Designers, Health Diagnosing and Treating Practitioners, All Other, Health Educators, Healthcare Social Workers, Historians, Human Resources Assistants, Except Payroll and Timekeeping, Human Resources Managers, Industrial-Organizational Psychologists, Instructional Coordinators, Insurance Sales Agents, Insurance Underwriters, Interior Designers, Interpreters and Translators, Judges, Magistrate Judges, and Magistrates, Judicial Law Clerks, Kindergarten Teachers, Except Special Education, Lawyers, Librarians, Loan Interviewers and Clerks, Loan Officers, Lodging Managers, Management Analysts, Managers, All Other, Market Research Analysts and Marketing Specialists, Marketing Managers, Marriage and Family Therapists, Medical Assistants, Medical and Health Services Managers, Meeting, Convention, and Event Planners, Mental Health Counselors, Mental Health and Substance Abuse Social Workers, Middle School Teachers, Except Special and Career/Technical Education, Music Directors and Composers, New Accounts Clerks, Occupational Therapists, Opticians, Dispensing, Paralegals and Legal Assistants, Personal Financial Advisors, Pharmacists, Police, Fire, and Ambulance Dispatchers, Political Scientists, Postmasters and Mail Superintendents, Preschool Teachers, Except Special Education, Private Detectives and Investigators, Probation Officers and Correctional Treatment Specialists, Procurement Clerks, Producers and Directors, Production, Planning, and Expediting Clerks, Property, Real Estate, and Community Association Managers, Psychiatric Technicians, Psychologists, All Other, Public Relations Specialists, Public Relations and Fundraising Managers, Purchasing Agents, Except Wholesale, Retail, and Farm Products, Purchasing Managers, Radio and Television Announcers, Real Estate Brokers, Real Estate Sales Agents, Recreation Workers, Recreational Therapists, Rehabilitation Counselors, Reporters and Correspondents, Residential Advisors, Sales Managers, Sales Representatives, Wholesale and Manufacturing, Except Technical and Scientific Products, Sales Representatives, Wholesale and Manufacturing, Technical and Scientific Products, Secondary School Teachers, Except Special and Career/Technical Education, Securities, Commodities, and Financial Services Sales Agents, Social Science Research Assistants, Social and Community Service Managers, Social and Human Service Assistants, Sociologists, Special Education Teachers, Middle School, Special Education Teachers, Secondary School, Speech-Language Pathologists, Substance Abuse and Behavioral Disorder Counselors, Survey Researchers, Tax Examiners and Collectors, and Revenue Agents, Tax Preparers, Training and Development Managers, Training and Development Specialists, Transportation, Storage, and Distribution Managers, Travel Agents, Travel Guides, Urban and Regional Planners, Wholesale and Retail Buyers, Except Farm Products, Writers and Authors

Yellow (1.02)	Aerospace Engineering and Operations Technicians, Agricultural Inspectors, Agricultural and Food Science Technicians, Aircraft Cargo Handling Supervisors, Aircraft Mechanics and Service Technicians, Airfield Operations Specialists, Airline Pilots, Copilots, and Flight Engineers, Animal Control Workers, Animal Trainers, Athletic Trainers, Audio and Video Equipment Technicians, Audio-Visual and Multimedia Collections Specialists, Avionics Technicians, Biological Technicians, Broadcast Technicians, Captains, Mates, and Pilots of Water Vessels, Cardiovascular Technologists and Technicians, Career/Technical Education Teachers, Secondary School, Chefs and Head Cooks, Chemical Technicians, Chiropractors, Commercial Divers, Commercial Pilots, Computer Numerically Controlled Machine Tool Programmers, Metal and Plastic, Computer Operators, Computer, Automated Teller, and Office Machine Repairers, Conservation Scientists, Construction and Building Inspectors, Correctional Officers and Jailers, Dental Assistants, Dental Hygienists, Dental Laboratory Technicians, Dentists, General, Desktop Publishers, Detectives and Criminal Investigators, Diagnostic Medical Sonographers, Dietetic Technicians, Electrical Power-Line Installers and Repairers, Electrical and Electronics Engineering Technicians, Electrical and Electronics Repairers, Commercial and Industrial Equipment, Electrical and Electronics Repairers, Powerhouse, Substation, and Relay, Electricians, Electro-Mechanical Technicians, Electronic Equipment Installers and Repairers, Motor Vehicles, Electronic Home Entertainment Equipment Installers and Repairers, Elevator Installers and Repairers, Embalmers, Emergency Medical Technicians and Paramedics, Engineering Technicians, Except Drafters, All Other, Environmental Engineering Technicians, Environmental Science and Protection Technicians, Including Health, Explosives Workers, Ordnance Handling Experts, and Blasters, Fabric and Apparel Patternmakers, Farmers, Ranchers, and Other Agricultural Managers, Fire Inspectors and Investigators, Firefighters, First-Line Supervisors of Construction Trades and Extraction Workers, First-Line Supervisors of Correctional Officers, First-Line Supervisors of Farming, Fishing, and Forestry Workers, First-Line Supervisors of Fire Fighting and Prevention Workers, First-Line Supervisors of Landscaping, Lawn Service, and Groundskeeping Workers, First-Line Supervisors of Mechanics, Installers, and Repairers, First-Line Supervisors of Police and Detectives, First-Line Supervisors of Production and Operating Workers, Fish and Game Wardens, Forensic Science Technicians, Forest Fire Inspectors and Prevention Specialists, Forest and Conservation Technicians, Forest and Conservation Workers, Foresters, Gaming Surveillance Officers and Gaming Investigators, Geological and Petroleum Technicians, Hazardous Materials Removal Workers, Heating, Air Conditioning, and Refrigeration Mechanics and Installers, Industrial Production Managers, Licensed Practical and Licensed Vocational Nurses, Life, Physical, and Social Science Technicians, All Other, Manufactured Building and Mobile Home Installers, Mechanical Engineering Technicians, Medical Appliance Technicians, Medical Equipment Preparers, Medical Equipment Repairers, Medical and Clinical Laboratory Technicians, Medical and Clinical Laboratory Technologists, Museum Technicians and Conservators, Nuclear Medicine Technologists, Nuclear Power Reactor Operators, Nuclear Technicians, Occupational Health and Safety Technicians, Occupational Therapy Assistants, Optometrists, Oral and Maxillofacial Surgeons, Orthodontists, Orthotists and Prosthetists, Pest Control Workers, Photographers, Physical Therapist Assistants, Physical Therapists, Physician Assistants, Podiatrists, Police and Sheriff's Patrol Officers, Power Distributors and Dispatchers, Radiation Therapists, Radio, Cellular, and Tower Equipment Installers and Repairers, Railroad Conductors and Yardmasters, Respiratory Therapists, Respiratory Therapy Technicians, Service Unit Operators, Oil, Gas, and Mining, Ship Engineers, Sound Engineering Technicians, Stationary Engineers and Boiler Operators, Surgical Technologists, Surveying and Mapping Technicians, Surveyors, Telecommunications Equipment Installers and Repairers, Except Line Installers, Traffic Technicians, Transit and Railroad Police, Transportation Attendants, Except Flight Attendants, Transportation Inspectors, Veterinarians, Veterinary Assistants and Laboratory Animal Caretakers, Veterinary Technologists and Technicians, Water and Wastewater Treatment Plant and System Operators, Zoologists and Wildlife Biologists

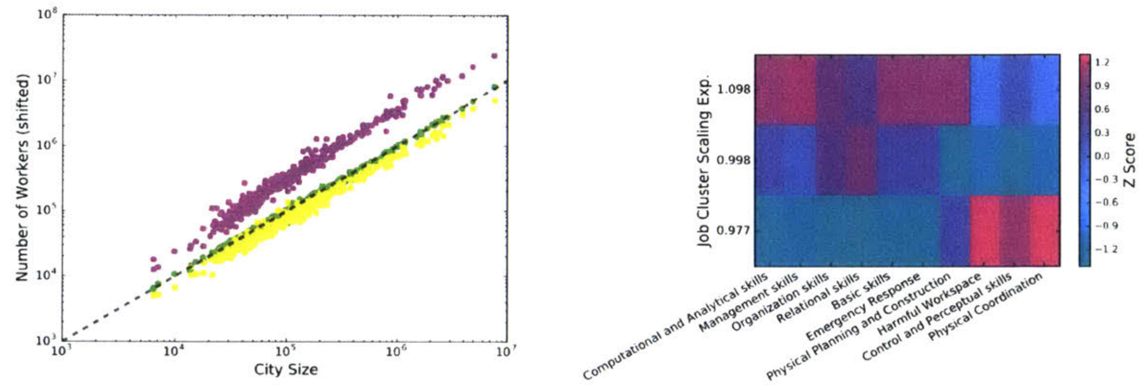
Red (0.98)	Adhesive Bonding Machine Operators and Tenders, Aircraft Structure, Surfaces, Rigging, and Systems Assemblers, Ambulance Drivers and Attendants, Except Emergency Medical Technicians, Animal Breeders, Automotive Body and Related Repairers, Automotive Glass Installers and Repairers, Automotive Service Technicians and Mechanics, Automotive and Watercraft Service Attendants, Bakers, Bicycle Repairers, Boilermakers, Brickmasons and Blockmasons, Bridge and Lock Tenders, Bus Drivers, School or Special Client, Bus Drivers, Transit and Intercity, Bus and Truck Mechanics and Diesel Engine Specialists, Butchers and Meat Cutters, Cabinetmakers and Bench Carpenters, Camera Operators, Television, Video, and Motion Picture, Camera and Photographic Equipment Repairers, Carpenters, Carpet Installers, Cement Masons and Concrete Finishers, Chemical Equipment Operators and Tenders, Chemical Plant and System Operators, Cleaners of Vehicles and Equipment, Cleaning, Washing, and Metal Pickling Equipment Operators and Tenders, Coating, Painting, and Spraying Machine Setters, Operators, and Tenders, Coil Winders, Tapers, and Finishers, Coin, Vending, and Amusement Machine Servicers and Repairers, Computer-Controlled Machine Tool Operators, Metal and Plastic, Construction Laborers, Continuous Mining Machine Operators, Control and Valve Installers and Repairers, Except Mechanical Door, Conveyor Operators and Tenders, Cooks, Restaurant, Cooling and Freezing Equipment Operators and Tenders, Couriers and Messengers, Craft Artists, Crane and Tower Operators, Crushing, Grinding, and Polishing Machine Setters, Operators, and Tenders, Cutters and Trimmers, Hand, Cutting and Slicing Machine Setters, Operators, and Tenders, Cutting, Punching, and Press Machine Setters, Operators, and Tenders, Metal and Plastic, Derrick Operators, Oil and Gas, Dishwashers, Drilling and Boring Machine Tool Setters, Operators, and Tenders, Metal and Plastic, Drywall and Ceiling Tile Installers, Earth Drillers, Except Oil and Gas, Electric Motor, Power Tool, and Related Repairers, Electrical and Electronic Equipment Assemblers, Electrical and Electronics Installers and Repairers, Transportation Equipment, Electromechanical Equipment Assemblers, Engine and Other Machine Assemblers, Etchers and Engravers, Excavating and Loading Machine and Dragline Operators, Extruding and Drawing Machine Setters, Operators, and Tenders, Metal and Plastic, Extruding and Forming Machine Setters, Operators, and Tenders, Synthetic and Glass Fibers, Extruding, Forming, Pressing, and Compacting Machine Setters, Operators, and Tenders, Fabric Menders, Except Garment, Fallers, Farm Equipment Mechanics and Service Technicians, Fence Erectors, Fiberglass Laminators and Fabricators, Fishers and Related Fishing Workers, Floor Layers, Except Carpet, Wood, and Hard Tiles, Floor Sanders and Finishers, Food Batchmakers, Food Cooking Machine Operators and Tenders, Food and Tobacco Roasting, Baking, and Drying Machine Operators and Tenders, Forging Machine Setters, Operators, and Tenders, Metal and Plastic, Foundry Mold and Coremakers, Furnace, Kiln, Oven, Drier, and Kettle Operators and Tenders, Furniture Finishers, Gas Compressor and Gas Pumping Station Operators, Gas Plant Operators, Glaziers, Grinding and Polishing Workers, Hand, Grinding, Lapping, Polishing, and Buffing Machine Tool Setters, Operators, and Tenders, Metal and Plastic, Heat Treating Equipment Setters, Operators, and Tenders, Metal and Plastic, Heavy and Tractor-Trailer Truck Drivers, Helpers–Brickmasons, Blockmasons, Stonemasons, and Tile and Marble Setters, Helpers–Carpenters, Helpers–Electricians, Helpers–Extraction Workers, Helpers–Installation, Maintenance, and Repair Workers, Helpers–Painters, Paperhangers, Plasterers, and Stucco Masons, Helpers–Pipelayers, Plumbers, Pipefitters, and Steamfitters, Helpers–Production Workers, Helpers–Roofers, Highway Maintenance Workers, Hoist and Winch Operators, Home Appliance Repairers, Industrial Machinery Mechanics, Industrial Truck and Tractor Operators, Insulation Workers, Floor, Ceiling, and Wall, Insulation Workers, Mechanical, Janitors and Cleaners, Except Maids and Housekeeping Cleaners, Jewelers and Precious Stone and Metal Workers, Laborers and Freight, Stock, and Material Movers, Hand, Landscaping and Groundskeeping Workers, Lathe and Turning Machine Tool Setters, Operators, and Tenders, Metal and Plastic, Laundry and Dry-Cleaning Workers, Layout Workers, Metal and Plastic, Light Truck or Delivery Services Drivers, Locksmiths and Safe Repairers, Locomotive Engineers, Log Graders and Scalers, Logging Equipment Operators, Machine Feeders and Offbearers, Machinists, Maintenance Workers, Machinery, Maintenance and Repair Workers, General, Meat, Poultry, and Fish Cutters and Trimmers, Mechanical Door Repairers, Metal-Refining Furnace Operators and Tenders, Meter Readers, Utilities, Milling and Planing Machine Setters, Operators, and Tenders, Metal and Plastic, Millwrights, Mine Cutting and Channeling Machine Operators, Mine Shuttle Car Operators, Mixing and Blending Machine Setters, Operators, and Tenders, Mobile Heavy Equipment Mechanics, Except Engines, Model Makers, Metal and Plastic, Model Makers, Wood, Molders, Shapers, and Casters, Except Metal and Plastic, Molding, Coremaking, and Casting Machine Setters, Operators, and Tenders, Metal and Plastic, Motion Picture Projectionists, Motorboat Mechanics and Service Technicians, Motorboat Operators, Motorcycle Mechanics, Multiple Machine Tool Setters, Operators, and Tenders, Metal and Plastic, Musical Instrument Repairers and Tuners, Office Machine Operators, Except Computer, Operating Engineers and Other Construction Equipment Operators, Ophthalmic Laboratory

Blue (0.94)	Actors, Amusement and Recreation Attendants, Athletes and Sports Competitors, Baggage Porters and Bellhops, Bailiffs, Barbers, Bartenders, Bill and Account Collectors, Billing and Posting Clerks, Bookkeeping, Accounting, and Auditing Clerks, Brokerage Clerks, Cashiers, Childcare Workers, Combined Food Preparation and Serving Workers, Including Fast Food, Cooks, Fast Food, Cooks, Institution and Cafeteria, Cooks, Short Order, Correspondence Clerks, Costume Attendants, Counter Attendants, Cafeteria, Food Concession, and Coffee Shop, Counter and Rental Clerks, Court Reporters, Court, Municipal, and License Clerks, Crossing Guards, Dancers, Data Entry Keyers, Demonstrators and Product Promoters, Dining Room and Cafeteria Attendants and Bartender Helpers, Door-to-Door Sales Workers, News and Street Vendors, and Related Workers, Driver/Sales Workers, Farm Labor Contractors, File Clerks, Fine Artists, Including Painters, Sculptors, and Illustrators, First-Line Supervisors of Food Preparation and Serving Workers, First-Line Supervisors of Housekeeping and Janitorial Workers, First-Line Supervisors of Retail Sales Workers, Fitness Trainers and Aerobics Instructors, Flight Attendants, Floral Designers, Food Preparation Workers, Food Servers, Nonrestaurant, Funeral Attendants, Gaming Cage Workers, Gaming Change Persons and Booth Cashiers, Gaming Dealers, Gaming and Sports Book Writers and Runners, Graders and Sorters, Agricultural Products, Hairdressers, Hairstylists, and Cosmetologists, Home Health Aides, Hosts and Hostesses, Restaurant, Lounge, and Coffee Shop, Hotel, Motel, and Resort Desk Clerks, Inspectors, Testers, Sorters, Samplers, and Weighers, Insurance Appraisers, Auto Damage, Insurance Claims and Policy Processing Clerks, Interviewers, Except Eligibility and Loan, Legal Secretaries, Library Assistants, Clerical, Library Technicians, Lifeguards, Ski Patrol, and Other Recreational Protective Service Workers, Locker Room, Coatroom, and Dressing Room Attendants, Maids and Housekeeping Cleaners, Mail Clerks and Mail Machine Operators, Except Postal Service, Makeup Artists, Theatrical and Performance, Manicurists and Pedicurists, Massage Therapists, Medical Records and Health Information Technicians, Medical Secretaries, Medical Transcriptionists, Merchandise Displayers and Window Trimmers, Models, Musicians and Singers, Nonfarm Animal Caretakers, Occupational Therapy Aides, Office Clerks, General, Order Clerks, Parking Enforcement Workers, Parts Salespersons, Payroll and Timekeeping Clerks, Personal Care Aides, Pharmacy Aides, Pharmacy Technicians, Photographic Process Workers and Processing Machine Operators, Physical Therapist Aides, Postal Service Clerks, Postal Service Mail Carriers, Postal Service Mail Sorters, Processors, and Processing Machine Operators, Prepress Technicians and Workers, Proofreaders and Copy Markers, Psychiatric Aides, Public Address System and Other Announcers, Radio Operators, Receptionists and Information Clerks, Reservation and Transportation Ticket Agents and Travel Clerks, Retail Salespersons, Secretaries and Administrative Assistants, Except Legal, Medical, and Executive, Security Guards, Self-Enrichment Education Teachers, Sewers, Hand, Shampooers, Shipping, Receiving, and Traffic Clerks, Skincare Specialists, Slot Supervisors, Stock Clerks and Order Fillers, Switchboard Operators, Including Answering Service, Teacher Assistants, Telemarketers, Telephone Operators, Tellers, Title Examiners, Abstractors, and Searchers, Tour Guides and Escorts, Umpires, Referees, and Other Sports Officials, Ushers, Lobby Attendants, and Ticket Takers, Waiters and Waitresses, Weighers, Measurers, Checkers, and Samplers, Recordkeeping, Word Processors and Typists

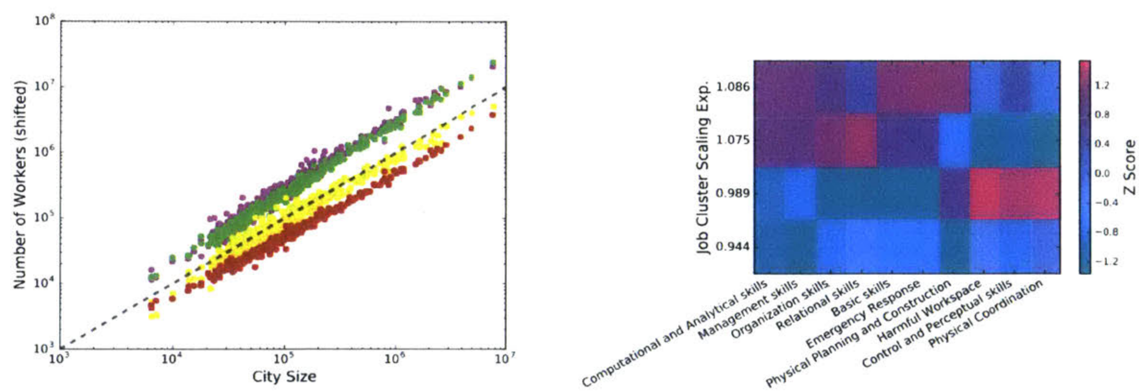
Alternative Job Groups using K-means

We demonstrate that our choice to focus on five groups of jobs according to skills produces results that are consistent for several alternative numbers of groups. Using K-means to identify between three and seven job groups continues demonstrate that computational/analytical and managerial skill are more indicative of super linear job growth, while physical skills are more indicative of linear or sub linear job growth. Likewise, our conclusions relating job scaling to expected job impact by comparing skills hold as well.

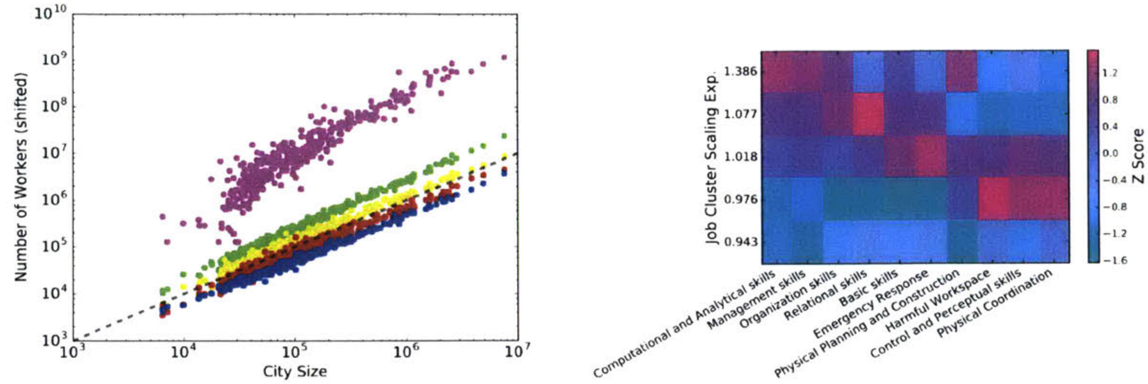
K-means clustering of similar jobs ($k = 3$)



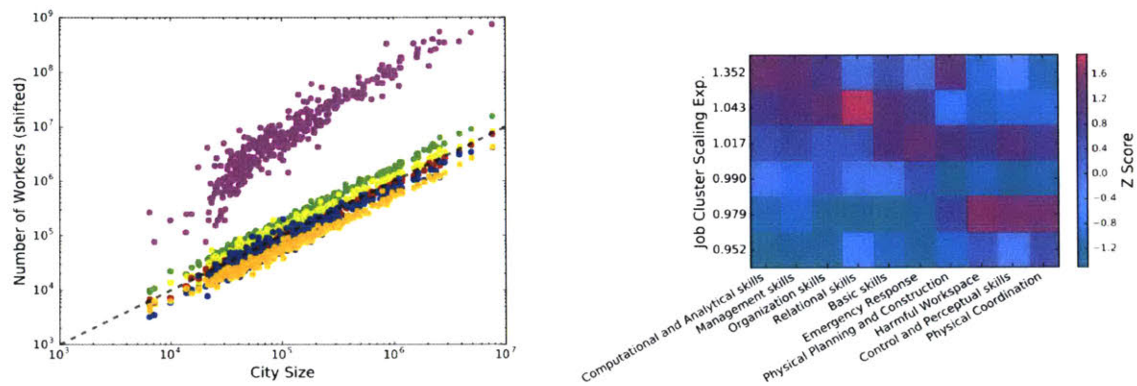
K-means clustering of similar jobs ($k = 4$)



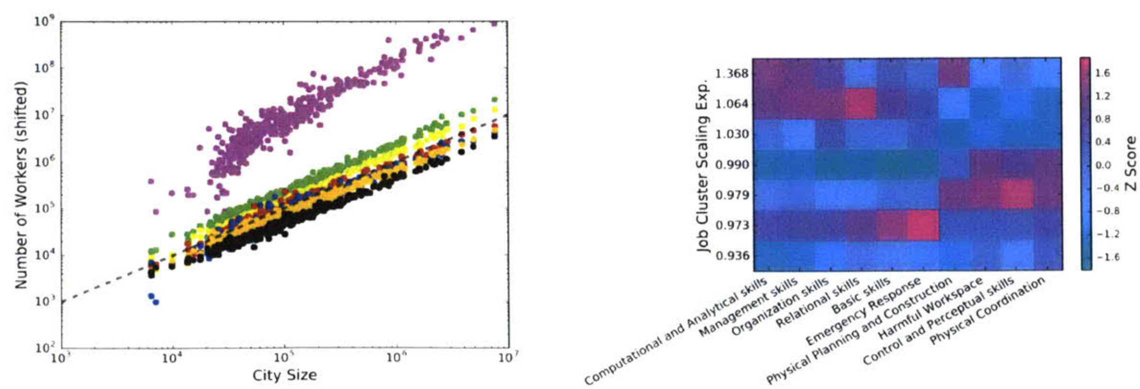
K-means clustering of similar jobs ($k = 5$)



K-means clustering of similar jobs ($k = 6$)



K-means clustering of similar jobs ($k = 7$)



Stability Testing for Job Groups

We want to test the stability of the scaling results we observe when using five job clusters obtained from k-means clustering. In particular, how robust to sub-sampling is our observation that one job clusters scales faster than the rest? For a single trial, we sub-sample from the complete list of BLS occupations (percent indicated in plot titles) to obtain a matrix where each row represents a single occupation which was sub-sampled and each column represents the raw O*NET importance of a skill to each occupation. We apply k-means to this occupation-skill matrix (i.e. occupations are instances and skills are features) to obtain five occupation clusters (note: examination of prescribing between three and seven clusters is discussed in the SM). We then measure the scaling exponent (β) of each occupation cluster and rank the occupation clusters according to scaling exponent (rank indicated by color in plots). We perform 100 independent trials for each sub-sampling proportion in $\{10\%, 20\%, \ldots, 90\%, 100\%\}$ and plot the resulting scaling exponent distributions. In good agreement with our original findings, we find that indeed one occupation cluster (indicated in purple) tends to grow much faster than the other occupation clusters despite varying sub-sampling of occupations.

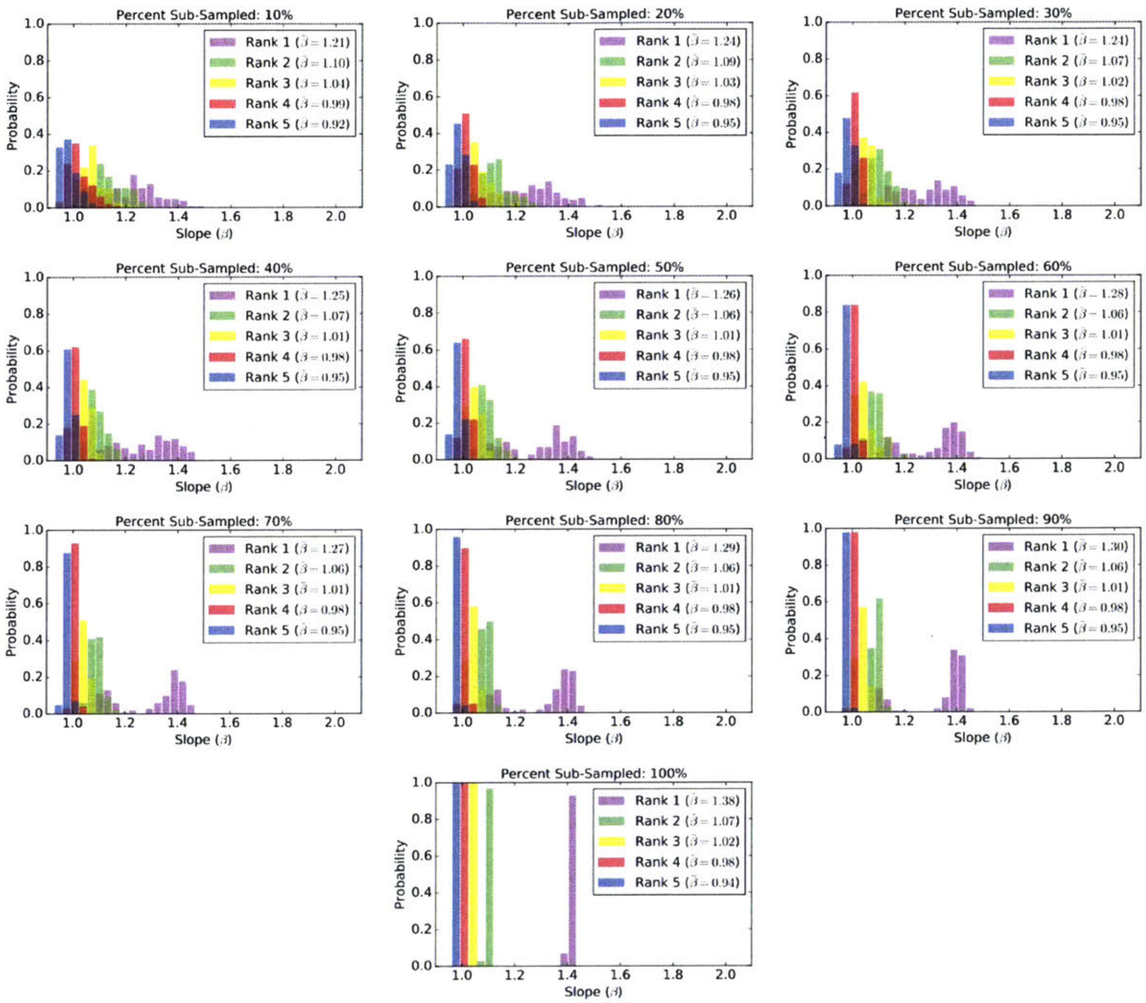


Figure B-15: Boot-strapping at various rates of sub-sampling demonstrates the stability of our result that one job cluster scales at a greater rate than the rest when using five clusters obtained from k-means clustering.

Checking the Statistical Robustness of Job Group Scaling

Readers who are familiar with the urban scaling literature may be aware of an ongoing debate about the statistical significance of exponent measurements and identification of underlying statistical models to explain that growth. For example, what model should one assume to test if a trend is significantly superlinear? Rather than solving this ongoing and important problem, the goal of this study is only to understand the relationship between automation and urbanization. Our narrative requires only that highly specialized occupations (represented by purple dots in Figure 3-3A) exhibit superlinear growth and be notably different from the growth exhibited by other occupations.

Recent work by Leitao et al. [140] proposes several statistical models that may explain urban scaling trends, and they apply them to a variety of datasets to test the models' ability to explain urban scaling. Here, we employ these same models to test if our requirements on the scaling of highly specialized occupations are met according to the five job groups discussed in the main analysis (i.e. K-means clustering with $k = 5$). As an example, Figure B-16 provides estimates of the scaling exponent along with standard errors for the scaling of each job group according to the unconstrained logarithm model. Table B.9 details the complete analysis in line with the methods in [140]. For each model tested, we find that the purple job group, which represents highly specialized occupations, exhibits significantly superlinear scaling, and, furthermore, consistently exhibits faster growth rates than other job groups.

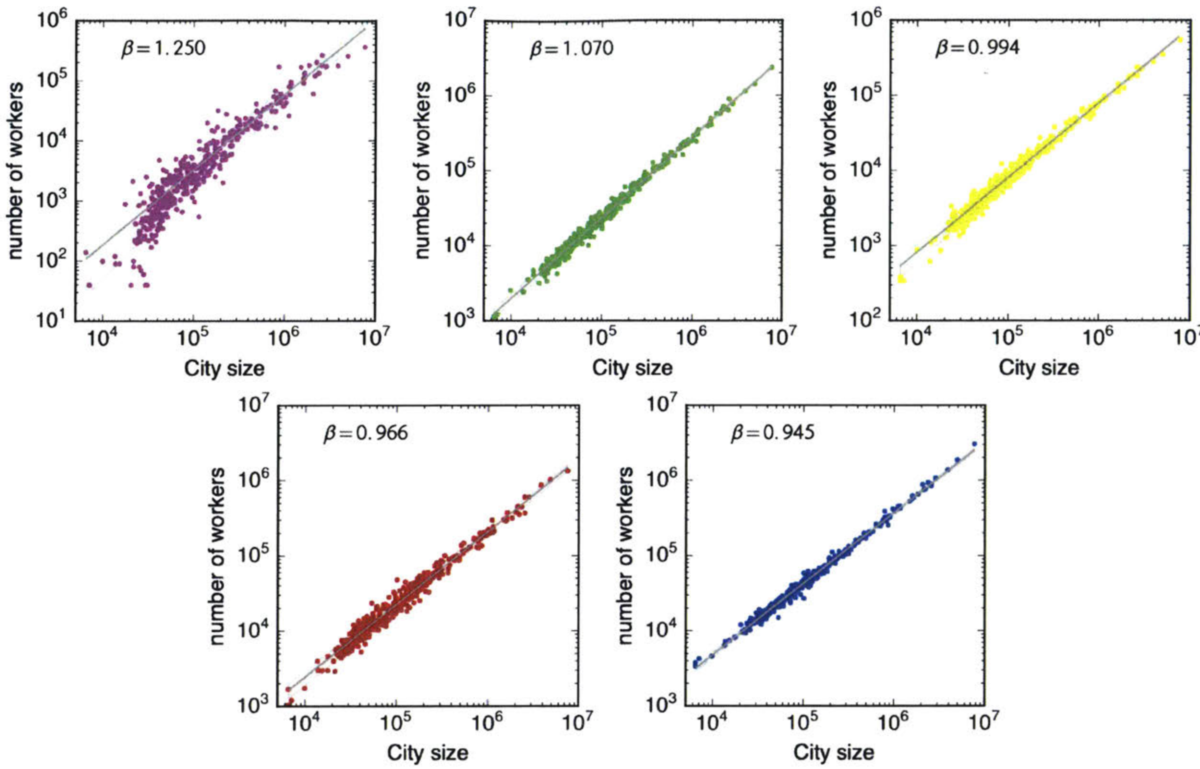


Figure B-16: Following the work of Leitao et al., we provide slope estimates along with standard errors for the scaling of each occupation cluster according to unconstrained logarithm model.

	city model				
	lognormal		Gaussian		person model
data	$\delta = 2$	$\delta \in [1,3]$	$\delta = 1$	$\delta \in [1,2]$	
purple	1.380 (0.226)$\nearrow$	1.250 (0.079)$\nearrow$	1.289 (0.234)$\nearrow$	1.284 (0.065)$\nearrow$	1.130 (0.102)$\nearrow$
green	1.075 (0.097)$\nearrow$	1.070 (0.044)$\nearrow$	1.065 (0.013)$\nearrow$	1.071 (0.009)$\nearrow^{*}$	1.061 (0.012)$\nearrow$
yellow	1.021 (0.060)$\circ^{*}$	0.994 (0.040)$\rightarrow$	0.980 (0.018)$\searrow$	0.996 (0.014)$\rightarrow$	0.976 (0.017)$\searrow$
red	0.976 (0.038)$\circ^{*}$	0.966 (0.013)$\searrow^{*}$	0.963 (0.022)$\searrow$	0.966 (0.015)$\searrow$	0.965 (0.020)$\searrow$
blue	0.943 (0.009)$\searrow$	0.945 (0.010)$\searrow^{*}$	0.974 (0.019)$\searrow$	0.945 (0.009)$\searrow$	0.971 (0.018)$\searrow$

Table B.9: An analysis of scaling exponents following the work of Leitao et al. The entries in the table represent the scaling exponent β. The value obtained through least-squares fitting in log-scale coincides with the value reported in the first column. The error bars were computed with bootstrap. The asterisk indicates that the model has a p-value higher than 0.05. If the difference ΔBIC between the BIC of each model with the same model with a fixed $\beta = 1$ is below 0, the model is linear ($\rightarrow$), between zero and six is inconclusive ($\circ$) and higher than six (strong evidence) is super-linear ($\nearrow$)/sublinear ($\searrow$).

B.6.4 Relating City Trends to O*NET Skill

Skill	Corr. to Job Impact	Corr to $H_{skill}(m)$	Corr. to Log_{10} City Size
Thinking Creatively	-0.71 ($< 10^{-56}$)	-0.24 ($< 10^{-4}$)	0.75 ($< 10^{-67}$)
Category Flexibility	-0.70 ($< 10^{-56}$)	-0.22 ($< 10^{-3}$)	0.63 ($< 10^{-41}$)
Inductive Reasoning	-0.70 ($< 10^{-56}$)	-0.27 ($< 10^{-5}$)	0.62 ($< 10^{-40}$)
Deductive Reasoning	-0.70 ($< 10^{-55}$)	-0.27 ($< 10^{-6}$)	0.57 ($< 10^{-32}$)
Active Learning	-0.70 ($< 10^{-55}$)	-0.32 ($< 10^{-9}$)	0.52 ($< 10^{-26}$)
Problem Sensitivity	-0.70 ($< 10^{-54}$)	-0.22 ($< 10^{-3}$)	0.63 ($< 10^{-41}$)
Originality	-0.70 ($< 10^{-54}$)	-0.32 ($< 10^{-9}$)	0.57 ($< 10^{-32}$)
Information Ordering	-0.70 ($< 10^{-54}$)	-0.20 ($< 10^{-3}$)	0.65 ($< 10^{-46}$)
Critical Thinking	-0.69 ($< 10^{-54}$)	-0.29 ($< 10^{-6}$)	0.61 ($< 10^{-38}$)
Complex Problem Solving	-0.69 ($< 10^{-53}$)	-0.18 ($< 10^{-2}$)	0.69 ($< 10^{-52}$)
Interpreting the Meaning of Information for Others	-0.69 ($< 10^{-53}$)	-0.27 ($< 10^{-6}$)	0.65 ($< 10^{-44}$)
Writing	-0.68 ($< 10^{-52}$)	-0.40 ($< 10^{-14}$)	0.39 ($< 10^{-13}$)
Written Comprehension	-0.68 ($< 10^{-51}$)	-0.38 ($< 10^{-12}$)	0.41 ($< 10^{-15}$)
Artistic	-0.68 ($< 10^{-51}$)	-0.32 ($< 10^{-9}$)	0.79 ($< 10^{-81}$)
Reading Comprehension	-0.68 ($< 10^{-50}$)	-0.37 ($< 10^{-11}$)	0.43 ($< 10^{-16}$)
Judgment and Decision Making	-0.68 ($< 10^{-50}$)	-0.26 ($< 10^{-5}$)	0.52 ($< 10^{-25}$)
Science	-0.68 ($< 10^{-50}$)	-0.16 ($< 10^{-1}$)	0.80 ($< 10^{-84}$)
Written Expression	-0.68 ($< 10^{-50}$)	-0.40 ($< 10^{-14}$)	0.36 ($< 10^{-11}$)
Computers and Electronics	-0.67 ($< 10^{-49}$)	-0.25 ($< 10^{-4}$)	0.57 ($< 10^{-33}$)
Fluency of Ideas	-0.67 ($< 10^{-49}$)	-0.34 ($< 10^{-9}$)	0.47 ($< 10^{-20}$)
Analyzing Data or Information	-0.67 ($< 10^{-48}$)	-0.21 ($< 10^{-3}$)	0.66 ($< 10^{-46}$)
Investigative	-0.66 ($< 10^{-47}$)	-0.12 ($< 10^{0}$)	0.82 ($< 10^{-92}$)
Making Decisions and Solving Problems	-0.66 ($< 10^{-47}$)	-0.19 ($< 10^{-2}$)	0.72 ($< 10^{-59}$)
Communications and Media	-0.66 ($< 10^{-46}$)	-0.47 ($< 10^{-20}$)	0.50 ($< 10^{-23}$)
History and Archeology	-0.66 ($< 10^{-46}$)	-0.34 ($< 10^{-10}$)	0.61 ($< 10^{-37}$)
Updating and Using Relevant Knowledge	-0.65 ($< 10^{-45}$)	-0.25 ($< 10^{-4}$)	0.66 ($< 10^{-46}$)
Processing Information	-0.64 ($< 10^{-43}$)	-0.23 ($< 10^{-4}$)	0.56 ($< 10^{-31}$)
Developing Objectives and Strategies	-0.63 ($< 10^{-41}$)	-0.20 ($< 10^{-3}$)	0.46 ($< 10^{-19}$)
Programming	-0.63 ($< 10^{-41}$)	-0.14 ($< 10^{-1}$)	0.63 ($< 10^{-41}$)
Getting Information	-0.63 ($< 10^{-40}$)	-0.32 ($< 10^{-8}$)	0.57 ($< 10^{-31}$)
Spend Time Sitting	-0.63 ($< 10^{-40}$)	-0.40 ($< 10^{-14}$)	0.51 ($< 10^{-24}$)
Operations Analysis	-0.62 ($< 10^{-40}$)	-0.14 ($< 10^{-1}$)	0.66 ($< 10^{-46}$)
Systems Analysis	-0.62 ($< 10^{-40}$)	-0.21 ($< 10^{-3}$)	0.37 ($< 10^{-12}$)
Systems Evaluation	-0.61 ($< 10^{-38}$)	-0.19 ($< 10^{-2}$)	0.35 ($< 10^{-10}$)
Flexibility of Closure	-0.61 ($< 10^{-38}$)	-0.07 ($< 10^{0}$)	0.83 ($< 10^{-97}$)
Fine Arts	-0.61 ($< 10^{-37}$)	-0.34 ($< 10^{-10}$)	0.79 ($< 10^{-81}$)
Near Vision	-0.61 ($< 10^{-37}$)	-0.16 ($< 10^{-1}$)	0.78 ($< 10^{-76}$)
Electronic Mail	-0.61 ($< 10^{-37}$)	-0.42 ($< 10^{-15}$)	0.25 ($< 10^{-5}$)

Documenting/Recording Information	-0.60 ($< 10^{-37}$)	-0.21 ($< 10^{-3}$)	0.47 ($< 10^{-20}$)
Identifying Objects, Actions, and Events	-0.60 ($< 10^{-37}$)	-0.24 ($< 10^{-4}$)	0.67 ($< 10^{-49}$)
Geography	-0.60 ($< 10^{-36}$)	-0.26 ($< 10^{-5}$)	0.61 ($< 10^{-37}$)
Technology Design	-0.60 ($< 10^{-36}$)	0.00 ($< 10^{0}$)	0.72 ($< 10^{-60}$)
Biology	-0.60 ($< 10^{-36}$)	-0.32 ($< 10^{-8}$)	0.63 ($< 10^{-41}$)
Freedom to Make Decisions	-0.60 ($< 10^{-36}$)	-0.21 ($< 10^{-3}$)	0.75 ($< 10^{-68}$)
Speed of Closure	-0.59 ($< 10^{-35}$)	-0.13 ($< 10^{0}$)	0.54 ($< 10^{-28}$)
Scheduling Work and Activities	-0.58 ($< 10^{-34}$)	-0.23 ($< 10^{-4}$)	0.31 ($< 10^{-8}$)
Selective Attention	-0.58 ($< 10^{-33}$)	0.03 ($< 10^{0}$)	0.84 ($< 10^{-101}$)
Education and Training	-0.58 ($< 10^{-33}$)	-0.08 ($< 10^{0}$)	0.34 ($< 10^{-10}$)
Estimating the Quantifiable Characteristics of Products, Events, or Information	-0.58 ($< 10^{-33}$)	-0.02 ($< 10^{0}$)	0.81 ($< 10^{-86}$)
Interacting With Computers	-0.57 ($< 10^{-32}$)	-0.30 ($< 10^{-7}$)	0.34 ($< 10^{-9}$)
Mathematical Reasoning	-0.56 ($< 10^{-31}$)	-0.19 ($< 10^{-2}$)	0.29 ($< 10^{-7}$)
Mathematics	-0.56 ($< 10^{-31}$)	-0.04 ($< 10^{0}$)	0.52 ($< 10^{-25}$)
Judging the Qualities of Things, Services, or People	-0.56 ($< 10^{-30}$)	-0.11 ($< 10^{0}$)	0.58 ($< 10^{-33}$)
English Language	-0.56 ($< 10^{-30}$)	-0.43 ($< 10^{-16}$)	0.17 ($< 10^{-2}$)
Provide Consultation and Advice to Others	-0.55 ($< 10^{-29}$)	-0.21 ($< 10^{-3}$)	0.27 ($< 10^{-6}$)
Monitoring	-0.55 ($< 10^{-29}$)	-0.07 ($< 10^{0}$)	0.26 ($< 10^{-5}$)
Physics	-0.53 ($< 10^{-27}$)	0.11 ($< 10^{0}$)	0.83 ($< 10^{-94}$)
Design	-0.53 ($< 10^{-26}$)	0.09 ($< 10^{0}$)	0.83 ($< 10^{-96}$)
Structured versus Unstructured Work	-0.53 ($< 10^{-26}$)	-0.34 ($< 10^{-9}$)	0.33 ($< 10^{-9}$)
Engineering and Technology	-0.52 ($< 10^{-26}$)	0.13 ($< 10^{0}$)	0.82 ($< 10^{-90}$)
Visualization	-0.52 ($< 10^{-25}$)	0.06 ($< 10^{0}$)	0.87 ($< 10^{-114}$)
Oral Comprehension	-0.52 ($< 10^{-25}$)	-0.48 ($< 10^{-21}$)	0.07 ($< 10^{0}$)
Memorization	-0.51 ($< 10^{-24}$)	-0.27 ($< 10^{-6}$)	0.08 ($< 10^{0}$)
Duration of Typical Work Week	-0.50 ($< 10^{-23}$)	0.17 ($< 10^{-2}$)	0.67 ($< 10^{-49}$)
Level of Competition	-0.50 ($< 10^{-23}$)	-0.16 ($< 10^{-1}$)	0.79 ($< 10^{-80}$)
Oral Expression	-0.49 ($< 10^{-22}$)	-0.48 ($< 10^{-21}$)	0.03 ($< 10^{0}$)
Organizing, Planning, and Prioritizing Work	-0.49 ($< 10^{-22}$)	-0.33 ($< 10^{-9}$)	0.09 ($< 10^{0}$)
Number Facility	-0.48 ($< 10^{-21}$)	-0.17 ($< 10^{-2}$)	0.21 ($< 10^{-3}$)
Active Listening	-0.48 ($< 10^{-21}$)	-0.49 ($< 10^{-22}$)	0.04 ($< 10^{0}$)
Instructing	-0.48 ($< 10^{-21}$)	-0.22 ($< 10^{-3}$)	-0.03 ($< 10^{0}$)
Telecommunications	-0.47 ($< 10^{-20}$)	-0.22 ($< 10^{-3}$)	0.42 ($< 10^{-15}$)
Indoors, Environmentally Controlled	-0.47 ($< 10^{-20}$)	-0.54 ($< 10^{-28}$)	0.21 ($< 10^{-3}$)
Learning Strategies	-0.47 ($< 10^{-20}$)	-0.26 ($< 10^{-5}$)	-0.06 ($< 10^{0}$)
Chemistry	-0.47 ($< 10^{-20}$)	0.13 ($< 10^{-1}$)	0.78 ($< 10^{-75}$)
Far Vision	-0.46 ($< 10^{-19}$)	0.08 ($< 10^{0}$)	0.71 ($< 10^{-57}$)

Monitor Processes, Materials, or Surroundings	-0.46 ($< 10^{-19}$)	0.14 ($< 10^{-1}$)	0.65 ($< 10^{-45}$)
Drafting, Laying Out, and Specifying Technical Devices, Parts, and Equipment	-0.45 ($< 10^{-18}$)	0.18 ($< 10^{-2}$)	0.80 ($< 10^{-84}$)
Perceptual Speed	-0.45 ($< 10^{-18}$)	0.16 ($< 10^{-1}$)	0.77 ($< 10^{-74}$)
Evaluating Information to Determine Compliance with Standards	-0.44 ($< 10^{-17}$)	-0.00 ($< 10^{0}$)	0.30 ($< 10^{-7}$)
Public Speaking	-0.44 ($< 10^{-17}$)	-0.31 ($< 10^{-8}$)	-0.03 ($< 10^{0}$)
Speaking	-0.43 ($< 10^{-16}$)	-0.47 ($< 10^{-21}$)	-0.07 ($< 10^{0}$)
Importance of Being Exact or Accurate	-0.42 ($< 10^{-15}$)	-0.07 ($< 10^{0}$)	0.80 ($< 10^{-84}$)
Visual Color Discrimination	-0.41 ($< 10^{-14}$)	0.09 ($< 10^{0}$)	0.84 ($< 10^{-101}$)
Communicating with Supervisors, Peers, or Subordinates	-0.40 ($< 10^{-14}$)	-0.16 ($< 10^{-1}$)	-0.01 ($< 10^{0}$)
Face-to-Face Discussions	-0.38 ($< 10^{-12}$)	-0.11 ($< 10^{0}$)	0.34 ($< 10^{-9}$)
Letters and Memos	-0.38 ($< 10^{-12}$)	-0.37 ($< 10^{-12}$)	-0.12 ($< 10^{0}$)
Impact of Decisions on Co-workers or Company Results	-0.37 ($< 10^{-12}$)	-0.14 ($< 10^{-1}$)	0.48 ($< 10^{-21}$)
Time Management	-0.37 ($< 10^{-11}$)	-0.27 ($< 10^{-5}$)	-0.18 ($< 10^{-2}$)
Training and Teaching Others	-0.36 ($< 10^{-11}$)	-0.04 ($< 10^{0}$)	0.02 ($< 10^{0}$)
Quality Control Analysis	-0.36 ($< 10^{-11}$)	0.25 ($< 10^{-5}$)	0.80 ($< 10^{-84}$)
Consequence of Error	-0.35 ($< 10^{-10}$)	0.16 ($< 10^{-1}$)	0.75 ($< 10^{-69}$)
Communicating with Persons Outside Organization	-0.35 ($< 10^{-10}$)	-0.50 ($< 10^{-23}$)	-0.03 ($< 10^{0}$)
Repairing and Maintaining Electronic Equipment	-0.35 ($< 10^{-10}$)	0.19 ($< 10^{-2}$)	0.81 ($< 10^{-88}$)
Speech Clarity	-0.35 ($< 10^{-10}$)	-0.47 ($< 10^{-20}$)	-0.18 ($< 10^{-2}$)
Philosophy and Theology	-0.34 ($< 10^{-9}$)	-0.35 ($< 10^{-10}$)	-0.06 ($< 10^{0}$)
Monitoring and Controlling Resources	-0.33 ($< 10^{-9}$)	-0.28 ($< 10^{-6}$)	0.01 ($< 10^{0}$)
Management of Personnel Resources	-0.33 ($< 10^{-9}$)	-0.18 ($< 10^{-2}$)	-0.19 ($< 10^{-2}$)
Sociology and Anthropology	-0.33 ($< 10^{-9}$)	-0.34 ($< 10^{-10}$)	-0.13 ($< 10^{0}$)
Equipment Selection	-0.28 ($< 10^{-6}$)	0.30 ($< 10^{-7}$)	0.79 ($< 10^{-80}$)
Law and Government	-0.27 ($< 10^{-6}$)	-0.27 ($< 10^{-5}$)	-0.08 ($< 10^{0}$)
Exposed to Radiation	-0.27 ($< 10^{-6}$)	-0.09 ($< 10^{0}$)	0.41 ($< 10^{-14}$)
Troubleshooting	-0.27 ($< 10^{-6}$)	0.32 ($< 10^{-8}$)	0.76 ($< 10^{-72}$)
Time Pressure	-0.27 ($< 10^{-6}$)	0.07 ($< 10^{0}$)	0.60 ($< 10^{-36}$)
Work Schedules	-0.26 ($< 10^{-5}$)	-0.01 ($< 10^{0}$)	0.71 ($< 10^{-57}$)
Operation Monitoring	-0.25 ($< 10^{-5}$)	0.33 ($< 10^{-9}$)	0.75 ($< 10^{-67}$)
Mechanical	-0.25 ($< 10^{-5}$)	0.38 ($< 10^{-12}$)	0.73 ($< 10^{-62}$)
Telephone	-0.23 ($< 10^{-4}$)	-0.44 ($< 10^{-17}$)	-0.17 ($< 10^{-2}$)
Persuasion	-0.23 ($< 10^{-4}$)	-0.32 ($< 10^{-8}$)	-0.27 ($< 10^{-6}$)

Management of Material Resources	-0.22 ($< 10^{-3}$)	-0.07 ($< 10^{0}$)	-0.24 ($< 10^{-4}$)
Production and Processing	-0.20 ($< 10^{-3}$)	0.35 ($< 10^{-10}$)	0.60 ($< 10^{-36}$)
Building and Construction	-0.20 ($< 10^{-3}$)	0.32 ($< 10^{-8}$)	0.51 ($< 10^{-24}$)
Wear Specialized Protective or Safety Equipment such as Breathing Apparatus, Safety Harness, Full Protection Suits, or Radiation Protection	-0.20 ($< 10^{-3}$)	0.32 ($< 10^{-8}$)	0.66 ($< 10^{-47}$)
Management of Financial Resources	-0.19 ($< 10^{-2}$)	-0.18 ($< 10^{-2}$)	-0.26 ($< 10^{-5}$)
Depth Perception	-0.18 ($< 10^{-2}$)	0.34 ($< 10^{-10}$)	0.73 ($< 10^{-61}$)
Establishing and Maintaining Interpersonal Relationships	-0.18 ($< 10^{-2}$)	-0.41 ($< 10^{-15}$)	-0.35 ($< 10^{-10}$)
Exposed to Hazardous Conditions	-0.17 ($< 10^{-2}$)	0.37 ($< 10^{-12}$)	0.69 ($< 10^{-53}$)
Hearing Sensitivity	-0.17 ($< 10^{-2}$)	0.40 ($< 10^{-14}$)	0.60 ($< 10^{-37}$)
Coaching and Developing Others	-0.16 ($< 10^{-1}$)	-0.13 ($< 10^{-1}$)	-0.39 ($< 10^{-13}$)
Finger Dexterity	-0.16 ($< 10^{-1}$)	0.24 ($< 10^{-4}$)	0.71 ($< 10^{-57}$)
Installation	-0.16 ($< 10^{-1}$)	0.37 ($< 10^{-11}$)	0.58 ($< 10^{-34}$)
Developing and Building Teams	-0.15 ($< 10^{-1}$)	-0.11 ($< 10^{0}$)	-0.39 ($< 10^{-13}$)
Guiding, Directing, and Motivating Subordinates	-0.15 ($< 10^{-1}$)	-0.08 ($< 10^{0}$)	-0.30 ($< 10^{-7}$)
Equipment Maintenance	-0.15 ($< 10^{-1}$)	0.38 ($< 10^{-12}$)	0.69 ($< 10^{-54}$)
Repairing	-0.14 ($< 10^{-1}$)	0.40 ($< 10^{-14}$)	0.68 ($< 10^{-51}$)
Medicine and Dentistry	-0.13 ($< 10^{-1}$)	-0.27 ($< 10^{-6}$)	-0.11 ($< 10^{0}$)
Third Interest High-Point	-0.13 ($< 10^{0}$)	-0.07 ($< 10^{0}$)	-0.18 ($< 10^{-2}$)
Negotiation	-0.12 ($< 10^{0}$)	-0.27 ($< 10^{-6}$)	-0.42 ($< 10^{-15}$)
Inspecting Equipment, Structures, or Material	-0.12 ($< 10^{0}$)	0.39 ($< 10^{-13}$)	0.64 ($< 10^{-43}$)
Operation and Control	-0.12 ($< 10^{0}$)	0.36 ($< 10^{-11}$)	0.69 ($< 10^{-52}$)
Realistic	-0.11 ($< 10^{0}$)	0.31 ($< 10^{-8}$)	0.70 ($< 10^{-55}$)
Coordination	-0.11 ($< 10^{0}$)	-0.20 ($< 10^{-3}$)	-0.50 ($< 10^{-23}$)
Controlling Machines and Processes	-0.10 ($< 10^{0}$)	0.38 ($< 10^{-12}$)	0.67 ($< 10^{-48}$)
Coordinating the Work and Activities of Others	-0.10 ($< 10^{0}$)	-0.10 ($< 10^{0}$)	-0.45 ($< 10^{-18}$)
Social Perceptiveness	-0.10 ($< 10^{0}$)	-0.30 ($< 10^{-7}$)	-0.46 ($< 10^{-19}$)
Auditory Attention	-0.09 ($< 10^{0}$)	0.47 ($< 10^{-20}$)	0.58 ($< 10^{-33}$)
Speech Recognition	-0.08 ($< 10^{0}$)	-0.38 ($< 10^{-12}$)	-0.50 ($< 10^{-23}$)
Repairing and Maintaining Mechanical Equipment	-0.08 ($< 10^{0}$)	0.43 ($< 10^{-17}$)	0.63 ($< 10^{-41}$)
Foreign Language	-0.07 ($< 10^{0}$)	-0.21 ($< 10^{-3}$)	-0.33 ($< 10^{-9}$)
Psychology	-0.07 ($< 10^{0}$)	-0.27 ($< 10^{-5}$)	-0.42 ($< 10^{-16}$)
Degree of Automation	-0.07 ($< 10^{0}$)	0.18 ($< 10^{-2}$)	0.36 ($< 10^{-11}$)

Administration and Management	-0.06 ($< 10^{0}$)	-0.11 ($< 10^{0}$)	-0.46 ($< 10^{-19}$)
Staffing Organizational Units	-0.06 ($< 10^{0}$)	-0.20 ($< 10^{-3}$)	-0.44 ($< 10^{-18}$)
Spend Time Using Your Hands to Handle, Control, or Feel Objects, Tools, or Controls	-0.06 ($< 10^{0}$)	0.22 ($< 10^{-3}$)	0.65 ($< 10^{-45}$)
Sales and Marketing	-0.03 ($< 10^{0}$)	-0.31 ($< 10^{-8}$)	-0.18 ($< 10^{-2}$)
Control Precision	-0.03 ($< 10^{0}$)	0.34 ($< 10^{-9}$)	0.64 ($< 10^{-43}$)
Performing Administrative Activities	-0.03 ($< 10^{0}$)	-0.29 ($< 10^{-7}$)	-0.51 ($< 10^{-25}$)
Exposed to Hazardous Equipment	-0.03 ($< 10^{0}$)	0.49 ($< 10^{-22}$)	0.56 ($< 10^{-30}$)
Exposed to High Places	-0.03 ($< 10^{0}$)	0.43 ($< 10^{-16}$)	0.47 ($< 10^{-20}$)
Rate Control	-0.02 ($< 10^{0}$)	0.42 ($< 10^{-15}$)	0.61 ($< 10^{-37}$)
Indoors, Not Environmentally Controlled	-0.01 ($< 10^{0}$)	0.52 ($< 10^{-26}$)	0.48 ($< 10^{-21}$)
Public Safety and Security	-0.00 ($< 10^{0}$)	0.21 ($< 10^{-3}$)	-0.03 ($< 10^{0}$)
Reaction Time	-0.00 ($< 10^{0}$)	0.44 ($< 10^{-17}$)	0.57 ($< 10^{-32}$)
Selling or Influencing Others	0.00 ($< 10^{0}$)	-0.30 ($< 10^{-7}$)	-0.34 ($< 10^{-9}$)
Pace Determined by Speed of Equipment	0.01 ($< 10^{0}$)	0.40 ($< 10^{-14}$)	0.53 ($< 10^{-27}$)
Clerical	0.02 ($< 10^{0}$)	-0.24 ($< 10^{-4}$)	-0.56 ($< 10^{-31}$)
Required Level of Education	0.02 ($< 10^{0}$)	-0.04 ($< 10^{0}$)	0.02 ($< 10^{0}$)
Glare Sensitivity	0.03 ($< 10^{0}$)	0.43 ($< 10^{-16}$)	0.50 ($< 10^{-23}$)
In an Enclosed Vehicle or Equipment	0.04 ($< 10^{0}$)	0.07 ($< 10^{0}$)	-0.00 ($< 10^{0}$)
Extremely Bright or Inadequate Lighting	0.04 ($< 10^{0}$)	0.45 ($< 10^{-18}$)	0.46 ($< 10^{-19}$)
Therapy and Counseling	0.04 ($< 10^{0}$)	-0.13 ($< 10^{0}$)	-0.49 ($< 10^{-22}$)
Frequency of Decision Making	0.04 ($< 10^{0}$)	-0.09 ($< 10^{0}$)	-0.06 ($< 10^{0}$)
Outdoors, Under Cover	0.05 ($< 10^{0}$)	0.25 ($< 10^{-5}$)	0.12 ($< 10^{0}$)
Personnel and Human Resources	0.06 ($< 10^{0}$)	-0.10 ($< 10^{0}$)	-0.59 ($< 10^{-35}$)
Transportation	0.06 ($< 10^{0}$)	0.29 ($< 10^{-6}$)	0.11 ($< 10^{0}$)
Wear Common Protective or Safety Equipment such as Safety Shoes, Glasses, Gloves, Hearing Protection, Hard Hats, or Life Jackets	0.06 ($< 10^{0}$)	0.52 ($< 10^{-25}$)	0.45 ($< 10^{-18}$)
On-the-Job Training	0.06 ($< 10^{0}$)	-0.02 ($< 10^{0}$)	-0.10 ($< 10^{0}$)
Responsibility for Outcomes and Results	0.07 ($< 10^{0}$)	0.27 ($< 10^{-5}$)	-0.17 ($< 10^{-2}$)
Work With Work Group or Team	0.07 ($< 10^{0}$)	-0.11 ($< 10^{0}$)	-0.56 ($< 10^{-31}$)
Arm-Hand Steadiness	0.07 ($< 10^{0}$)	0.30 ($< 10^{-7}$)	0.51 ($< 10^{-24}$)
Coordinate or Lead Others	0.08 ($< 10^{0}$)	-0.04 ($< 10^{0}$)	-0.63 ($< 10^{-41}$)
Exposed to Contaminants	0.09 ($< 10^{0}$)	0.50 ($< 10^{-23}$)	0.44 ($< 10^{-17}$)
Related Work Experience	0.09 ($< 10^{0}$)	0.16 ($< 10^{-1}$)	-0.34 ($< 10^{-9}$)
Spatial Orientation	0.10 ($< 10^{0}$)	0.42 ($< 10^{-16}$)	0.34 ($< 10^{-9}$)
Sounds, Noise Levels Are Distracting or Uncomfortable	0.10 ($< 10^{0}$)	0.56 ($< 10^{-30}$)	0.34 ($< 10^{-9}$)
Wrist-Finger Speed	0.11 ($< 10^{0}$)	0.37 ($< 10^{-12}$)	0.44 ($< 10^{-17}$)

Response Orientation	0.11 ($< 10^{0}$)	0.45 ($< 10^{-18}$)	0.40 ($< 10^{-14}$)
Night Vision	0.12 ($< 10^{0}$)	0.42 ($< 10^{-15}$)	0.33 ($< 10^{-9}$)
Cramped Work Space, Awkward Positions	0.12 ($< 10^{0}$)	0.46 ($< 10^{-19}$)	0.37 ($< 10^{-12}$)
Spend Time Climbing Ladders, Scaffolds, or Poles	0.12 ($< 10^{0}$)	0.47 ($< 10^{-20}$)	0.26 ($< 10^{-5}$)
On-Site or In-Plant Training	0.12 ($< 10^{0}$)	-0.05 ($< 10^{0}$)	-0.25 ($< 10^{-5}$)
Operating Vehicles, Mechanized Devices, or Equipment	0.13 ($< 10^{0}$)	0.49 ($< 10^{-22}$)	0.33 ($< 10^{-9}$)
Economics and Accounting	0.13 ($< 10^{-1}$)	-0.21 ($< 10^{-3}$)	-0.57 ($< 10^{-31}$)
Sound Localization	0.14 ($< 10^{-1}$)	0.49 ($< 10^{-23}$)	0.31 ($< 10^{-7}$)
Manual Dexterity	0.16 ($< 10^{-1}$)	0.33 ($< 10^{-9}$)	0.44 ($< 10^{-17}$)
Peripheral Vision	0.17 ($< 10^{-1}$)	0.46 ($< 10^{-19}$)	0.26 ($< 10^{-5}$)
Importance of Repeating Same Tasks	0.19 ($< 10^{-2}$)	0.04 ($< 10^{0}$)	-0.08 ($< 10^{0}$)
Outdoors, Exposed to Weather	0.20 ($< 10^{-3}$)	0.27 ($< 10^{-5}$)	-0.06 ($< 10^{0}$)
Service Orientation	0.21 ($< 10^{-3}$)	-0.28 ($< 10^{-6}$)	-0.75 ($< 10^{-68}$)
Resolving Conflicts and Negotiating with Others	0.23 ($< 10^{-4}$)	-0.14 ($< 10^{-1}$)	-0.75 ($< 10^{-68}$)
Social	0.24 ($< 10^{-4}$)	-0.21 ($< 10^{-3}$)	-0.73 ($< 10^{-61}$)
Spend Time Making Repetitive Motions	0.25 ($< 10^{-5}$)	0.17 ($< 10^{-2}$)	0.18 ($< 10^{-2}$)
Multilimb Coordination	0.26 ($< 10^{-5}$)	0.42 ($< 10^{-15}$)	0.28 ($< 10^{-6}$)
Deal With External Customers	0.26 ($< 10^{-5}$)	-0.30 ($< 10^{-7}$)	-0.66 ($< 10^{-47}$)
Time Sharing	0.27 ($< 10^{-5}$)	0.00 ($< 10^{0}$)	-0.68 ($< 10^{-50}$)
In an Open Vehicle or Equipment	0.27 ($< 10^{-6}$)	0.56 ($< 10^{-30}$)	0.09 ($< 10^{0}$)
Exposed to Whole Body Vibration	0.28 ($< 10^{-6}$)	0.46 ($< 10^{-19}$)	-0.06 ($< 10^{0}$)
Customer and Personal Service	0.28 ($< 10^{-6}$)	-0.22 ($< 10^{-3}$)	-0.74 ($< 10^{-64}$)
Performing for or Working Directly with the Public	0.29 ($< 10^{-7}$)	-0.32 ($< 10^{-8}$)	-0.62 ($< 10^{-40}$)
Exposed to Disease or Infections	0.32 ($< 10^{-9}$)	-0.10 ($< 10^{0}$)	-0.69 ($< 10^{-53}$)
Contact With Others	0.33 ($< 10^{-9}$)	-0.12 ($< 10^{0}$)	-0.73 ($< 10^{-62}$)
Exposed to Minor Burns, Cuts, Bites, or Stings	0.38 ($< 10^{-12}$)	0.56 ($< 10^{-30}$)	0.03 ($< 10^{0}$)
Handling and Moving Objects	0.38 ($< 10^{-12}$)	0.43 ($< 10^{-16}$)	0.10 ($< 10^{0}$)
Assisting and Caring for Others	0.38 ($< 10^{-12}$)	-0.05 ($< 10^{0}$)	-0.78 ($< 10^{-77}$)
Gross Body Equilibrium	0.38 ($< 10^{-13}$)	0.49 ($< 10^{-22}$)	-0.02 ($< 10^{0}$)
Frequency of Conflict Situations	0.38 ($< 10^{-13}$)	0.01 ($< 10^{0}$)	-0.84 ($< 10^{-98}$)
Deal With Physically Aggressive People	0.39 ($< 10^{-13}$)	0.04 ($< 10^{0}$)	-0.73 ($< 10^{-61}$)
Explosive Strength	0.39 ($< 10^{-13}$)	0.20 ($< 10^{-3}$)	-0.44 ($< 10^{-17}$)
Very Hot or Cold Temperatures	0.40 ($< 10^{-13}$)	0.54 ($< 10^{-27}$)	-0.05 ($< 10^{0}$)
Enterprising	0.40 ($< 10^{-14}$)	-0.13 ($< 10^{0}$)	-0.86 ($< 10^{-109}$)
First Interest High-Point	0.40 ($< 10^{-14}$)	-0.13 ($< 10^{-1}$)	-0.89 ($< 10^{-127}$)

Performing General Physical Activities	0.41 ($< 10^{-15}$)	0.48 ($< 10^{-21}$)	0.01 ($< 10^{0}$)
Spend Time Keeping or Regaining Balance	0.45 ($< 10^{-18}$)	0.47 ($< 10^{-20}$)	-0.11 ($< 10^{0}$)
Static Strength	0.46 ($< 10^{-19}$)	0.43 ($< 10^{-16}$)	-0.07 ($< 10^{0}$)
Physical Proximity	0.47 ($< 10^{-20}$)	-0.03 ($< 10^{0}$)	-0.61 ($< 10^{-38}$)
Food Production	0.47 ($< 10^{-20}$)	0.04 ($< 10^{0}$)	-0.73 ($< 10^{-63}$)
Dynamic Strength	0.48 ($< 10^{-21}$)	0.45 ($< 10^{-18}$)	-0.10 ($< 10^{0}$)
Responsible for Others' Health and Safety	0.48 ($< 10^{-22}$)	0.50 ($< 10^{-23}$)	-0.49 ($< 10^{-22}$)
Trunk Strength	0.53 ($< 10^{-26}$)	0.41 ($< 10^{-14}$)	-0.21 ($< 10^{-3}$)
Deal With Unpleasant or Angry People	0.53 ($< 10^{-27}$)	0.03 ($< 10^{0}$)	-0.83 ($< 10^{-98}$)
Spend Time Kneeling, Crouching, Stooping, or Crawling	0.53 ($< 10^{-27}$)	0.44 ($< 10^{-17}$)	-0.36 ($< 10^{-11}$)
Dynamic Flexibility	0.53 ($< 10^{-27}$)	0.25 ($< 10^{-4}$)	-0.53 ($< 10^{-26}$)
Conventional	0.54 ($< 10^{-28}$)	0.07 ($< 10^{0}$)	-0.88 ($< 10^{-120}$)
Extent Flexibility	0.55 ($< 10^{-29}$)	0.45 ($< 10^{-19}$)	-0.21 ($< 10^{-3}$)
Spend Time Bending or Twisting the Body	0.57 ($< 10^{-31}$)	0.47 ($< 10^{-20}$)	-0.26 ($< 10^{-5}$)
Spend Time Standing	0.57 ($< 10^{-32}$)	0.37 ($< 10^{-11}$)	-0.42 ($< 10^{-15}$)
Gross Body Coordination	0.58 ($< 10^{-33}$)	0.41 ($< 10^{-15}$)	-0.36 ($< 10^{-11}$)
Speed of Limb Movement	0.60 ($< 10^{-37}$)	0.46 ($< 10^{-20}$)	-0.37 ($< 10^{-12}$)
Stamina	0.60 ($< 10^{-37}$)	0.38 ($< 10^{-13}$)	-0.44 ($< 10^{-17}$)
Second Interest High-Point	0.65 ($< 10^{-44}$)	0.20 ($< 10^{-3}$)	-0.71 ($< 10^{-58}$)
Spend Time Walking and Running	0.67 ($< 10^{-49}$)	0.33 ($< 10^{-9}$)	-0.66 ($< 10^{-46}$)

B.6.5 Skill Types

We provide example O*NET skills from each of ten skill types. These groups of skills are obtained from the co-occurrence of skills across jobs. The left column provides a subjective labelling for each skill type based on the skills comprising that cluster of skills.

Skill Types	**O*NET Skills**
Computational and Analytical Skills	Active Learning, Analyzing Data or Information, Communications and Media, Complex Problem Solving, Computers and Electronics, Developing Objectives and Strategies, Documenting/Recording Information, Fluency of Ideas, Instructing, Interacting With Computers, Interpreting the Meaning of Information for Others, Judgement and Decision Making, Learning Strategies, Making Decisions and Solving Problems, Mathematical Reasoning, Memorization, Number Facility, Originality, Processing Information, Provide Consultation and Advice to Others, Systems Analysis, Systems Evaluation, Updating and Using Relevant Knowledge
Physical Planning and Construction	Building and Construction, Chemistry, Design, Drafting, Engineering and Technology, Estimating the Quantifiable Characteristics of Products, Events, or Information, Explosive Strength, Far Vision, Installation, Perceptual Speed, Physics, Production and Processing, Public Safety and Security,Transportation, Visualization

Harmful Workspace Management Skills	Cramped Work Space, Awkward Positions, Dynamic Flexibility, Exposed to Contaminants, Exposed to Hazardous Conditions, Exposed to High Places, Exposed to Minor Burns, Cuts, Bites, or Stings, Exposed to Whole Body Vibration, Extremely Bright or Inadequate Lighting, Finger Dexterity, Outdoors, Exposed to Weather, Pace Determined by Speed of Equipment, Responsible for Others' Health and Safety, Sounds, Noise Levels Are Distracting or Uncomfortable, Spend Time Bending or Twisting the Body, Spend Time Climbing Ladders, Scaffolds, or Poles, Spend Time Keeping or Regaining Balance, Spend Time Kneeling, Crouching, Stooping, or Crawling, Spend Time Making Repetitive Motions, Spend Time Standing, Spend Time Using Your Hands to Handle, Control, or Feel Objects, Tools, or Controls, Spend Time Walking and Running, Very Hot or Cold Temperatures, Wear Specialized Protective or Safety Equipment such as Breathing Apparatus, Safety Harness, Full Protection Suits, or Radiation Protection
Relational/Social Skills	Customer and Personal Service, Education and Training, Fine Arts, Foreign Language, History and Archeology, Identifying Objects, Actions, and Events, Law and Government, Performing for or Working Directly with the Public, Philosophy and Theology, Psychology, Resolving Conflicts and Negotiating with Others , Sales and Marketing, Selling or Influencing Others, Sociology and Anthropology, Therapy and Counseling, Training and Teaching Others
Control and Perceptual Skills	Auditory Attention, Depth Perception, Equipment Maintenance, Equipment Selection, Glare Sensitivity, Hearing Sensitivity, Inspecting Equipment, Structures, or Material, Mechanical, Night Vision, Operating Vehicles, Mechanized Devices, or Equipment, Operation Monitoring, Operation and Control, Peripheral Vision, Quality Control Analysis, Rate Control, Repairing, Repairing and Maintaining Electronic Equipment, Repairing and Maintaining Mechanical Equipment, Response Orientation, Sound Localization, Spatial Orientation, Troubleshooting, Visual Color Discrimination
Emergency Response	Consequence of Error, Contact With Others ,Coordinate or Lead Others, Deal With External Customers, Deal With Physically Aggressive People, Deal With Unpleasant or Angry People, Degree of Automation, Exposed to Disease or Infections, Exposed to Radiation, Frequency of Conflict Situations, Importance of Being Exact or Accurate, Importance of Repeating Same Tasks, Indoors, Environmentally Controlled, Near Vision, Responsibility for Outcomes and Results, Selective Attention
Basic skills	Artistic, Assisting and Caring for Others, Biology, Conventional, Duration of Typical, Work Week, Electronic Mail, Enterprising, Face-to-Face Discussions, First Interest High-Point, Flexibility of Closure, Food Production, Freedom to Make Decisions, Frequency of Decision Making, Impact of Decisions on Co-workers or Company Results, In an Enclosed Vehicle or Equipment, In an Open Vehicle or Equipment, Indoors, Not Environmentally Controlled, Investigative, Letters and Memos, Level of Competition, Medicine and Dentistry, Monitor Processes, Materials, or Surroundings, On-Site or In-Plant Training, On-the-Job Training, Outdoors, Under Cover, Physical Proximity, Public Speaking, Realistic, Related Work Experience, Required Level of Education, Second Interest High-Point, Social, Structured versus Unstructured Work, Technology Design, Telecommunications, Telephone, Third Interest High-Point, Time Pressure, Time Sharing,Work Schedules, Work With Work Group or Team

Organization Skills	Active Listening, Category Flexibility, Clerical, Communicating with Persons Outside, Organization, Communicating with Supervisors, Peers, or Subordinates, Coordination, Critical Thinking, Deductive Reasoning, English Language, Establishing and Maintaining Interpersonal Relationships, Getting Information, Inductive Reasoning, Information, Ordering, Monitoring, Negotiation, Oral Comprehension, Oral Expression, Organizing, Planning, and Prioritizing Work, Performing Administrative Activities, Persuasion, Problem, Sensitivity, Reading Comprehension, Service Orientation, Social Perceptiveness, Speaking, Speech Clarity, Speech Recognition, Spend Time Sitting, Time Management, Writing, Written Comprehension, Written Expression
Management Skills	Administration and Management, Coaching and Developing Others, Coordinating the Work and Activities of Others, Developing and Building Teams, Economics and Accounting, Evaluating Information to Determine Compliance with Standards, Geography, Guiding, Directing, and Motivating Subordinates, Judging the Qualities of Things, Services, or People, Management of Financial Resources, Management of Material Resources, Management of Personnel Resources, Mathematics, Monitoring and Controlling Resources, Operations Analysis, Personnel and Human Resources, Programming, Scheduling Work and Activities, Science, Speed of Closure, Staffing Organizational Units, Thinking Creatively
Physical Coordination	Arm-Hand Steadiness, Control Precision, Controlling Machines and Processes, Dynamic Strength, Exposed to Hazardous Equipment, Extent Flexibility, Gross Body Coordination, Gross Body Equilibrium, Handling and Moving Objects, Manual Dexterity, Multilimb Coordination, Performing General Physical Activities, Reaction Time, Speed of Limb Movement, Stamina, Static Strength, Trunk Strength, Wear Common Protective or Safety Equipment such as Safety Shoes, Glasses, Gloves, Hearing Protection, Hard Hats, or Life Jackets, Wrist-Finger Speed

THIS PAGE INTENTIONALLY LEFT BLANK

Appendix C

Unpacking the polarization of workplace skills

C.1 Exploring Occupations and their Constituent Skills

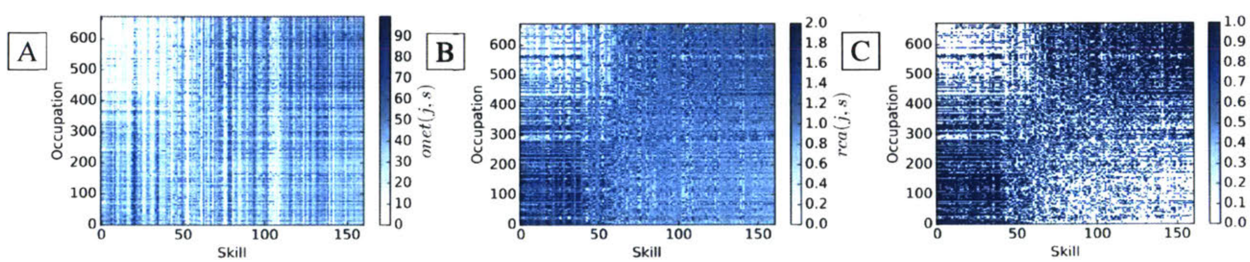


Figure C-1: Transforming raw O*NET data with RCA. **(A)** The raw occupation-skill matrix, $onet(j,s)$, **(B)** the RCA occupation-skill matrix, $rca(j,s)$, and **(C)** the thresholded RCA job-skill matrix, $I(j,s)$, for 2014. Here, $I(j,s) = 1$ if and only if $rca(j,s) > 1$. Occupations (y-axis) are ordered by the sum of threshold RCA skill values, and skills (x-axis) are ordered by the correlation of their thresholded RCA values across occupations to the occupational sums.

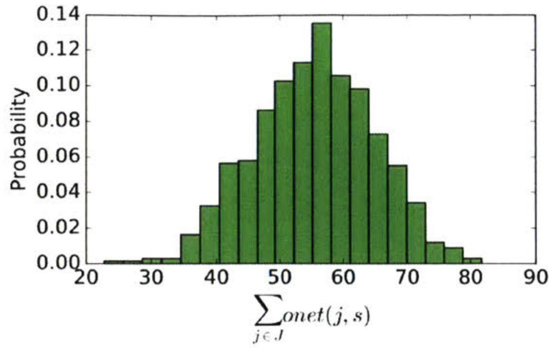


Figure C-2: The distribution of aggregate skill importance by summing raw O*NET values of each occupation.

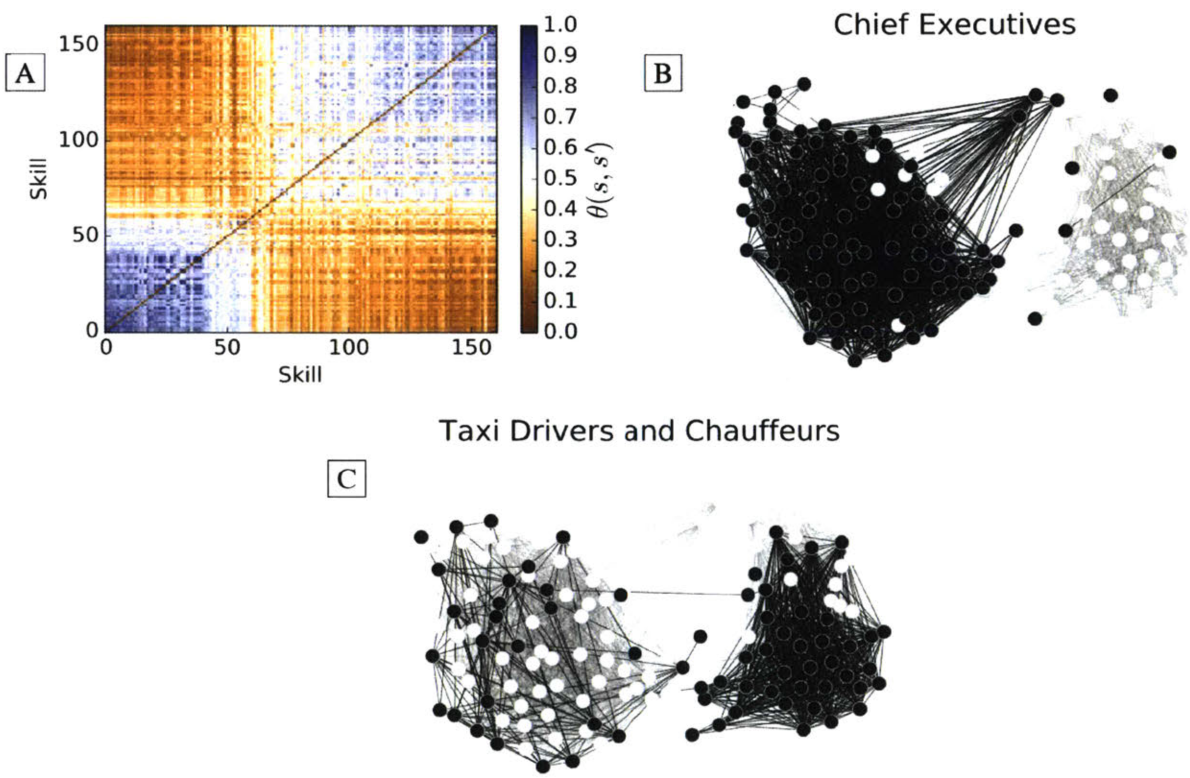


Figure C-3: Projecting occupational skill requirements onto the polarized skill network.labelsep The skill-skill matrix **(A)** defining the Skillscape from 2014 data. The projections of the occupations of Chief Executive **(C)** and Taxi Driver **(D)** onto the Skillscape based on their effectively used skills (black).

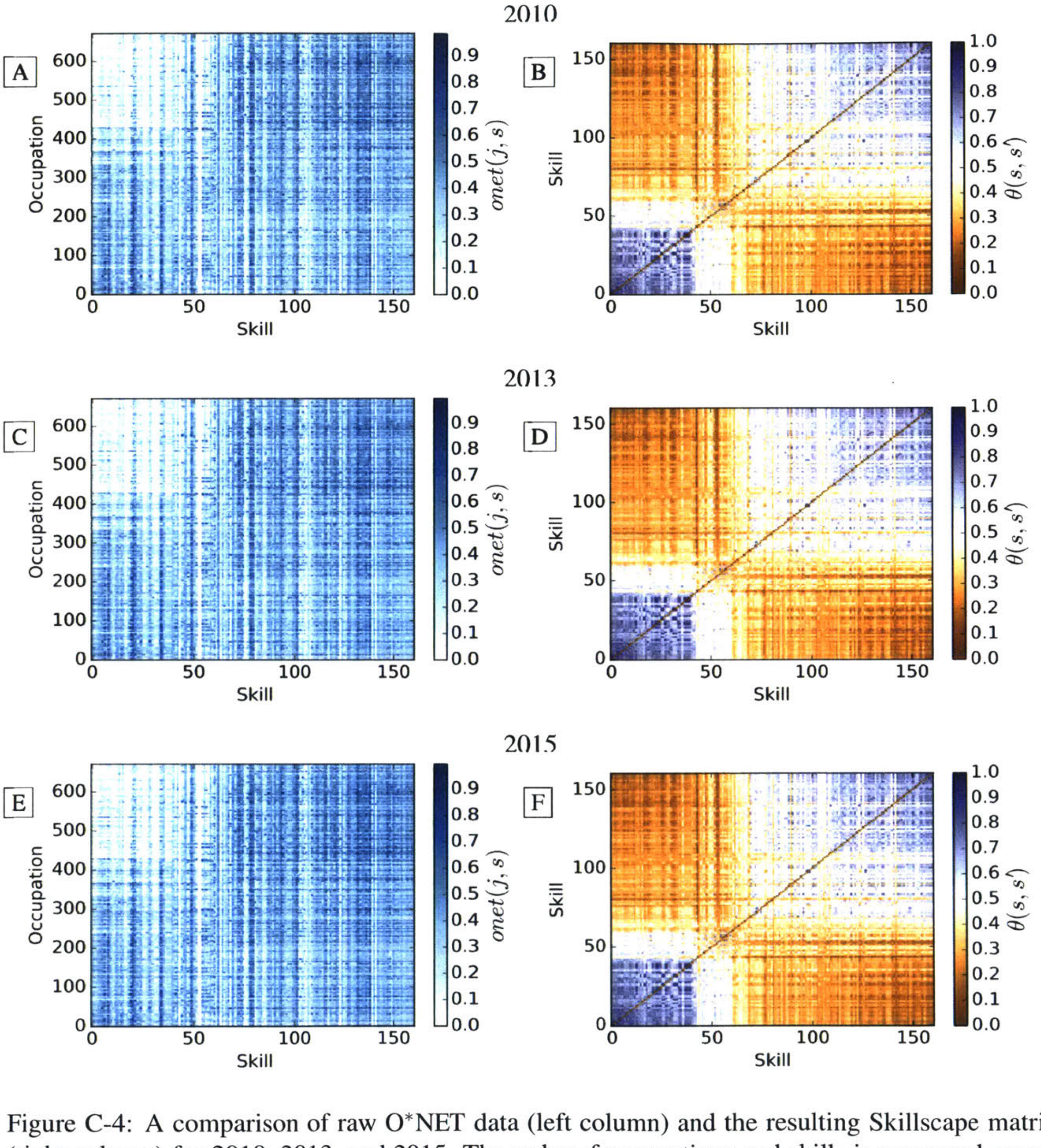


Figure C-4: A comparison of raw O*NET data (left column) and the resulting Skillscape matrix (right column) for 2010, 2013, and 2015. The order of occupations and skills is preserved across plots for easy comparison. The polarized skill structure is stable across years.

Skill Type	O*NET Skill
Socio-Cognitive Skills	Active Learning, Active Listening, Complex Problem Solving, Coordination, Critical Thinking, Instructing Judgment and Decision Making, Learning Strategies, Management of Financial Resources, Management of Material Resources, Management of Personnel Resources, Mathematics, Monitoring, Negotiation, Operations Analysis, Persuasion, Programming, Reading Comprehension, Science, Service Orientation, Social Perceptiveness, Speaking, Systems Analysis, Systems Evaluation, Time Management, Writing, Category Flexibility, Deductive Reasoning, Fluency of Ideas, Inductive Reasoning, Information Ordering, Mathematical Reasoning, Memorization, Near Vision, Number Facility, Oral Comprehension, Oral Expression, Originality, Problem Sensitivity, Speech Clarity, Speech Recognition, Speed of Closure, Written Comprehension, Written Expression, Administration and Management, Biology, Clerical, Communications and Media, Computers and Electronics, Customer and Personal Service, Economics and Accounting, Education and Training, English Language, Fine Arts, Foreign Language, Geography, History and Archeology, Law and Government, Mathematics Knowledge, Medicine and Dentistry, Personnel and Human Resources, Philosophy and Theology, Psychology, Sales and Marketing, Sociology and Anthropology, Telecommunications, Therapy and Counseling, Analyzing Data or Information, Assisting and Caring for Others, Coaching and Developing Others, Communicating with Persons Outside Organization, Communicating with Supervisors, Peers, or Subordinates, Coordinating the Work and Activities of Others, Developing Objectives and Strategies, Developing and Building Teams, Documenting/Recording Information, Establishing and Maintaining Interpersonal Relationships, Evaluating Information to Determine Compliance with Standards, Getting Information, Guiding, Directing, and Motivating Subordinates, Identifying Objects, Actions, and Events, Interacting With Computers, Interpreting the Meaning of Information for Others, Judging the Qualities of Things, Services, or People, Making Decisions and Solving Problems, Monitoring and Controlling Resources, Organizing, Planning, and Prioritizing Work, Performing Administrative Activities, Performing for or Working Directly with the Public, Processing Information, Provide Consultation and Advice to Others, Resolving Conflicts and Negotiating with Others, Scheduling Work and Activities, Selling or Influencing Others, Staffing Organizational Units, Thinking Creatively, Training and Teaching Others, Updating and Using Relevant Knowledge
Sensory-Physical Skills	Equipment Maintenance, Equipment Selection, Installation, Operation Monitoring, Operation and Control, Quality Control Analysis, Repairing, Technology Design, Troubleshooting, Arm-Hand Steadiness, Auditory Attention, Control Precision, Depth Perception, Dynamic Flexibility, Dynamic Strength, Explosive Strength, Extent Flexibility, Far Vision, Finger Dexterity, Flexibility of Closure, Glare Sensitivity, Gross Body Coordination, Gross Body Equilibrium, Hearing Sensitivity, Manual Dexterity, Multilimb Coordination, Night Vision, Perceptual Speed, Peripheral Vision, Rate Control, Reaction Time, Response Orientation, Selective Attention, Sound Localization, Spatial Orientation, Speed of Limb Movement, Stamina, Static Strength, Time Sharing, Trunk Strength, Visual Color Discrimination, Visualization, Wrist-Finger Speed, Building and Construction, Chemistry, Design, Engineering and Technology, Food Production, Mechanical, Physics, Production and Processing, Public Safety and Security, Transportation, Controlling Machines and Processes, Drafting, Laying Out, and Specifying Technical Devices, Parts, and Equipment, Estimating the Quantifiable Characteristics of Products, Events, or Information, Handling and Moving Objects, Inspecting Equipment, Structures, or Material, Monitor Processes, Materials, or Surroundings, Operating Vehicles, Mechanized Devices, or Equipment, Performing General Physical Activities, Repairing and Maintaining Electronic Equipment, Repairing and Maintaining Mechanical Equipment

Table C.1: The skills comprising each skill community on the Skillscape.

C.2 Skill Complementarity propensities and clusters

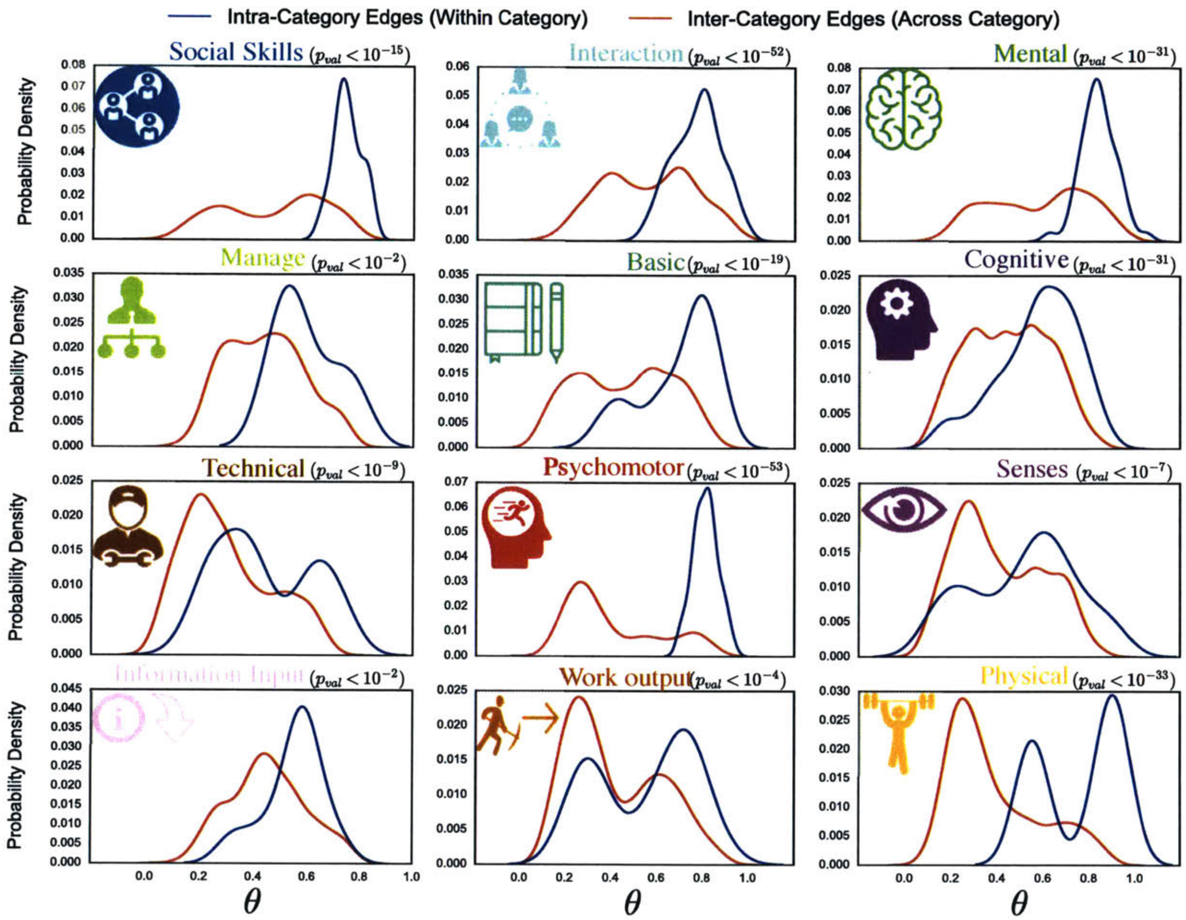


Figure C-5: The Skillscape network respects skill categorization from experts. For each O*NET skill category, we measure the distribution of θ's for pairs of skills within a category (blue) and compare to the distribution of θ's for each edge connecting a skill within the category to a skill outside of the category (red). The complementarity for skills within a category is significantly stronger according the the KS statistic (title) then the complementarity for inter-category pairs of skills.

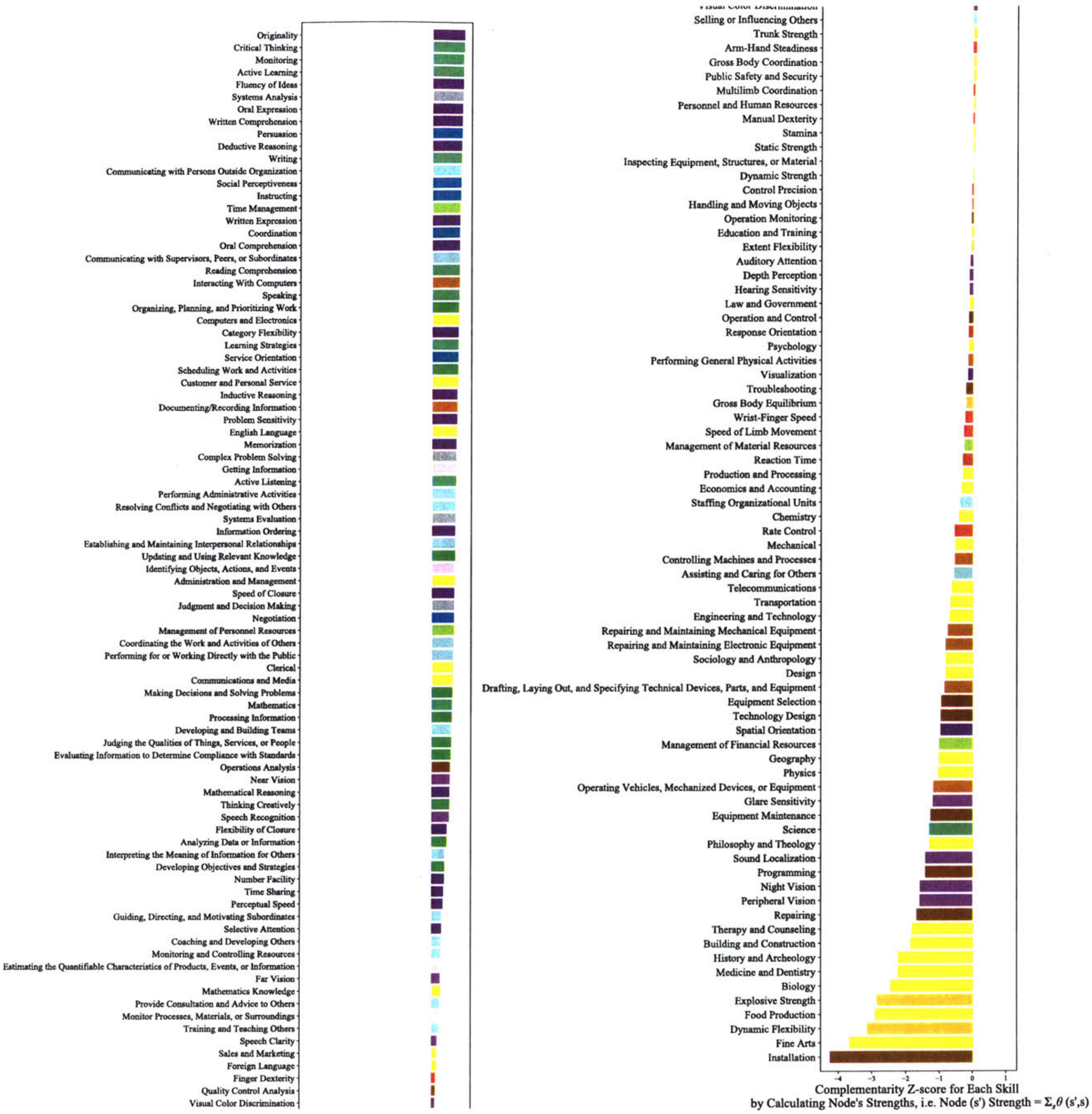


Figure C-6: Complementarity scores for every individual Skill (node in the network). That is, the Z-score of each node's strength (sum of it's edges, or "complementarity weights" θ). Color represents skill category.

C.3 How Educational Requirements Relate to Skill Requirements for Occupations

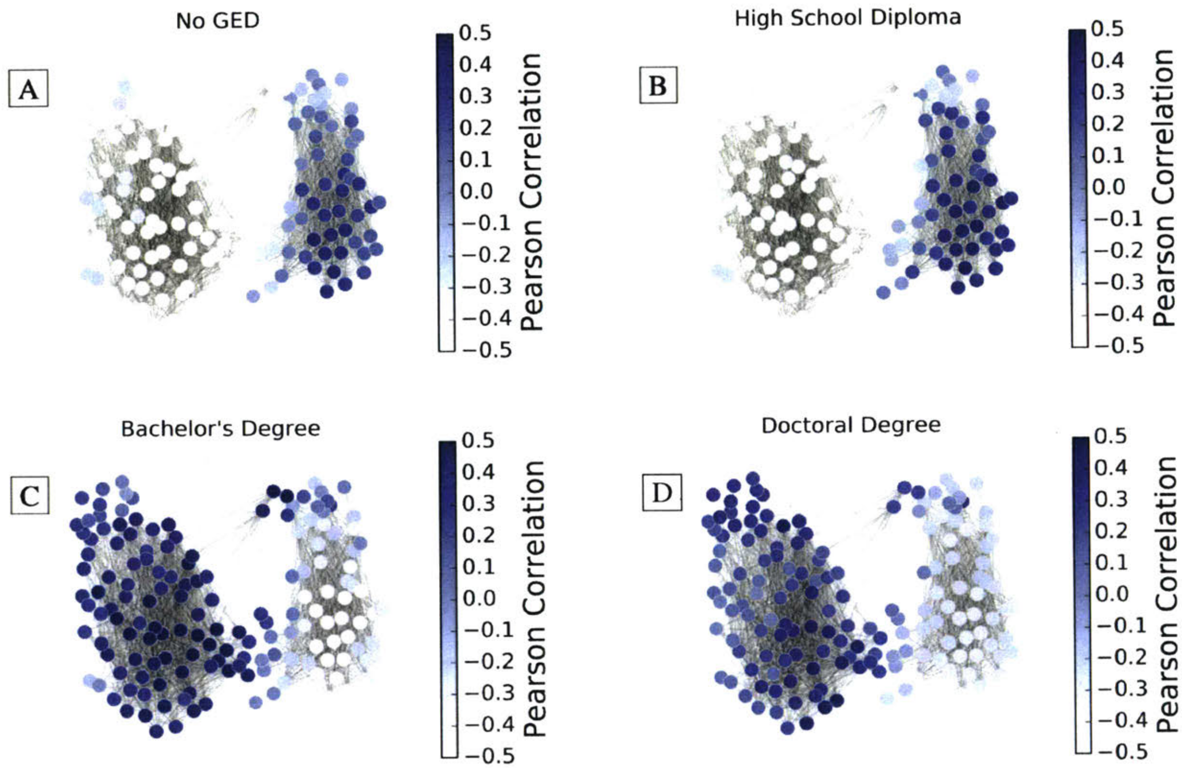


Figure C-7: The skill requirements of an occupation indicate the education required. In each panel, we plot the Skillscape network thresholding edges with $\theta > 0.6$. Nodes (or skills) are colored according to the Pearson correlation between $onet(j,s)$ and the proportion of workers of each occupation with a given degree (title).

C.4 Validating Skill Polarization

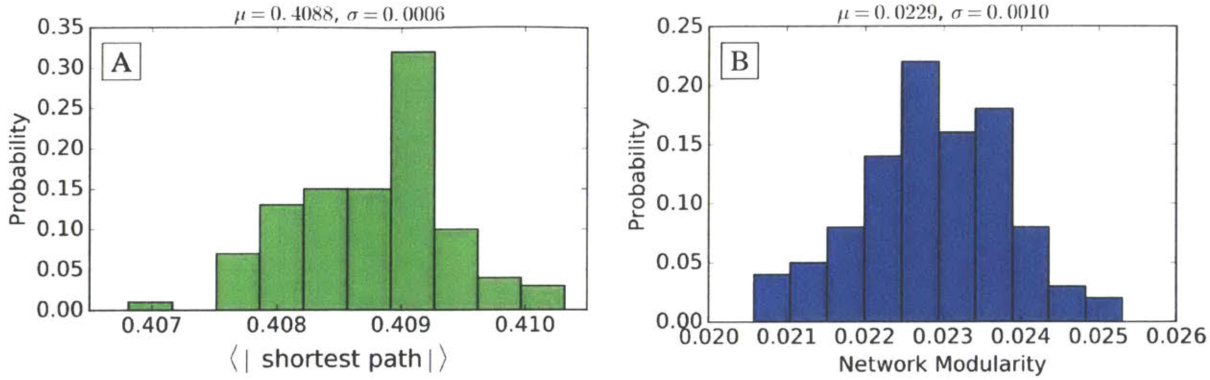


Figure C-8: Testing the significance of Skillscape polarization. We perform 100 trials of randomly shuffling Skillscape edge weights (i.e. θ for modularity, and $1-\theta$ for shortest path) and measuring the resulting **(A)** average shortest path length and **(B)** network modularity. Empirically, the average shortest path is 0.567 and the network modularity is 0.159; both clear outliers given the values resulting from randomization.

C.5 Projecting Urban Workforces onto the Skillscape

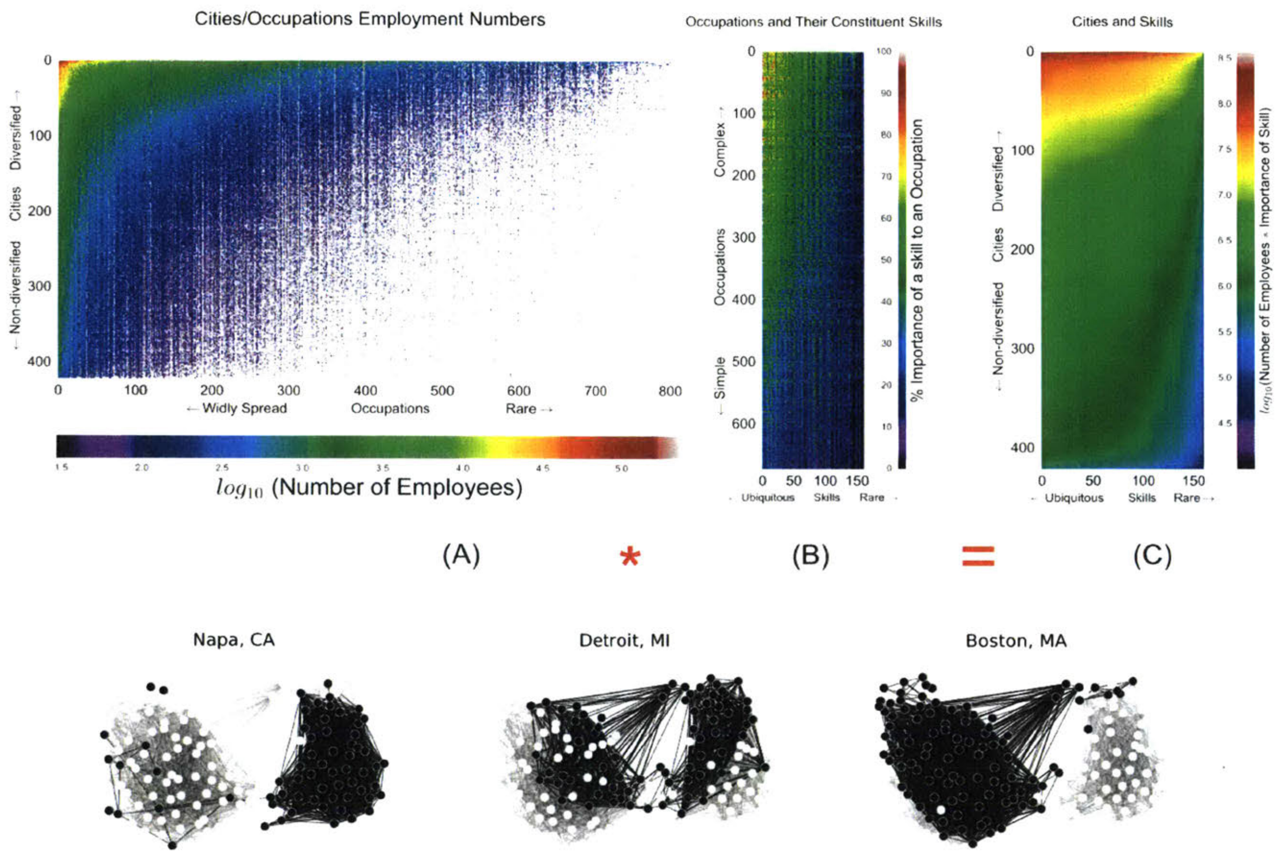


Figure C-9: Identifying the skill sets of urban workforces. Using 2014 data to construct a matrix relating skills to cities from the dot product of the **(A)** city-job matrix with the **(B)** job-skill matrix to produce a **(C)** city-skill matrix. Note that (B) is the same matrix as Fig. C-1A only reordered. Logarithmic values are presented in (A) and (C) only for visual appeal. Finally, we project Napa, Detroit, and Boston onto the Skillscape based on their effectively used skills.

Combining the relationships between occupations and skills according to the O*NET database with the distributions of occupations in cities according to BLS captures how strongly each urban workforce relies on each skill. Denoting the number of workers in city c with occupation j using $bls(c,j)$, we combine the two data sets according to

$$CS(c,s) = \sum_{j \in J} bls(c,j) \cdot onet(j,s), \tag{C.1}$$

where $CS(c,s)$ denotes city c's reliance on workplace skill s.

As with the raw O*NET data, certain jobs and certain skills are ubiquitous across many cities. We again apply RCA to $CS(c,s)$ to identify which skills are effectively used in each city according to

$$rca(c,s) = \frac{CS(c,s)/\sum_{s\in S} CS(c,s)}{\sum_{c\in C} CS(c,s)/\sum_{c\in C,s\in S} CS(c,s)}. \tag{C.2}$$

Similar to occupations, $rca(c,s) > 1$ indicates the effective use of s in c. Fig.C-10 demonstrates the resulting effectively used skills in some example cities.

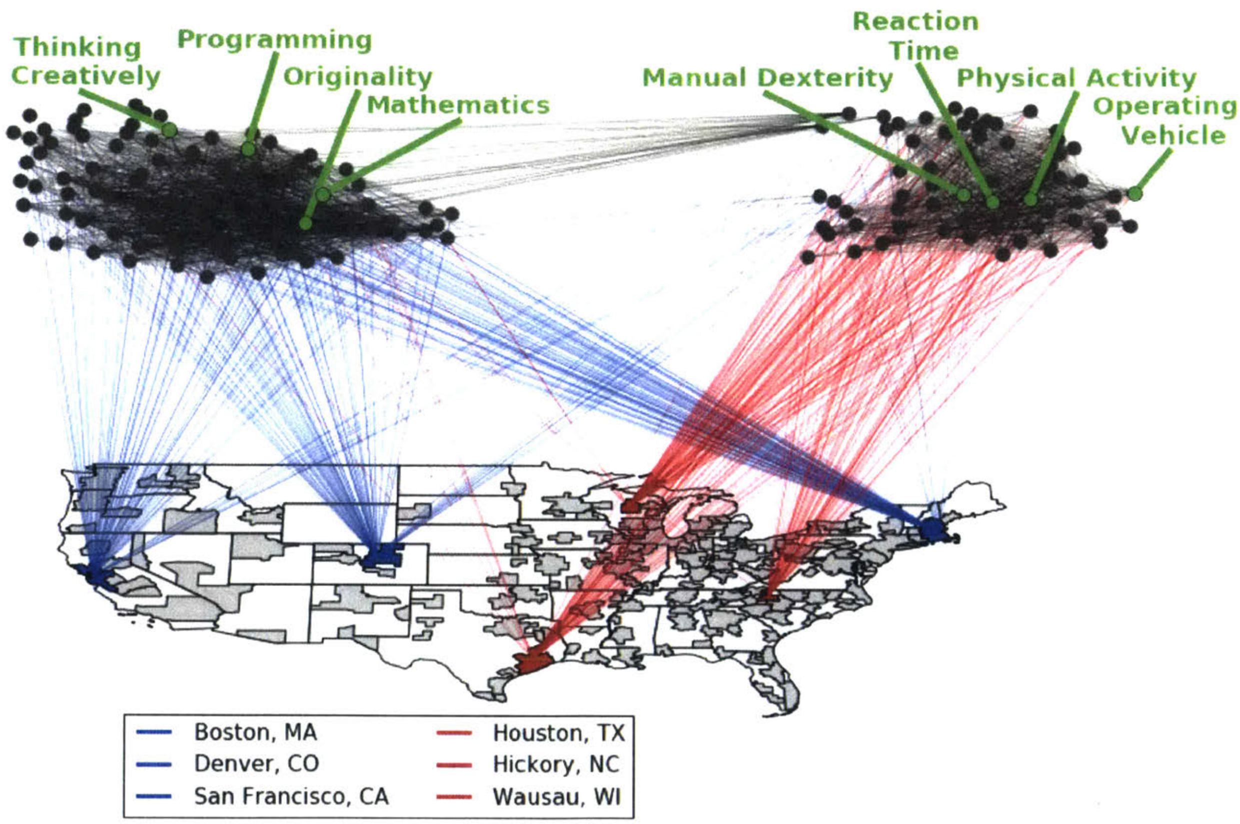


Figure C-10: Example cities projected onto the Skillscape according to effective use of skills. Blue cities rely on socio-cognitive skills, while red cities rely on sensory-physical skills.

C.6 Predicting Economic Well-Being with Socio-Cognitive Skills

Label	Industry Description
Occupation 1	Management, Business, & Financial Occupations
Occupation 2	Computer, Engineering, & Science Occupations
Occupation 3	Education, Legal, Community Service, & Arts Occupations
Occupation 4	Healthcare Practitioners and Technical Occupations
Occupation 5	Service Occupations
Occupation 6	Sales & Office Occupations
Occupation 7	Natural Resources, Construction, & Maintenance Occupations
Occupation 8	Production, Transportation, & Material Moving Occupations

Table C.2: Descriptions of each occupation type indicator variable used in regression models. For each occupation, the indicator variable is 1 if and only if the occupation SOC code belongs to that occupation category. Each occupation belongs to exactly one occupation category.

C.6.1 Predicting Annual Wages of Occupations

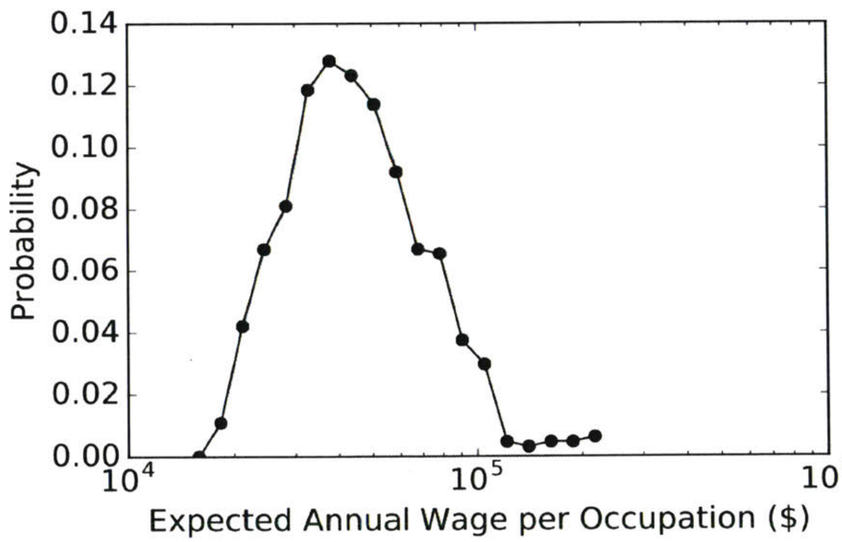


Figure C-11: The distribution of expected annual wages across occupations.

	Dependent Variable Annual Wages of Occupations				
Variable	Model 1	Model 2	Model 3	Model 4	Model 5
$cognitive_j$	0.387***		0.403***		0.372***
Occupation 1		0.490***	0.050		
Occupation 2		0.663***	0.544***	0.970***	0.648***
Occupation 3		0.112	−0.203*		
Occupation 4		1.320***	1.251***	1.627***	1.351***
Occupation 5		−0.785***	−0.674***	−0.478***	−0.588***
Occupation 6		−0.575***	−0.837***	−0.268*	−0.722***
Occupation 7		−0.483***	0.036		
Occupation 8		−0.582***	−0.108		
Intercept	0.000	0.160***	0.059*	−0.147***	−0.027
R^2	0.150	0.386	0.429	0.312	0.424
adj. R^2	0.149	0.380	0.421	0.308	0.419
$p_{val} < 0.1^{*}$, $p_{val} < 0.01^{**}$, $p_{val} < 0.001^{***}$					

Table C.3: Linear regression using standardized $cognitive_j$ for each occupation and occupation type indicator variables.

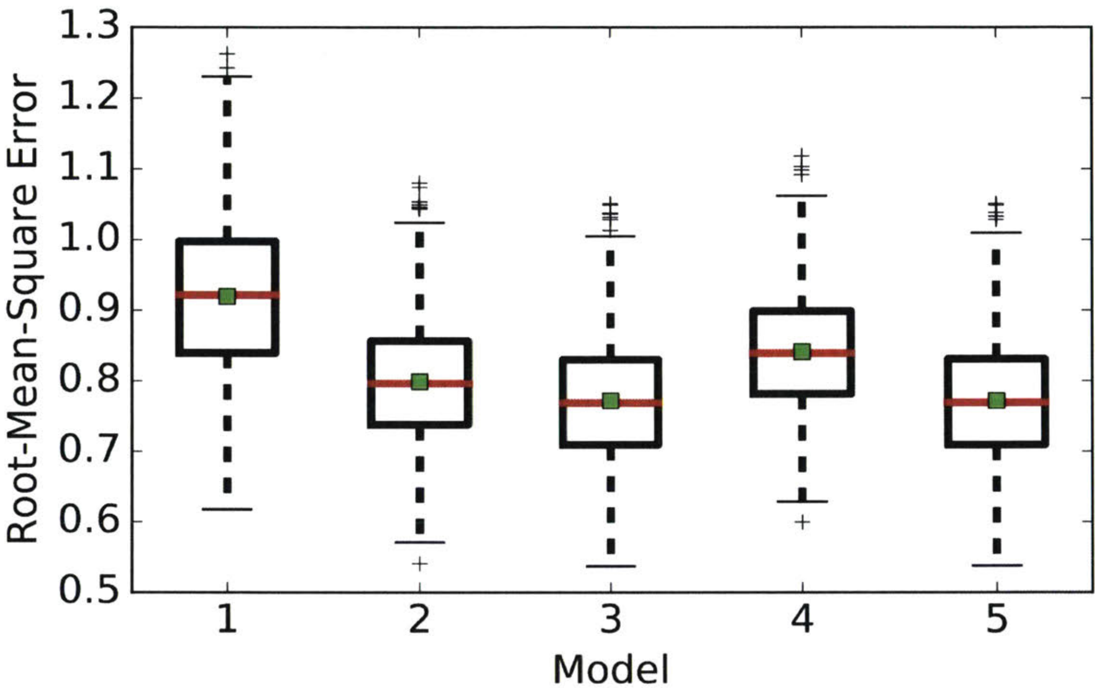


Figure C-12: Out of sample testing of model performance from Table C.3. For each model, 1,000 trials are run where 75% of the data is randomly selected as training data and the remaining 25% of data is used as validation. The distribution root-mean-square errors for each model is reported. Medians are represented by a red line, while the mean error is represented by the green square.

Variable	Dependent Variable Annual Wages of Occupations Model 1	Model 2	Model 3
Intercept	0.000	−0.000	0.000
$cognitive_j$	0.387^{***}		0.355^{***}
No B.D. Employment		-0.264^{***}	-0.234^{***}
B.D. Employment		0.216^{***}	0.094^{*}
R^2	0.150	0.090	0.203
adj. R^2	0.149	0.087	0.199

$p_{val} < 0.1^{*}$, $p_{val} < 0.01^{**}$, $p_{val} < 0.001^{***}$

Table C.4: Linear regression using $cognitive_j$ and employment in each occupation with a bachelor's degree (denoted B.D. Employment) and without a bachelor's degree (denoted No B.D. Employment). Each variable has been standardized. Employment by level of education for each occupation is taken from O*NET data.

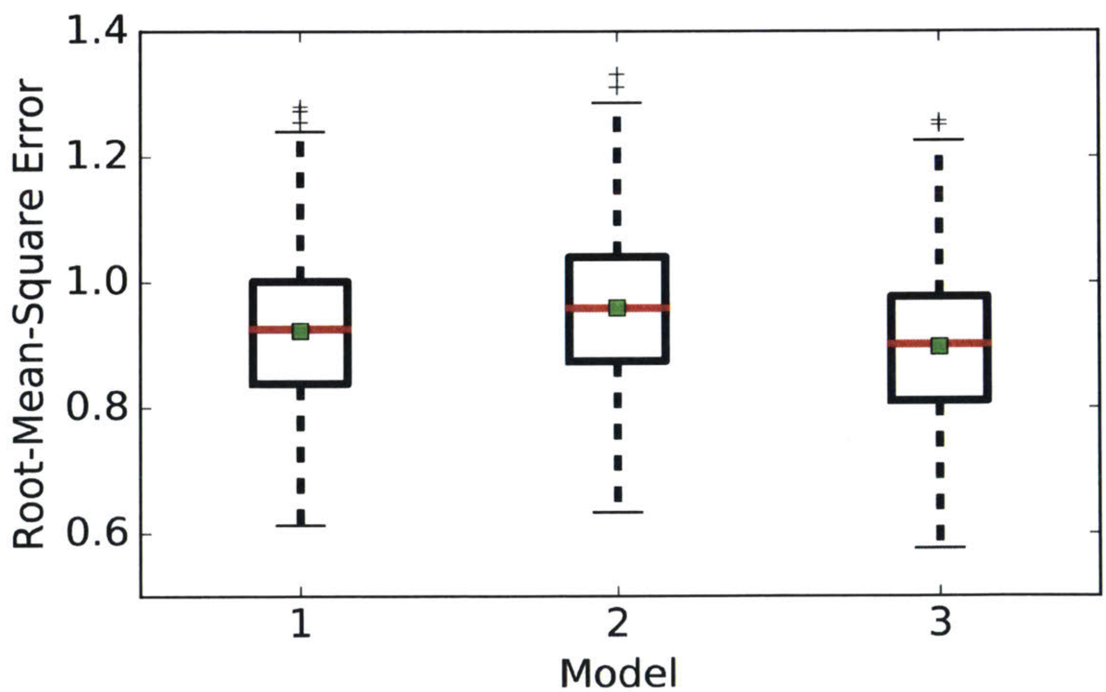


Figure C-13: Out of sample testing of model performance from Table C.4. For each model, 1,000 trials are run where 75% of the data is randomly selected as training data and the remaining 25% of data is used as validation. The distribution root-mean-square errors for each model is reported. Medians are represented by a red line, while the mean error is represented by the green square.

C.6.2 Predicting Median Household Income of Cities

	Dependent Variable Median Household Income of Cities				
Variable	Model 1	Model 2	Model 3	Model 4	Model 5
Intercept	0.000	0.000	−0.000	0.000	−0.000
$cognitive_c$	0.216***		−0.162		−0.147*
Occupation 1		0.354***	0.402***	0.338***	0.407***
Occupation 2		0.229**	0.247**	0.233***	0.287***
Occupation 3		−0.076	−0.040		
Occupation 4		−0.084	−0.066		
Occupation 5		−0.005	−0.013		
Occupation 6		−0.248***	−0.229**	−0.240***	−0.181**
Occupation 7		0.118	0.053		
Occupation 8		−0.089	−0.165		
R^2	0.046	0.456	0.458	0.424	0.435
adj. R^2	0.043	0.439	0.440	0.417	0.426
$p_{val} < 0.1^{*}$, $p_{val} < 0.01^{**}$, $p_{val} < 0.001^{***}$					

Table C.5: Linear regression using standardized $cognitive_c$ for each city and employment in that city of each occupation type. All variables have been standardized.

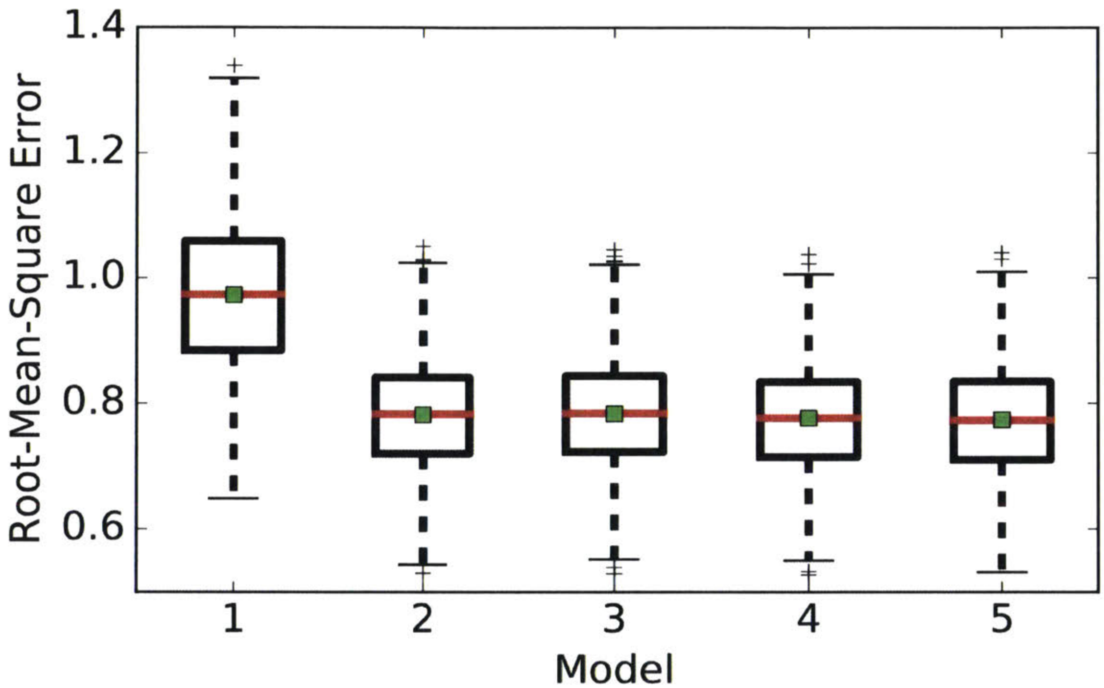


Figure C-14: Out of sample testing of model performance from Table C.5. For each model, 1,000 trials are run where 75% of the data is randomly selected as training data and the remaining 25% of data is used as validation. The distribution root-mean-square errors for each model is reported. Medians are represented by a red line, while the mean error is represented by the green square.

	Dependent Variable Median Household Income of Cities				
Variable	Model 1	Model 2	Model 3	Model 4	Model 5
Intercept	0.000	0.000	−0.000	0.000	−0.000
$cognitive_c$	0.216***		−0.278***		−0.266***
No GED		−0.048	0.007		
H.S. Diploma		−0.340***	−0.402***	−0.351***	−0.423***
Associate's Degree		−0.258***	−0.175**	−0.257***	−0.183**
Bachelor's Degree		0.317***	0.436***	0.318***	0.426***
Master's Degree		−0.061	−0.001		
Doctoral Degree		0.047	0.056		
R^2	0.046	0.351	0.381	0.348	0.378
adj. R^2	0.043	0.337	0.364	0.341	0.369
$p_{val} < 0.1^{*}$, $p_{val} < 0.01^{**}$, $p_{val} < 0.001^{***}$					

Table C.6: Linear regression using $cognitive_c$ and education variables. Education variables represent the employment in each city by highest educational degree attainment. All variables have been standardized.

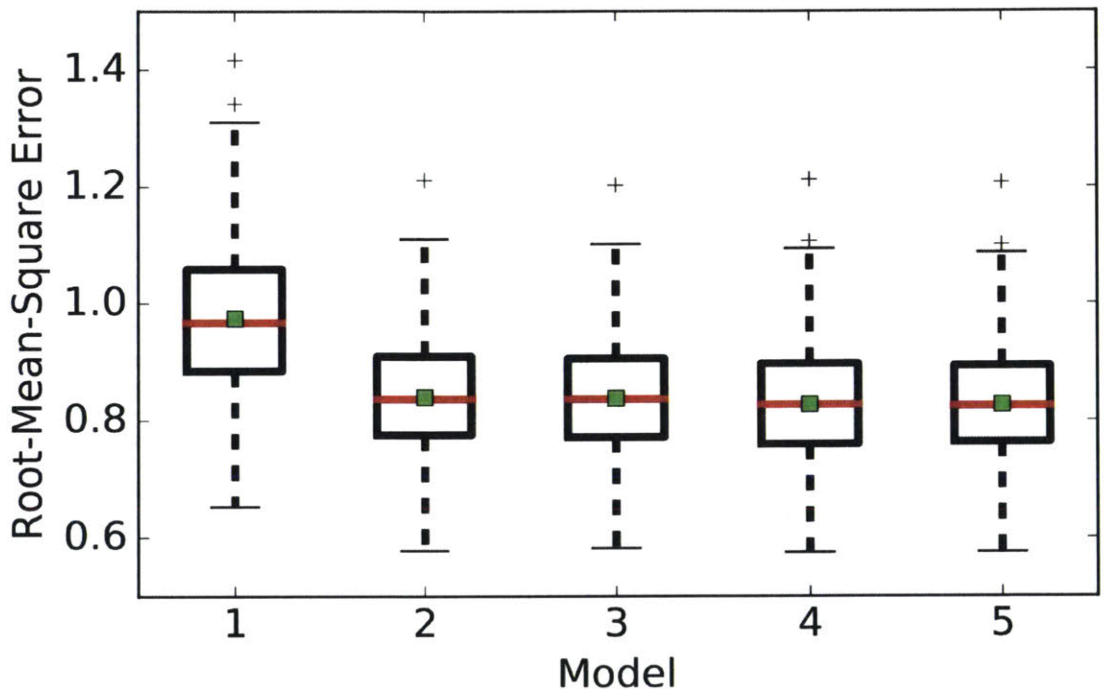


Figure C-15: Out of sample testing of model performance from Table C.6. For each model, 1,000 trials are run where 75% of the data is randomly selected as training data and the remaining 25% of data is used as validation. The distribution root-mean-square errors for each model is reported. Medians are represented by a red line, while the mean error is represented by the green square.

C.7 Using Skillscape Proximity to Predict Labor Dynamics

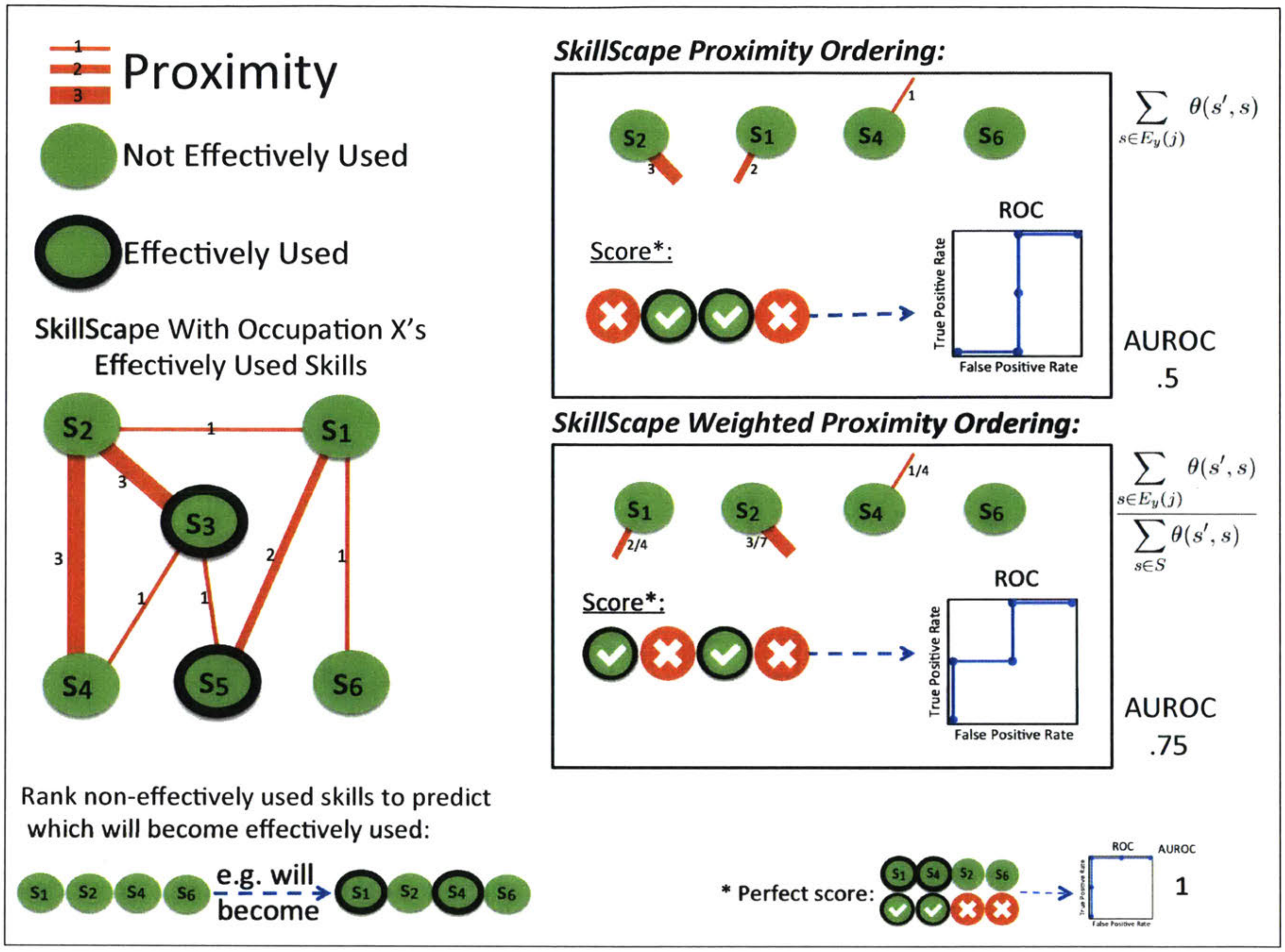


Figure C-16: A cartoon example of AUROC calculation.

Let $\lambda \in [0,1]$ denote the RCA upper limit for skills to consider. Given the set of skills which are effectively used by occupation j in year y (denoted $E_y(j) = \{s \in S \mid rca_y(j,s) > 1\}$), we consider the unweighted proximity of each remaining skill (denoted $\bar{E}_{y,\lambda}(j) = \{s \in S \setminus E_y(j) \mid rca_y(j,s) < \lambda\}$) to the skills in $E_y(j)$. For each $s' \in \bar{E}_{y,\lambda}(j)$, we calculate

$$\overline{proximity}(s') = \sum_{s\in E_y(j)} \theta(s',s). \tag{C.3}$$

We control for occupations relying on ubiquitous skills by calculating weighted Skillscape

A

B

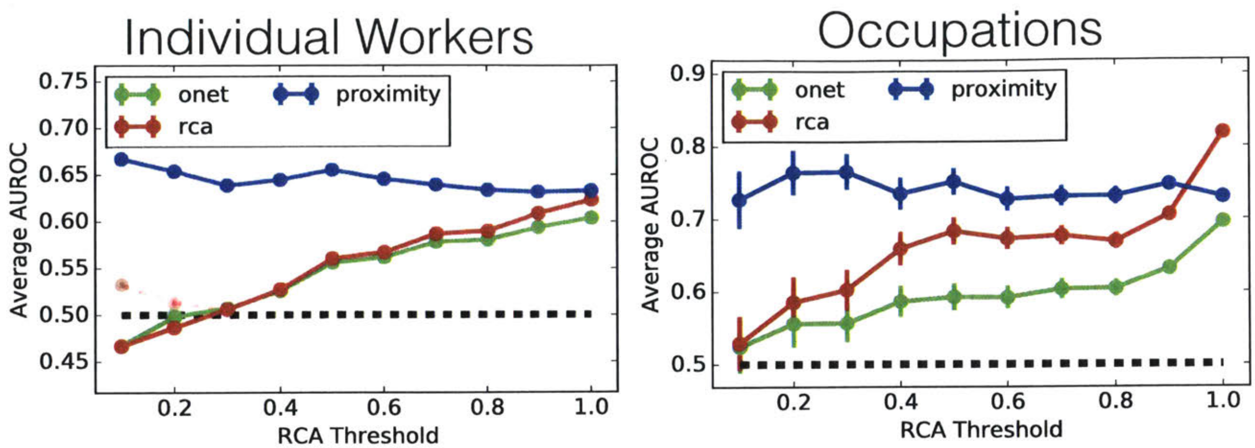


Figure C-17: Worker mobility and occupation redefinition are constrained by skill complementarity and polarization. Each point is the average AUROC for each predictor, using a wide variety of starting RCA threshold values shown on the x-Axis, and standard error bars are provided. **(A)** Skillscape complementarity proximity (blue) better predicts the future (2015) skill requirements of an worker's occupation from their old (2014) occupation requirements, compared to O*NET data (green), and RCA (red). **(B)** Similar to (A), the effectively used skills comprising each occupation in 2015 compared to 2010 are best predicted by Skillscape proximity.

proximity (hereafter "proximity") according to

$$proximity(s') = \frac{\sum_{s \in E_y(j)} \theta(s', s)}{\sum_{s \in S} \theta(s', s)} \tag{C.4}$$

Note that analogous calculations can determine Skillscape proximity from urban workforces by considering $rca(c,s)$ instead of $rca(j,s)$.

We rank each skill in $\bar{E}_{y,\lambda}(j)$ according to their proximity to skills in $E_y(j)$, and examine $\bar{E}_{y,\lambda}(j) \cap E_{2015}(j)$. This produces a receiver operating characteristic (ROC) curve from which we calculate the area under the ROC curve (commonly referred to as AUROC). AUROC allows us to assess the predictive power of Skillscape proximity without prescribing a minimum proximity threshold, and allows us to compare several different predictors with potentially different ranges of values. Figure C-16 provides a cartoon explanation of this calculation.

The U.S. Census Bureau's Current Population Survey (CPS) details the occupation transitions of individual workers from occupation j_A in 2014 to occupation j_B in 2015.

We note the skills which are not effectively used by j_A (i.e. $rca_{2014}(j_A, s) \leq 1$) and predict the effectively used skills of j_B (i.e. $rca_{2015}(j_B, s) > 1$). Ordering these skills according to $onet_{2014}(j_A, s)$, $rca_{2014}(j_A, s)$, or Skillscape proximity, produces receiver operating characteristic curves from which the area under the curve (AUROC) represents the ability of each predictor. Recalling that a skill is not effectively used if $rca(j, s) \leq 1$, the problem of predicting which skills will be acquired becomes more difficult as we restrict ourselves to skills with lower $rca(j, s)$ in the year of interest (e.g. consider only $rca_{2014}(j, s) < .9$, see Fig. 17A). Similarly, we use the 2010 and 2015 O*NET data to test Skillscape proximity's ability to predict changes in the constituent skills of each occupation (see Fig. 17B). Skillscape proximity outperforms $rca(j, s)$, $onet(j, s)$, and a random model at predicting skill acquisition of both occupations and individual workers, which suggests that worker mobility and occupational redefinition occur locally on the Skillscape. Additional figures detailing proximity's predictive ability for several subsets of workers are provided in the remainder of this Section.

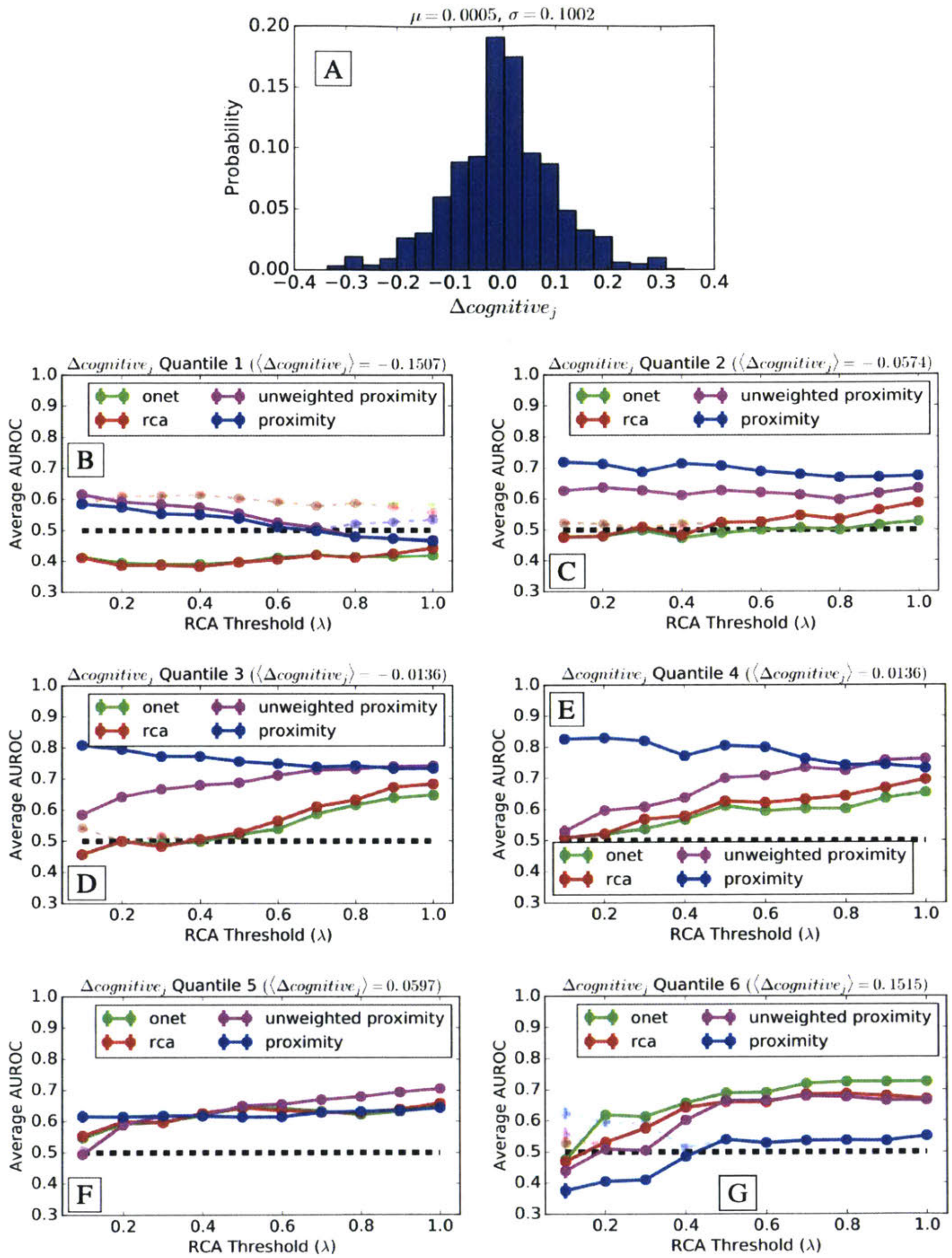


Figure C-18: Predicting changes in cognitive skill fraction of individual workers binning transitions by the magnitude of change. **(A)** The distribution of changes in cognitive skill fraction (denoted $\Delta cognitive_j$) associated with occupation transitions of individual workers from the CPS data. **(B)-(G)** The performance of *onet*, *rca*, and Skillscape proximity for predicting changes in effectively used skills after binning CPS occupation transitions into six quantiles ($N = 900$ transitions per bin) according to the change in cognitive skill fraction.

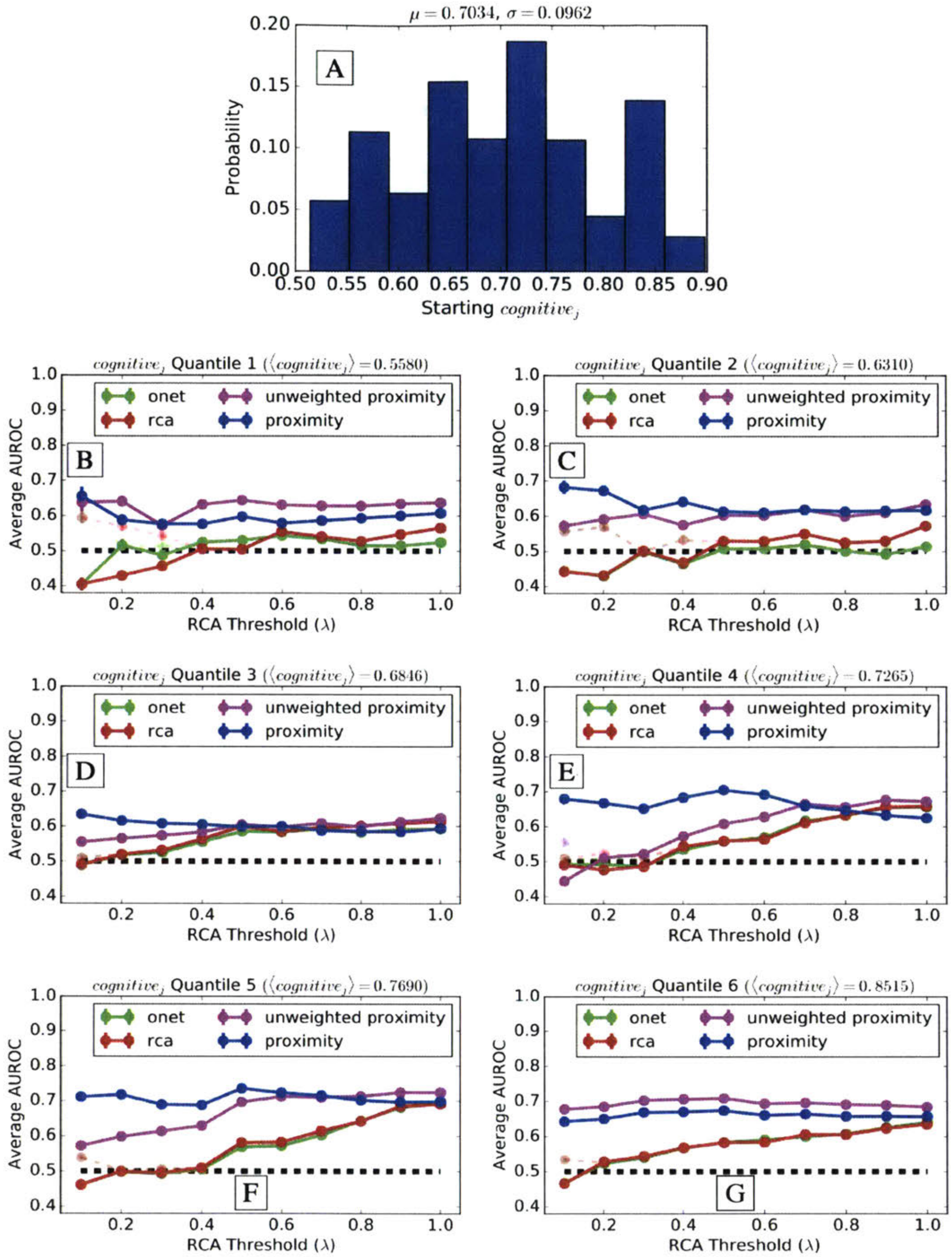


Figure C-19: Predicting changes in cognitive skill fraction of individual workers binning transitions by their starting cognitive skill fraction. **(A)** The distribution of starting cognitive skill fraction associated with occupation transitions of individual workers from the CPS data. **(B)-(G)** The performance of *onet*, *rca*, and Skillscape proximity for predicting changes in effectively used skills after binning CPS occupation transitions into six quantiles ($N = 900$ transitions per bin) according to the cognitive skill fraction for the worker's original occupation.

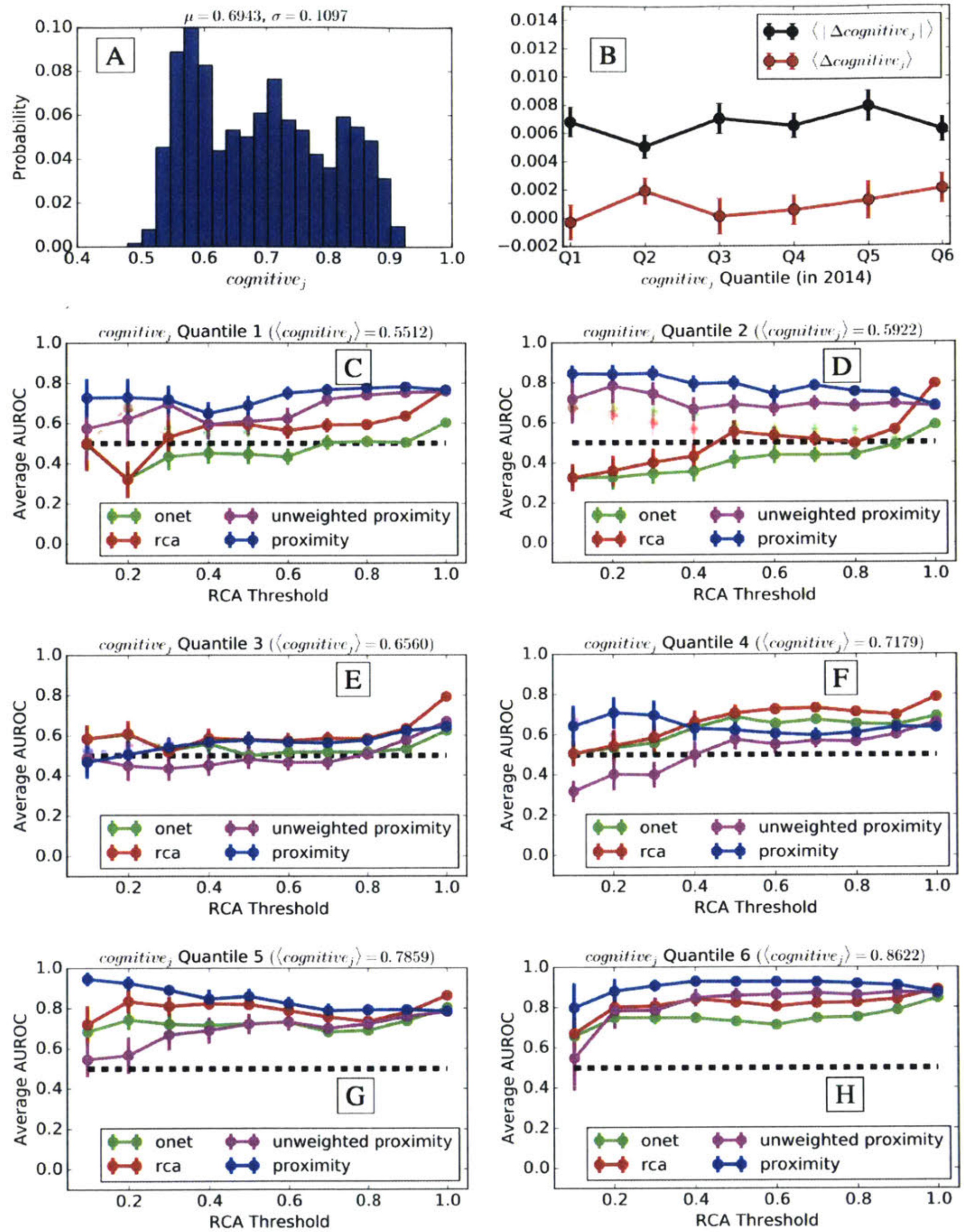


Figure C-20: Predicting changes to the cognitive skill fraction of occupations. **(A)** The distribution of starting cognitive skill fraction associated with each occupation in 2014. **(B)** The average change (red) and average magnitude of change (black) in cognitive skill fraction for occupations from 2014 to 2015 for each quantile of occupations according to $cognitive_j$ in 2014 ($N = 100$ occupations per bin). **(C)-(H)** The performance of *onet*, *rca*, and Skillscape proximity for predicting changes in effectively used skills after binning occupations into six quantiles according to the cognitive skill fraction of the occupation in 2014.

C.7.1 Occupation Transitions of Urban Workforces

Analogous to our investigation of individual worker occupation transitions and occupation redefinition, we ask if changes in the effectively used skills of urban workforces is predicted by complementarity of skills. To this end, we identify the effectively used skills of each city in 2010 and predict from the remaining skills which ones will become effectively used in 2015. Again, we compare *onet*, *rca*, and Skillscape proximity measures while varying the RCA threshold (see Fig. C-21A). The first observation is that the RCA values of skills in cities do not change much from 2010 to 2015 (see Fig. C-21B). This is not to say that urban workforces do not change their constituent workplace skills, but, rather, it may take longer time-scales (i.e. longer than 5-years) for appreciable labor dynamics to present themselves in urban workforces. All the same, Skillscape proximity narrowly outperforms the other measures for most choices of RCA threshold; however, if you allow for the $1-onet$ predictor (indicated by pale green in Fig. C-21A), then this method tends to outperform Skillscape proximity. Although the $1-onet$ method may be reasonably predictive, we suspect that all results presented in Figure C-21 are not conclusive due to the small amounts of observable dynamics in the constituent skills of urban workforces at the five year time-scale.

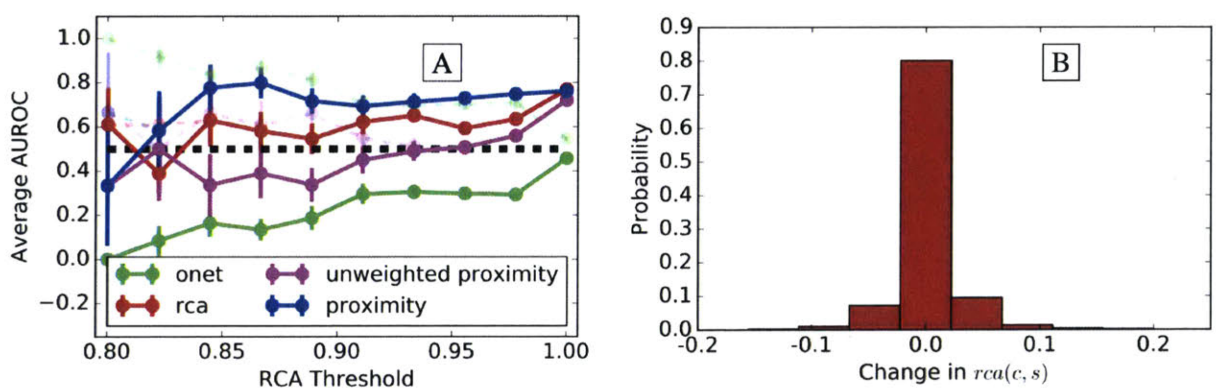


Figure C-21: Predicting the effectively used skills of cities over time. **(A)** Predicting the effectively used skills of urban workforces in 2015 from the effectively used skills of urban workforces in 2010. **(B)** The distribution of changes in $rca(c,s)$ comparing 2010 to 2015.

C.7.2 Worker Occupation Mobility is Constrained by Skill Polarization

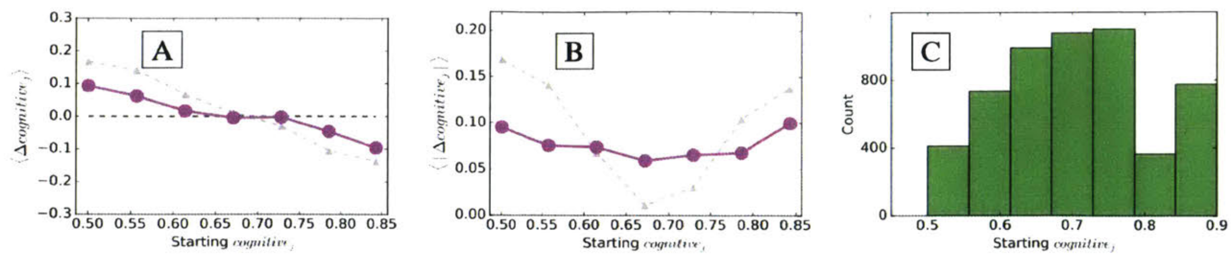


Figure C-22: Workers exhibit greater career mobility when leveraging exclusively socio-cognitive or sensory-physical skills. Binning individual worker occupation transitions into evenly spaced bins according to starting $cognitive_j$, we provide the **(A)** expected change in cognitive skill fraction and **(B)** the expected magnitude of change in cognitive skill fraction resulting from the occupation change. For both (A) & (B), standard error bars are plotted, but are negligible. **(C)** The number of observations in each bin.

Here, we investigate the changes in cognitive skill fraction (i.e. $cognitive_j$) associated with each occupation transition in the CPS dataset. Recall that occupations with higher $cognitive_j$ tend to have higher annual salaries (see Fig. 2B of the main text), but we do not find that workers necessarily increase $cognitive_j$ when they transition between occupations. Clearly, workers do not have complete freedom in selecting their next occupation.

Therefore, the task is to identify feasible null models that might explain the trends we observe in occupation transitions. First, we consider the case where individual workers select their new occupation at random with probability proportional to the number of workers of each occupation according to national labor statistics. In this case, the expected cognitive skill fraction of the newly obtained occupation is $cognitive_j = 0.70$. Figure 3 C&D in the main text demonstrate the average change and the average magnitude of change in cognitive skill fraction after binning occupation transitions into quantiles (i.e. evenly populated bins) in comparison to the null model. Figure C-22 is the analogous plot using evenly spaced bins for the $cognitive_j$ of the original occupation in each occupation transition. In both cases, we find that workers transitioning away from occupations with low cognitive skill fraction tend to transition into occupations with higher $cognitive_j$, and workers transitioning away from occupations with high cognitive skill fraction tend to transition into occupations with lower $cognitive_j$, but these transitions appear to be smaller than we would

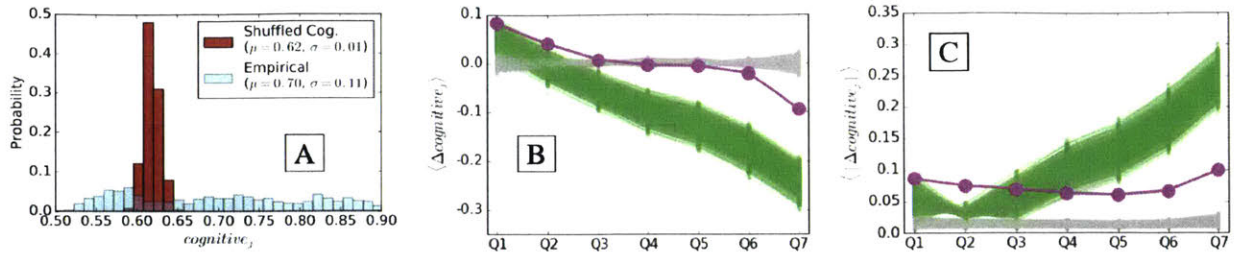


Figure C-23: The effects of randomly selecting "cognitive skills" as a null model alternative to Louvain community detection. For 1,000 trials, we randomly select skills to be considered as cognitive skills (denoted C^*) and calculate the updated cogntive skill fraction, $cognitive^*_j$, for each occupation. **(A)** The $cognitive^*_j$ of one randomization (red) is provided for comparison to the empirical distribution of $cognitive_j$ (blue). For each cognitive skill randomization, we calculate the **(B)** expected change in $cognitive_j$ and **(C)** the expected change in $|cognitive_j|$ using C^* for the cognitive skill fraction of just new occupations (green) and for both new and old occupations (grey) from CPS data, and compare to the empirical results (purple).

expect given random occupation selection. On the other hand, we find that the magnitude of change in cognitive skill fraction (i.e. $\langle| \Delta cognitive_j |\rangle$. See Fig. C-22B) is maximized relative to the null model for workers transitioning away from middle-quantile occupations.

Combined, these results suggest constraints on worker mobility that we would expect given the polarized skill network. Specifically, we expect workers transitioning away from low (high) $cognitive_j$ occupations to obtain occupations with higher (lower) $cognitive_j$ just by chance due to the bounded nature of $cognitive_j$. However, our results demonstrate that the mobility for low and high $cognitive_j$ workers is actually more constrained than we would expect from pure randomness. This constraint is a result of the divide between sensory-physical skills and socio-cognitive skills we observe in the skill network. Conversely, workers transitioning away from occupations relying on skills that straddle this skills divide (i.e. middle $cognitive_j$ occupations) have greater ability to transition to either low or high $cognitive_j$ occupations because of their privileged starting location according to the Skillscape structure. Thus, we conclude that occupation transitions are not achieved through randomly selecting a worker's next occupation, but, rather, these transitions are constrained in a way that makes sense given the Skillscape structure

It remains to show that our results about occupation transitions are not an artifact of how cognitive skills are selected empirically. We carry out 1,000 trials of randomly selecting skills to be "cognitive skills", and we measure the change in cognitive skill fraction when the new occupation's $cognitive_j$ is calculated from the randomized skill selection and when both the old and new occupations' $cognitive_j$ are calculated from the randomized skills. Figure C-23A demonstrates that the resulting distribution of $cognitive_j$ across occupations after randomization is very narrow by comparison to the empirical distribution of $cognitive_j$. Also, Fig. C-23 B&C demonstrate that the changes in $cognitive_j$ and $|cognitive_j|$ that we observe empirically for occupation transitions is not well-captured by this null model either. These results suggest that the selection of cognitive skills is not a result of randomness and reveals something intrinsic about how individual workers move about the Skillscape. We conclude that the Skillscape and $cognitive_j$ captured something meaningful about workers' abilities and their freedom to change occupations.

C.7.3 The Trimodal Employment Distribution over $cognitive_j$ is Robust

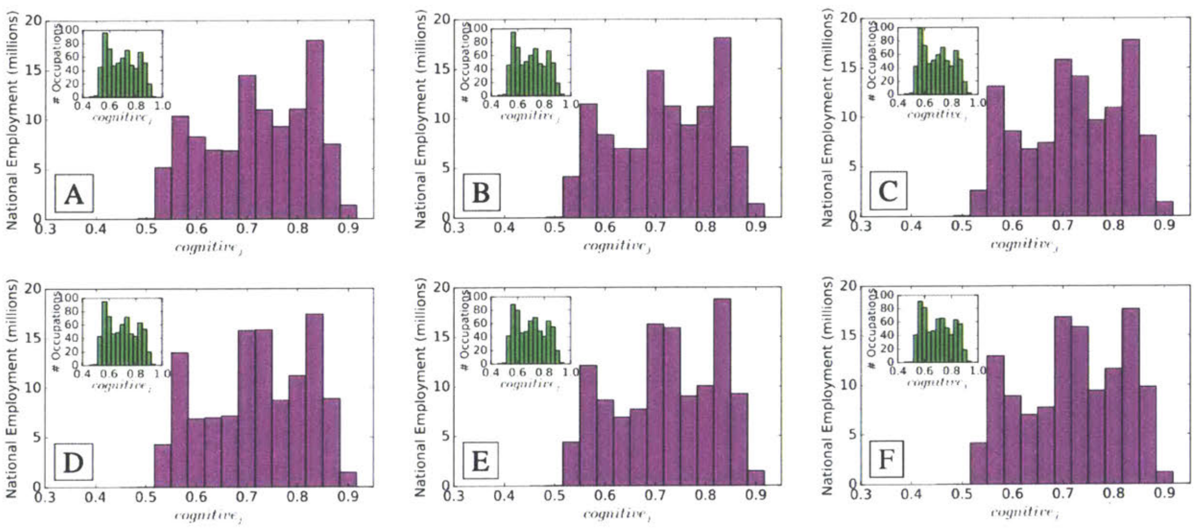


Figure C-24: The distribution of national employment, and of individual occupations as an inset, after binning by $cognitive_j$ in **(A)** 2010, **(B)** 2011, **(C)** 2012, **(D)** 2013, **(E)** 2014, and **(F)** 2015.

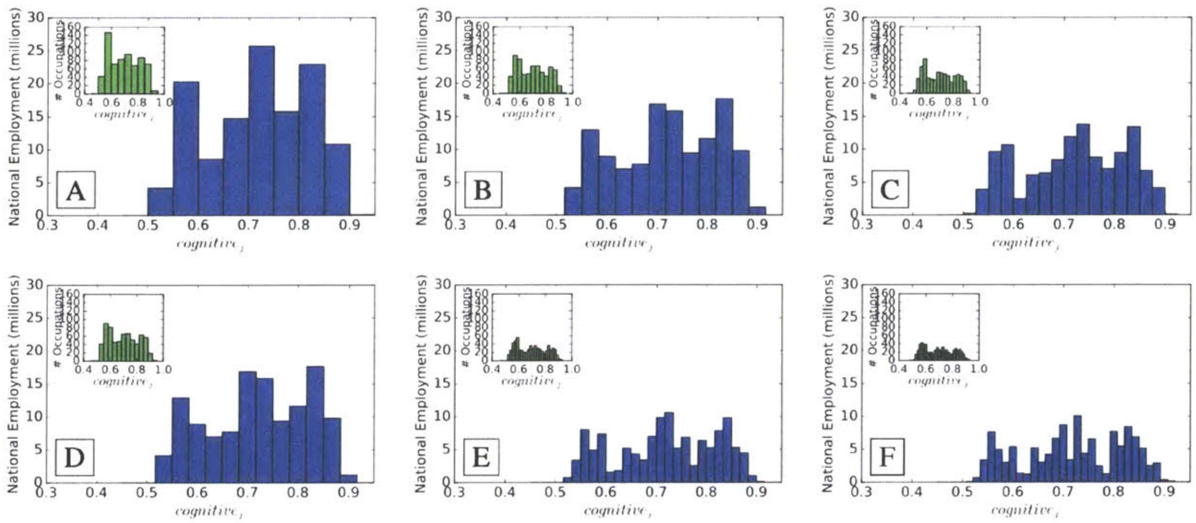


Figure C-25: The distribution of national employment in 2015, and of individual occupations as an inset, after binning by $cognitivce_j$ while varying the number of bins between **(A)** 10, **(B)** 15, **(C)** 20, **(D)** 25, **(E)** 30, and **(F)** 35 bins.

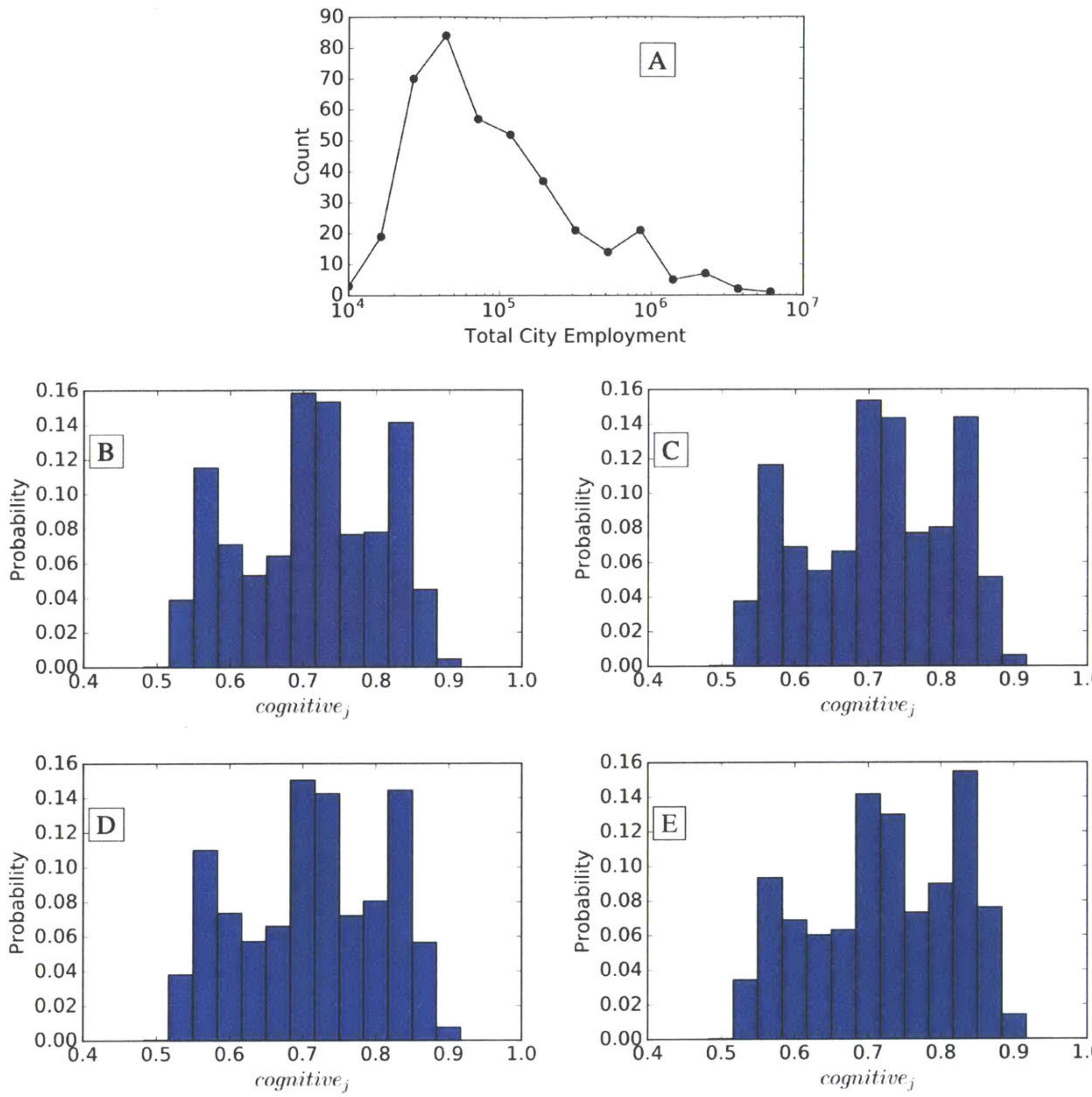


Figure C-26: Binning employment according to cognitive skill fraction reveals a trimodal distribution across cities of all sizes. **(A)** The distribution of total employment in cites (i.e. metropolitan statistical areas). The distribution of city-level employment binned by $cognitive_j$ after dividing cities into quartiles according to total employment in the city: **(B)** total employment between 7,000 and 45,000, **(C)** total employment between 45,000 and 84,000, **(D)** total employment between 84,000 and 2,000,000, and **(E)** total employment between 2,000,000 and 8,900,000.

C.7.4 Bar plots for main paper figure 3 including 95% confidence intervals.

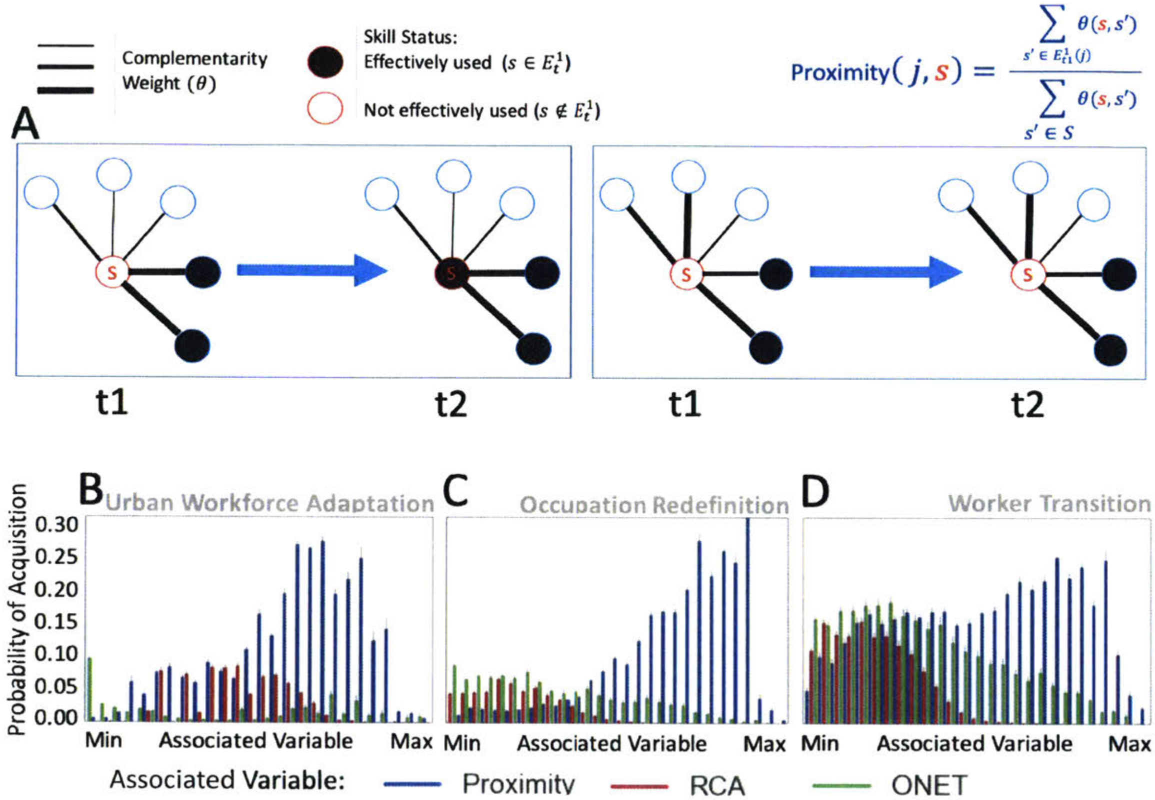


Figure C-27: Skill proximity predicts skill acquisition for individual workers transitioning between occupations, for the skill requirements of occupations, and for labor markets of cities. Bar plot figure including 95% confidence intervals. **(A)** Example demonstrating Skillscape proximity ($proximity(j,s)$) as a means to asses the strength of connection between effectively used skills and other skills. Our model suggests that skill s will be an occupation j if has high proximity to the occupation's effectively used skills. **(B)** Skills with high proximity to the effectively used skills of an occupation in 2010 are more likely to be effectively used by that occupation in 2015. **(C)** Skills with high proximity to the effectively used skill of an urban labor market in 2010 are more likely to be effectively used by that work force in 2015. **(D)** The effectively used skills of a workers occupation in 2014 are more likely to be effectively used by the workers next occupation in 2015.

Appendix D

Hidden constraints on career mobility: How workplace skills determine a worker's next move

D.1 Constructing the Job Network

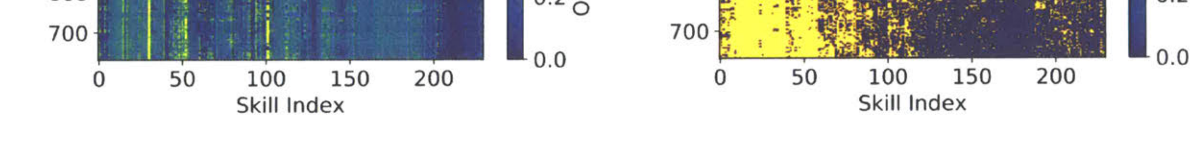


Figure D-1: **(A)** The matrix of O*NET skill values. **(B)** The matrix of characteristic skills of each occupation after normalizing according to revealed comparative advantage. Here, we use $I(j,s) = 1$ if $rca_{j,s} > 1$ and $I(j,s) = 0$ otherwise as an indicator function for characteristic skills.

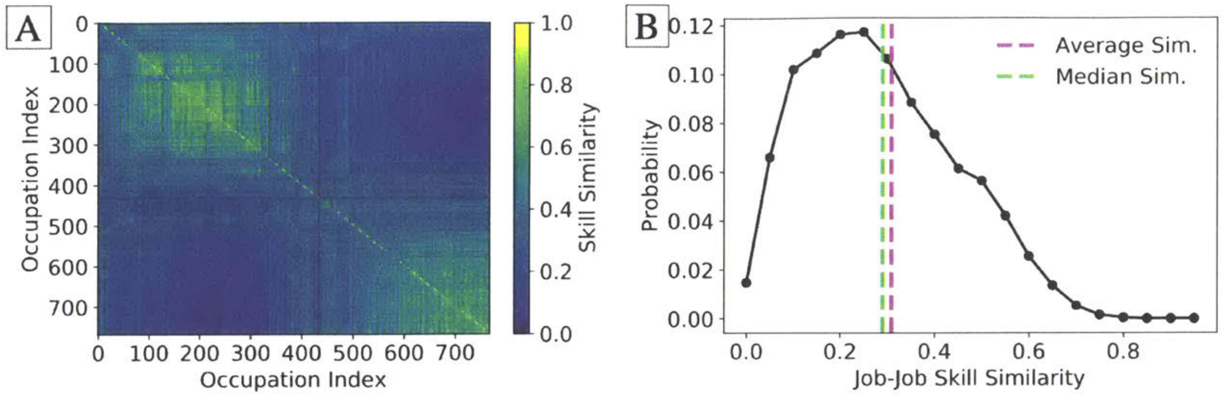


Figure D-2: **(A)** The skill similarity similarity matrix calculated from Jaccard similarity scores. This matrix is equivalent to the job network discussed in the main text. **(B)** The distribution of skill similarity scores.

D.2 Testing Job Network Modularity

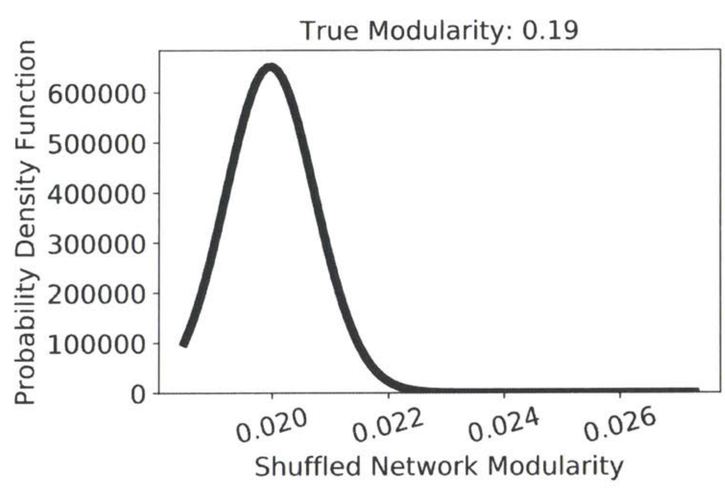


Figure D-3: The distribution of network modularity scores resulting from 100 test networks each constructed by randomly permuting skill similarity scores. The empirically observed job network modularity score of 0.19 is significantly greater than the randomized network scores.

D.3 Tracking Career Mobility through the Community Population Survey

The U.S. Census Bureau and BLS produces monthly Current Population Survey (CPS) data through a continuous survey process that produces representative samples of the U.S. population. Providing high resolution labor statistics is one of the primary goals of CPS, and, in particular, CPS records changes in occupations of survey participants over a 1.5 year period for which that participant is an active contributor to the survey. For our purpose, we are interested only in participants who reported one occupation when they were first surveyed in 2014 and reported working a different occupation when they were surveyed one year later in 2015. There are several methods for joining different time periods of the CPS data [143], so we employed strict merging criteria including participant ID, gender, sex, state of residency, and age to verify the validity of our occupation transitions. The result is a dataset of 5,400 occupation transitions for individual U.S. workers from 2014 to 2015.

D.4 Aggregations of Workplace Skills

While the present study argues for the value of granular workplace skills for studying worker mobility, several existing studies instead utilize broad aggregations of skills (e.g. cognitive versus physical skills) to study labor. These simplifications are necessary for tractable analytical work, but has the potential to stymie empirical forecasting of worker flows between occupations. Therefore, our analysis of worker flows (i.e. Figure 2 of the main text) compares a model using the granular O*NET taxonomy of skills to alternative baseline models using expertly-derived aggregations of skills.

We measure an occupations reliance on an aggregated skill category C according to

$$reliance_{j,C} = \sum_{s \in C} onet_{j,s}. \tag{D.1}$$

Similar to our granular skills model discussed in the main text, we use the Jaccard coefficient to measure the skill similarity between occupation pairs based on occupations' reliance on the aggregated skill categories. Finally, we use the skill-by-skill summands of the calculation for Jaccard similarity (see main text) in combination with linear regression to produce the final baseline model.

D.4.1 O*NET Task Groups

An alternative simplification of the raw O*NET skills is the O*NET Task Groups, which represent collections of similar work activities. We provide the definitions for these task groups in Table D.1. These task groups have been used to investigate the task connectivity of urban labor markets in relation to employment growth [136].

Task Group	O*NET Skills
Information Input	Getting Information, Monitor Processes, Materials, or Surroundings, Identifying Objects, Actions, and Events, Inspecting Equipment, Structures, or Material, Estimating the Quantifiable Characteristics of Products, Events, or Information
Mental Process	Judging the Qualities of Things, Services, or People, Processing Information, Evaluating Information to Determine Compliance with Standards, Analyzing Data or Information, Making Decisions and Solving Problems, Thinking Creatively, Updating and Using Relevant Knowledge, Developing Objectives and Strategies, Scheduling Work and Activities, Organizing, Planning, and Prioritizing Work
Work Output	Performing General Physical Activities, Handling and Moving Objects, Controlling Machines and Processes, Operating Vehicles, Mechanized Devices, or Equipment, Interacting With Computers, Drafting, Laying Out, and Specifying Technical Devices. Parts. and Equipment, Repairing and Maintaining Mechanical Equipment, Repairing and Maintaining Electronic Equipment, Documenting or Recording Information
Interacting with Others	Interpreting the Meaning of Information for Others, Communicating with Supervisors, Peers, or Subordinates, Communicating with Persons Outside Organization, Establishing and Maintaining Interpersonal Relationships, Assisting and Caring for Others, Selling or Influencing Others, Resolving Conflicts and Negotiating with Others, Performing for or Working Directly with the Public, Coordinating the Work and Activities of Others, Developing and Building Teams, Training and Teaching Others, Guiding, Directing, and Motivating Subordinates, Coaching and Developing Others, Provide Consultation and Advice to Others, Performing Administrative Activities, Staffing Organizational Units, Monitoring and Controlling Resources

Table D.1: The O*NET skills comprising each Task Group.

D.4.2 Cognitive/Physical & Routine/Nonroutine Tasks

Autor et al. [24, 25] identify workplace tasks according to their type and how routine the task is. They find that non-routine tasks are becoming increasingly important to workers relative to routine tasks. We provide the definitions for these task groups in Table D.2.

Task Type	**O*NET Skills**
Non-routine Analytic	Mathematical Reasoning, Mathematics, Deductive Reasoning, Number Facility, Physics, Programming
Non-routine Interactive	Design, Administration and Management, Economics and Accounting, Equipment Selection, Estimating the Quantifiable Characteristics of Products, Events, or Information, Importance of Being Exact or Accurate, Management of Financial Resources, Management of Material Resources, Management of Personnel Resources, Organizing, Planning, and Prioritizing Work, Personnel and Human Resources, Quality Control Analysis, Sales and Marketing, Scheduling Work and Activities, Technology Design, Visualization
Routine Cognitive	Consequence of Error, Control Precision, Controlling Machines and Processes, Documenting/Recording Information, Evaluating Information to Determine Compliance with Standards, Inspecting Equipment, Structures, or Material, Operation and Control, Quality Control Analysis
Routine Manual	Finger Dexterity, Manual Dexterity, Arm-Hand Steadiness, Wrist-Finger Speed
Non-routine Manual	Reaction Time, Response Orientation, Cramped Work Space, Awkward Positions, Dynamic Flexibility, Spatial Orientation, Transportation, Coordination

Table D.2: The O*NET skills comprising each Task Type.

D.5 Exploring Inter-city Mobility Patterns With Labor Data

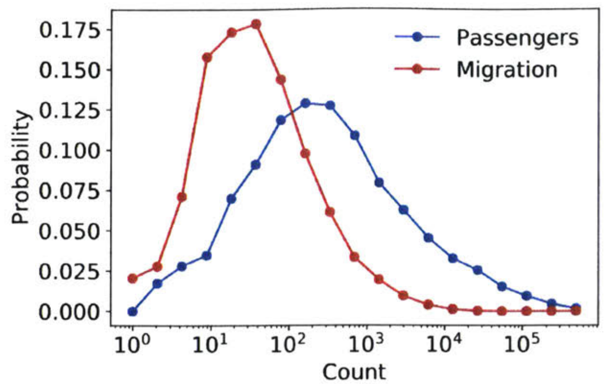


Figure D-4: The probability distributions of the number of enplaned passengers and the number of migrants flowing between city pairs.

According to the 2015 World Migration Report [1], employment opportunities are the leading factor in urban migration. Yet, modern models for urban migration typically study urban migration using the size of cities and the spatial distance between cities combined into a gravity model [35]. This strategy makes no use of employment information or the latent skills of local economies. On the other hand, Figure 1 in the main text demonstrates how the skill distance between occupations can approximate worker migration between those occupations, and, therefore, provides a meaningful notion for the "distance" between a pair of occupations. So, can we extend this notion to approximate the skill distance between city pairs? If so, then we can leverage the employment in cities to predict human mobility between city pairs. In the remainder of this section, we will describe the empirical mobility data in our study and introduce the gravity model for urban migration.

D.5.1 Introduction to Mobility Data

We focus on two data sources each representing a different type of human mobility between cities. First, we use migration between cities according the U.S. Census Bureau 2011-2015 American Community Survey. This data source provides a representative sample of the U.S. population between 2011 and 2015 through a continual survey methodology. Part of this survey asks participants where they resided one year prior to the survey. This information is aggregated by the U.S. Census Bureau to approximate the flows of people migration between combined statistical areas (hereafter referred to as "cities"). In our study, we are interested in the undirected total flow of people between cities (i.e. the sum of migrants from city c to city c' with the number of migrants from city c' to city c). We consider urban migration according to this data to represent costly permanent relocations.

Alternatively, the Bureau of Transportation Statistics, which resides in the U.S. Department of Transportation, reports that nearly 16% of all travel in the U.S. is business related, and they report that around 208,000 business flights occurred in 2016. Unlike Census migration data, business travel is often non-permanent relocation, but does represent a much more common reason for mobility between cities (e.g. there are ~ 1.3 million business trips in the U.S. each day according to Global Business Travel Association). Here, we account for business flights between cities in 2017 using the number of enplaned passengers according to the Origin and Destination Survey from the Bureau of Transportation Statistics. As with the migration data, we consider the undirected total flow of enplaned passengers between city pairs.

D.5.2 Introduction to the Gravity Model for Mobility

As a baseline model for mobility between city pairs, we consider the commonly-used gravity model. The gravity model has been used as a baseline model for migration in several studies, including the study of the migratory effects of climate [28], the study of migration in rural areas [101], and to understand how educated workers find employment opportunities in OECD countries [73]. In physical systems, the gravitational pull between two objects with masses m_1 and m_2 can be modelled according to

$$G = \frac{m_1 \cdot m_2}{distance},$$

where *distance* is the spatial distance between the two masses. Analogously, we can use the size of cities c and c' (in place of mass) to understand the undirected flow of mobility between cities.

Specifically, the gravity model gives us

$$flow_{c,c'} = \alpha_0 \cdot \frac{(employment_c \cdot employment_{c'})^{\alpha_1}}{distance_{c,c'}^{\alpha_2}}, \tag{D.2}$$

where $flow_{c,c'}$ is the undirected total flow of human mobility between cities c and c, $employment_c$ is the total employment in city c (similarly for $employment'_c$), and $distance_{c,c'}$ is the spatial distance between cities c and c'. Each α_i represents a free parameter that can be used to tune model performance. For our purposes, we apply a logarithmic transform to eq. (D.2) to obtain

$$\log_{10}(flow_{c,c'}) = \alpha_0 + \alpha_1 \cdot \log(employment_c \cdot employment_{c'}) - \alpha_2 \cdot \log(distance_{c,c'}). \tag{D.3}$$

After this transformation, all parameters can be approximated using linear regression and we can add additional variables to the model as additional terms in the summation; for instance, we might bolster the predictive performance of the gravity model with city fixed effects (using I_c as a dummy variable for city c) to yield

$$\log_{10}(flow_{c,c'}) = \alpha_0 + \alpha_1 \cdot \log(employment_c \cdot employment_{c'}) - \alpha_2 \cdot \log(distance_{c,c'}) + \sum_{c^* \in Cities} \alpha_{c^*} \cdot I_{c^*}. \tag{D.4}$$

Note that in practice, we calculate $\log(employment_c) + \log(employment_{c'})$ instead of $\log(employment_c \cdot employment_{c'})$ to avoid numerical issues.

D.6 Explaining Differences in City's Job Tightness

Since job tightness bolsters models for spatial mobility between cities, how can job tightness explain differences in a city's connection to the rest of the economy? We approximate this by a city's average tightness given by

$$\begin{aligned} \widehat{tightness}(c) &= \langle tightness(c,c') \rangle_{c' \in Cities} \\ &= \frac{1}{|Cities|} \sum_{c' \in Cities} tightness(c,c'). \end{aligned} \tag{D.5}$$

As seen in Figure 4A of the main text, large cities tend to be more tightly connected to other cities—perhaps adding to explanations for large cities' privileged role as centers for economic productivity and innovation.

Given two cities c and c' we can quantify their difference in average job tightness as

$$
\begin{aligned}
\Delta Tightness(c,c') &= \widehat{tightness}(c) - \widehat{tightness}(c') \\
&= \frac{1}{|Cities|} \sum_{c^{*}*\in Cities} \left[\sum_{j,j'\in J^2} \frac{skillsim(j,j') \cdot \big(I(c,j)+I(c^{*},j)\big)}{2 \cdot \sum_{i,i'\in J^2} skillsim(i,i')} \right. \\
&\quad \left. - \sum_{j,j'\in J^2} \frac{skillsim(j,j') \cdot \big(I(c',j)+I(c^{*},j)\big)}{2 \cdot \sum_{i,i'\in J^2} skillsim(i,i')} \right] \\
&= \frac{1}{|Cities|} \sum_{c^{*}*\in Cities} \left[\sum_{j,j'\in J^2} \frac{skillsim(j,j') \cdot \big(I(c,j)-I(c',j)\big)}{2 \cdot \sum_{i,i'\in J^2} skillsim(i,i')} \right] \\
&= \sum_{j,j'\in J^2} \frac{skillsim(j,j') \cdot \big(I(c,j)-I(c',j)\big)}{2 \cdot \sum_{i,i'\in J^2} skillsim(i,i')} \\
&= \sum_{j,j'\in J^2} \frac{skillsim(j,j') \cdot \Delta Emp(c,c',j)}{2 \cdot \sum_{i,i'\in J^2} skillsim(i,i')}
\end{aligned}
\tag{D.6}
$$

where $\Delta Emp(c,c',j) = I(c,j) - I(c',j)$ captures the differences in characteristic occupations between c and c'.

We can rewrite this result in a few useful ways. First to capture the contribution to difference of each O*NET skill, we examine the summands of

$$
\Delta Tightness(c,c') = \sum_{s\in S} \beta_s \left[\sum_{j,j'\in J^2} \frac{\gamma(j,j',s) \cdot \Delta Emp(c,c',j)}{2 \cdot \sum_{i,i'\in J^2} skillsim(i,i')} \right]. \tag{D.7}
$$

Second, we can instead quantify to contribution of individual occupations to the difference in average tightness by examining the summands of

$$
\begin{aligned}
\Delta Tightness(c,c') &= \sum_{j\in J}\left[\left(\Delta Emp(c,c',j)\right)\left(\frac{\sum\limits_{s\in S}\beta_s\sum\limits_{j'\in J}\gamma(j,j',s)}{2\cdot\sum\limits_{i,i'\in J^2} skillsim(i,i')}\right)\right] \\
&= \sum_{j\in J}\left[\left(\text{employment differences}\right)\left(\text{skill connectivity}\right)\right]
\end{aligned}
\tag{D.8}
$$

In particular, we can decompose each summand in equation (D.8) to understand how an individual occupation contributes to the difference in average job tightness in terms of employment differences and skill connectivity. The sign of each of these terms can be either positive or negative. That is, the occupation may be characteristic of one city and not the other city, and these differences in employment can contribute positively or negatively to the net difference in average tightness depending on the connectivity of the occupation to other occupations and the weight (i.e. regression coefficient β_s) associated with each skill. We can summarize the possible contributions of a single occupation as follows:

- if j is neither characteristic of c nor of c' (i.e. ($\Delta Emp(c,c',j) = 0$), then j does not contribute to $\Delta Tightness(c,c')$

- if j is characteristic of both c and c' (i.e. $\Delta Emp(c,c',j) = 0$), then j does not contribute to $\Delta Tightness(c,c')$

- if j is characteristic of c and not c' (i.e. $\Delta Emp(c,c',j) = 1$) and j' shared skills with other occupations tend to promote spatial mobility, then j increases $\Delta Tightness(c,c')$

- if j is characteristic of c and not c' (i.e. $\Delta Emp(c,c',j) = 1$) and j' shared skills with other occupations tend to hinder spatial mobility, then j decreases $\Delta Tightness(c,c')$

- if j is not characteristic of c but is characteristic of c' (i.e. $\Delta Emp(c,c',j) = -1$) and j' shared skills with other occupations tend to promote spatial mobility, then j decreases $\Delta Tightness(c,c')$

- if j is not characteristic of c but is characteristic of c' (i.e. $\Delta Emp(c, c', j) = -1$) and j' shared skills with other occupations tend to hinder spatial mobility, then j increases $\Delta Tightness(c, c')$

We use a job shift graph to examine how individual occupations contribute to the difference in average job tightness between two cities. Figure D-5, Figure D-6, and Figure D-7 provide example plots. Additionally, we measure the contribution of each O*NET skill to $\Delta Tightness$ using equation (D.7).

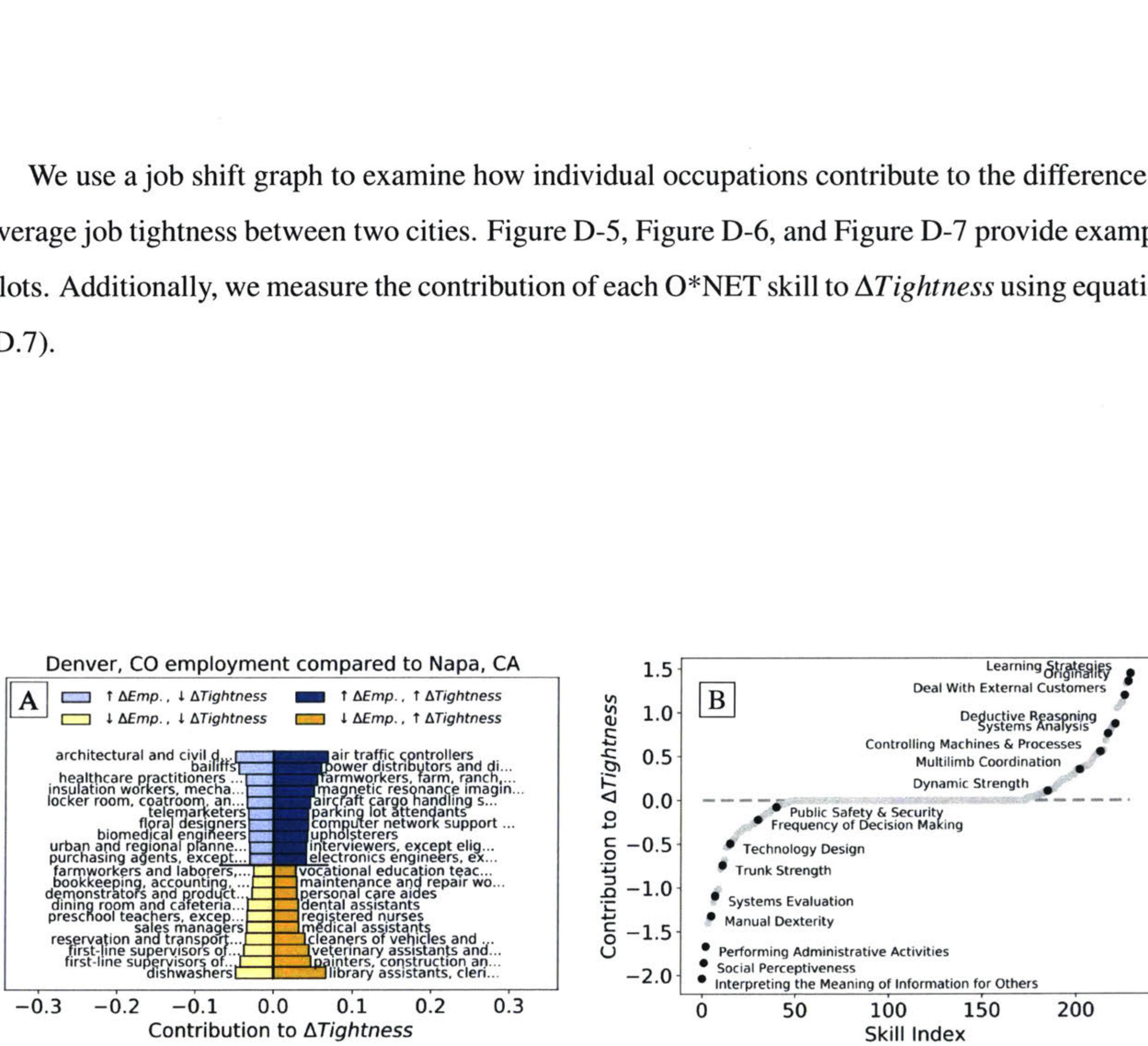


Figure D-5: Explaining the difference in the average job tightness of Denver, CO and Napa, CA. **(A)** Occupation's contribution to $\Delta Tightness$ is quantified by the bar size. Colors specify how the difference in employment contributes to the difference in average job tightness. **(B)** Quantifying how differences in employment for different O*NET skills contribute to $\Delta Tightness$.

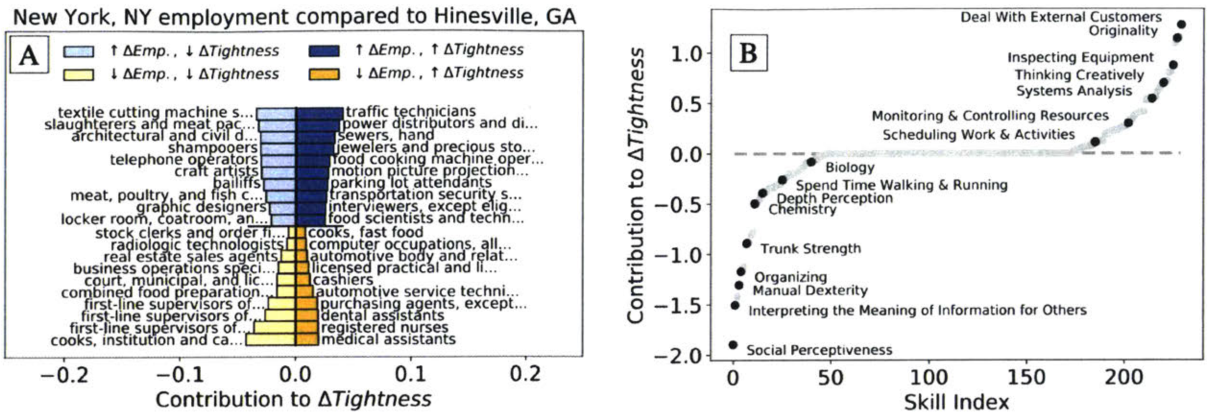


Figure D-6: Explaining the difference in the average job tightness of New York, NY and Hinesville, GA. **(A)** Occupation's contribution to $\Delta Tightness$ is quantified by the bar size. Colors specify how the difference in employment contributes to the difference in average job tightness. **(B)** Quantifying how differences in employment for different O*NET skills contribute to $\Delta Tightness$.

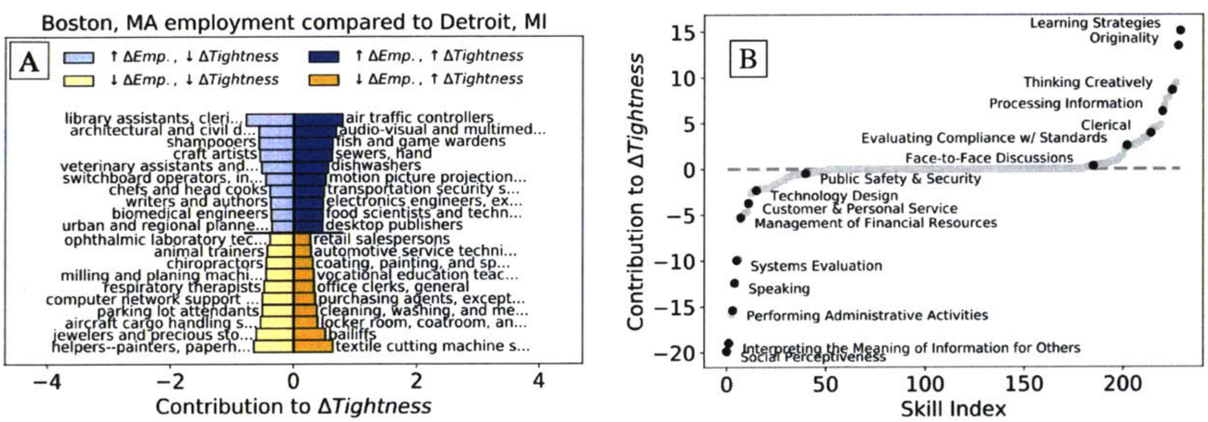


Figure D-7: Explaining the difference in the average job tightness of Boston, MA and Detroit, MI. **(A)** Occupation's contribution to $\Delta Tightness$ is quantified by the bar size. Colors specify how the difference in employment contributes to the difference in average job tightness. **(B)** Quantifying how differences in employment for different O*NET skills contribute to $\Delta Tightness$.

THIS PAGE INTENTIONALLY LEFT BLANK

Appendix E

The evolution of citation graphs in artificial intelligence research

E.1 Microsoft Academic Graph Fields of Study

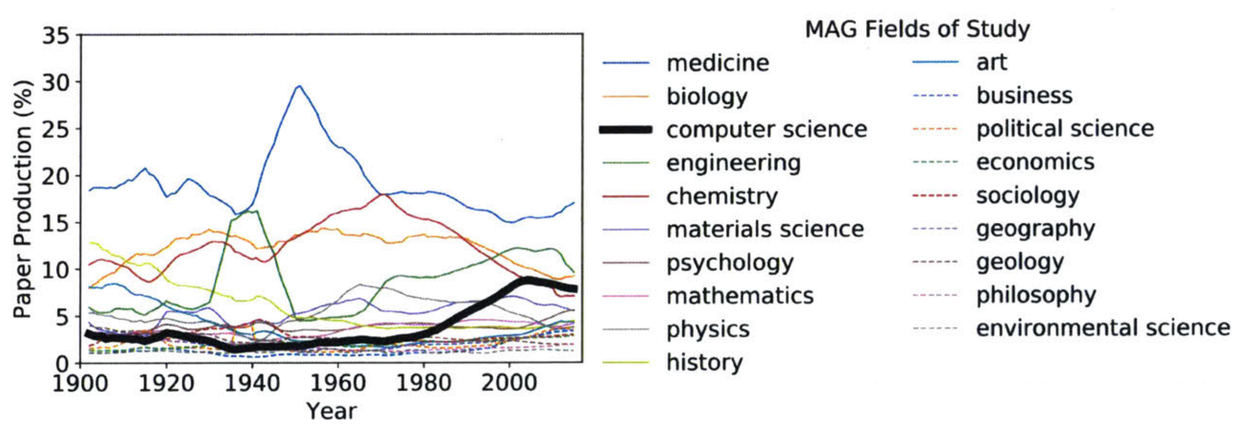


Figure E-1: Annual paper produce by top-level field of study. In the legend, fields of study are ordered according to their share of overall paper production in the final year of analysis.

The Microsoft Academic Graph (MAG) data assigns fields of study (FOS) to each publication in the dataset. FOS are selected from a hierarchical taxonomy of fields, including biology, mathematics, and, in particular, computer science as FOS at the top of the hierarchy. Figure E-1 demonstrates the share of annual paper production assigned to each top-level FOS from 1900 to 2018. Computer science has risen to the fourth most productive FOS in the last few decades beginning around 1950.

E.1.1 Computer Science Subfields

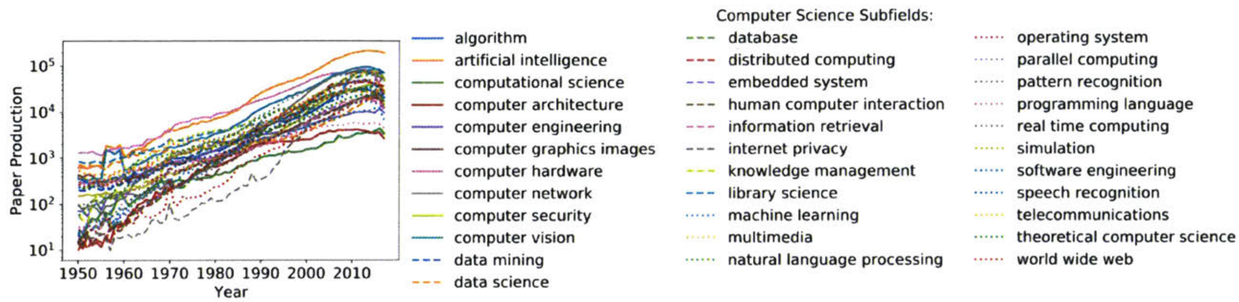


Figure E-2: Annual paper production by subfield of Computer Science.

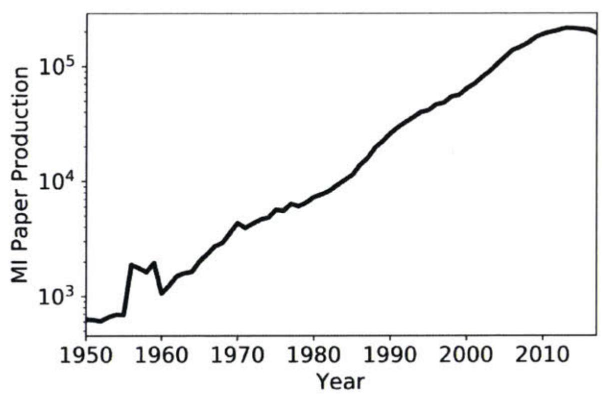


Figure E-3: Annual artificial intelligence paper production.

Each FOS is divided into subfields. We are particularly interested in the subfields of Computer Science. Figure E-2 demonstrates the annual paper production by Computer Science subfield.

E.2 Computer Science Subfield Citation Networks by Decade

Today, the phrases *machine intelligence* and *artificial intelligence* (AI) are most commonly used in reference for machine learning, but this was not always the case. Over the past 60 years, AI has been closely related to various Computer Science subfields, including Computer Vision, Machine Learning, Natural Language Processing , and Pattern Recognition.

To see this, we construct citation networks from the papers published in each Computer Science subfield (see Figures E-4-E-10). In these networks, nodes are CS subfields and node size corresponds to paper production in that subfield (note: one paper may belong to multiple subfields). The connections between subfields have width proportional to the number of references made between papers in a pair of fields. After constructing this raw citation network, we apply community detecting (according to [164]) to identify clusters of Computer Science subfields based on how these fields reference each other. In the citation networks, we use color to identify these clusters and encode the number of references between clusters in the width of the arrows.

The strength of association between AI and Computer Vision, Natural Language Processing (NLP), Machine Learning, and Pattern Recognition change dynamically over time. In fact, we can see the number of references between AI papers and NLP papers slowly diminish over the past several decades until NLP is actually contained in a separate community of Computer Science subfields. We also observe interesting dynamics around the subfield of Theoretical Computer Science and the emergence of the World Wide Web. Based on this analysis, we use papers in the following Computer Science subfields as a proxy for publications on AI-related fields of study: Artificial Intelligence, Machine Learning, Natural Language Processing, Computer Vision, and Pattern Recognition.

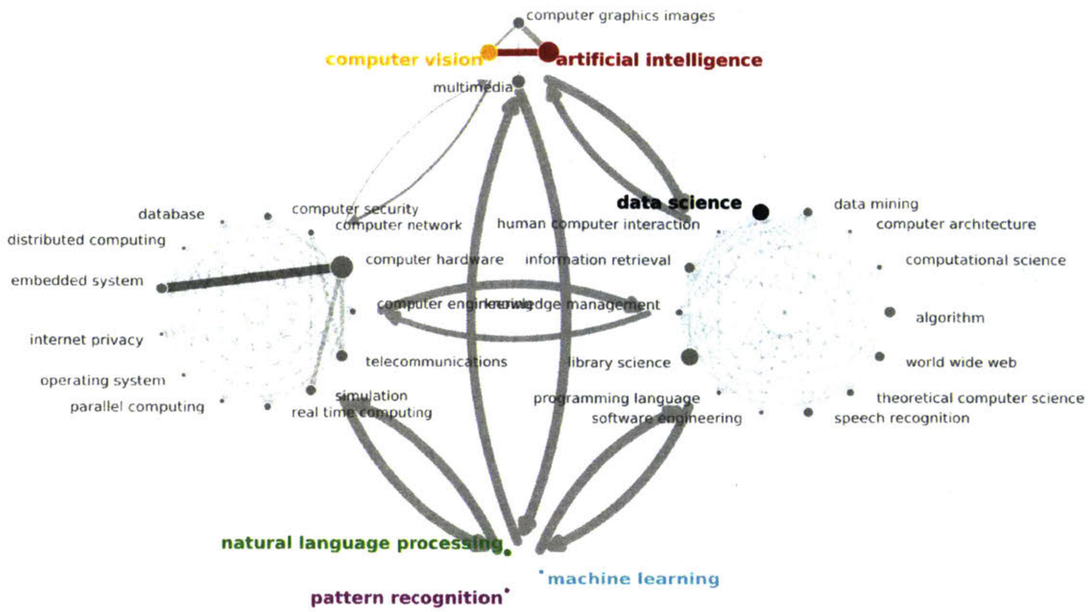


Figure E-4: Citation network for Computer Science (CS) subfields constructed from papers published in the 1950's.

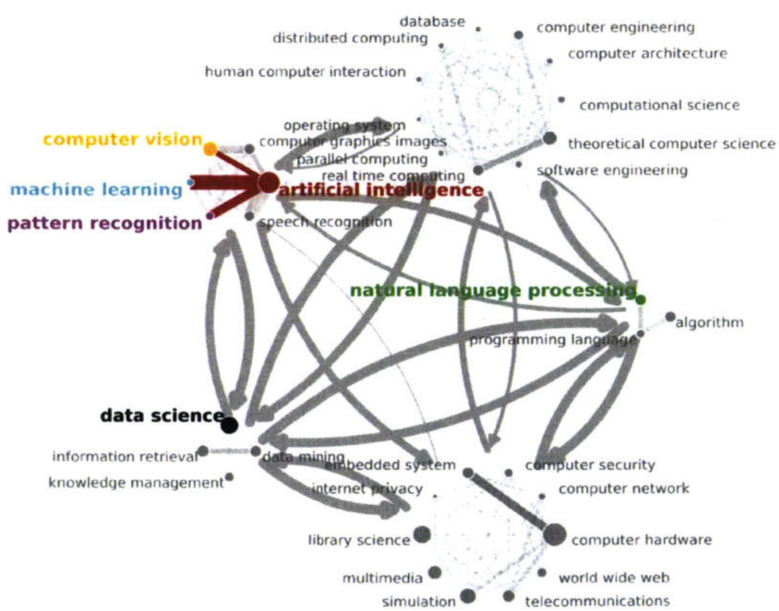


Figure E-5: Citation network for Computer Science (CS) subfields constructed from papers published in the 1960's.

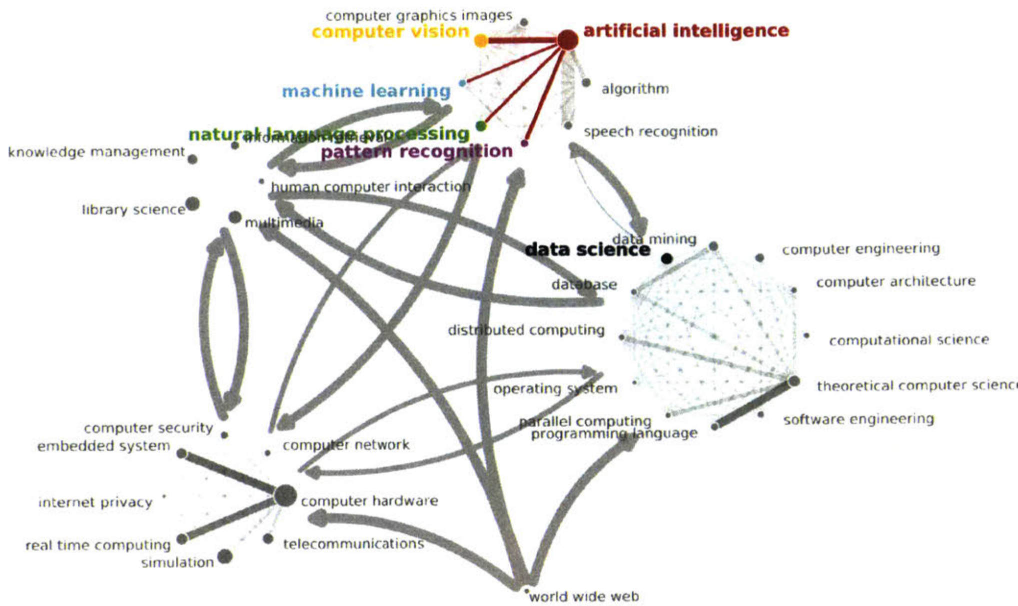


Figure E-6: Citation network for Computer Science (CS) subfields constructed from papers published in the 1970's.

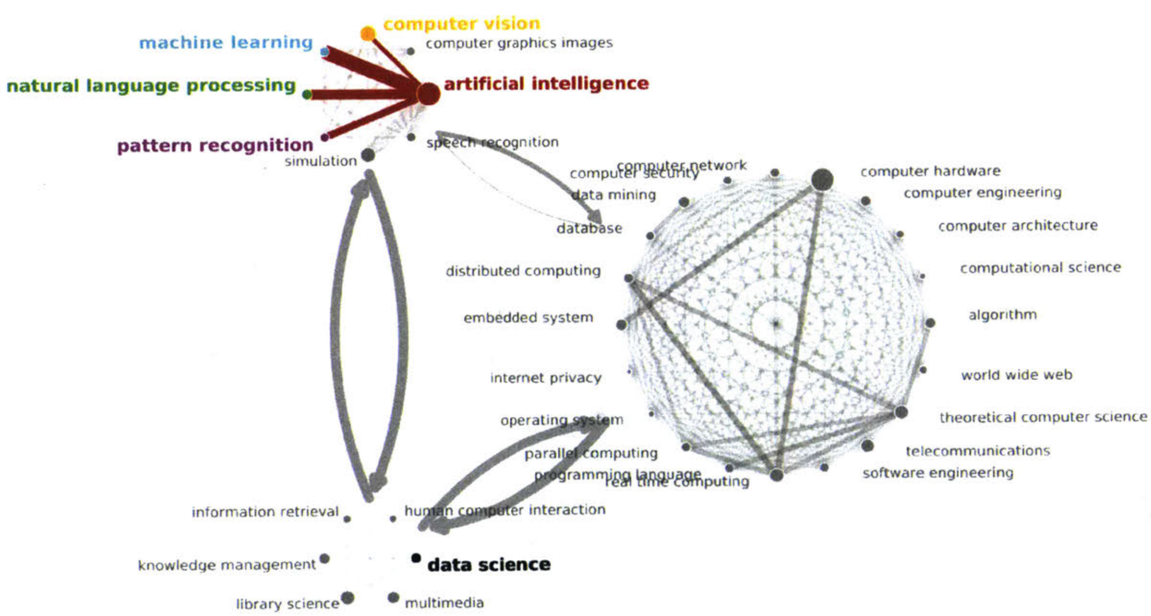


Figure E-7: Citation network for Computer Science (CS) subfields constructed from papers published in the 1980's.

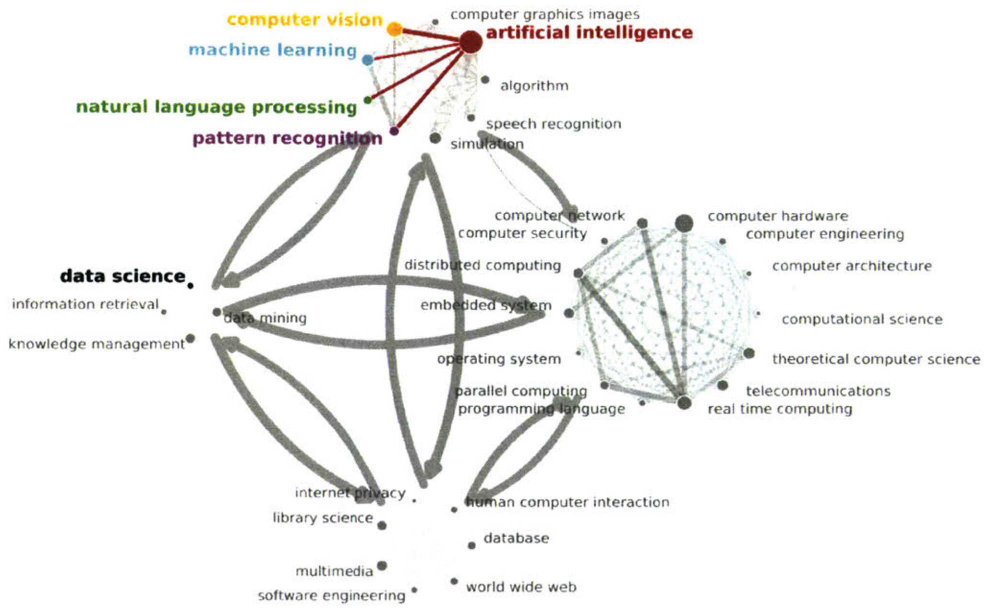


Figure E-8: Citation network for Computer Science (CS) subfields constructed from papers published in the 1990's.

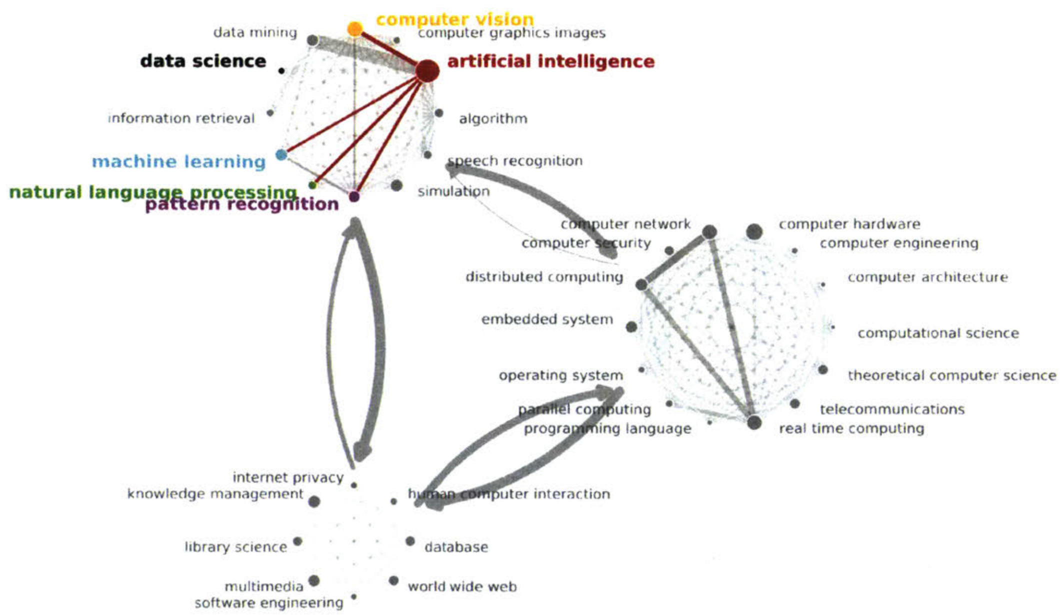


Figure E-9: Citation network for Computer Science (CS) subfields constructed from papers published in the 2000's.

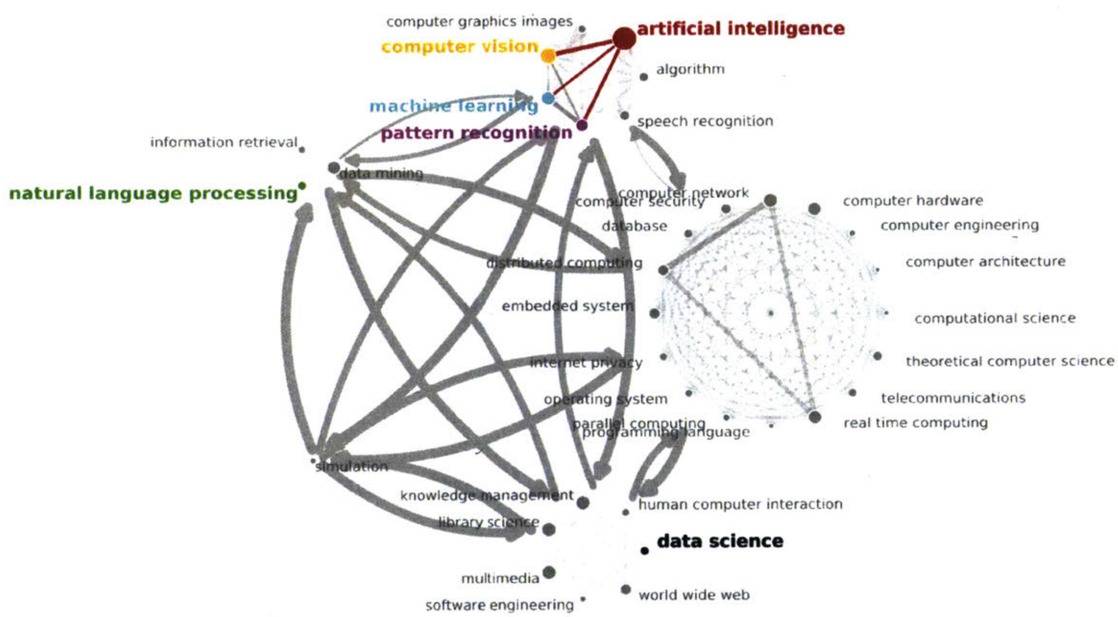


Figure E-10: Citation network for Computer Science (CS) subfields constructed from papers published in the 2010's.

E.3 Distribution of Artificial Intelligence Productivity by Research Institution

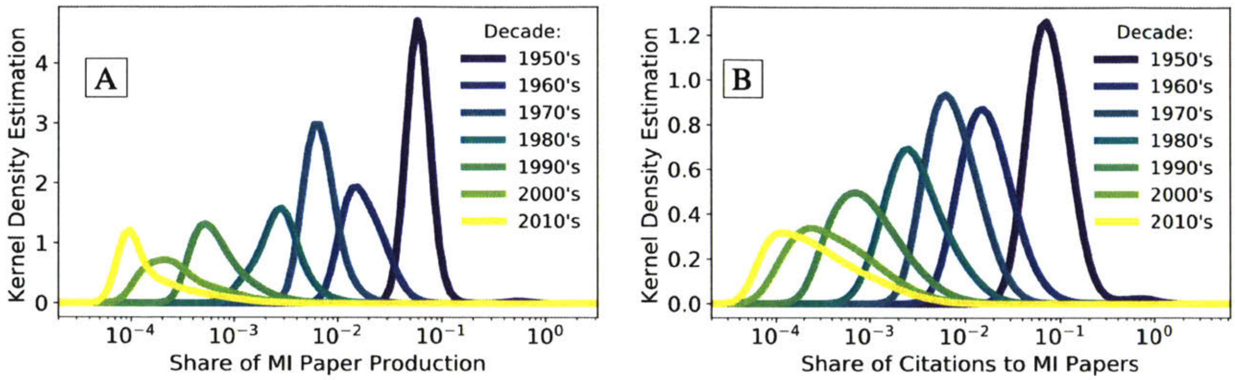


Figure E-11: For each decade, we plot the distribution of **(A)** artificial intelligence paper production across each research institution producing at least one AI paper, and **(B)** the distribution of citations to AI research after 10 years across research institutions producing at least one AI publication. All curves are approximated using a Gaussian kernel density estimator.

E.4 The Preference of Academic Fields for Industry Artificial Intelligence Publications

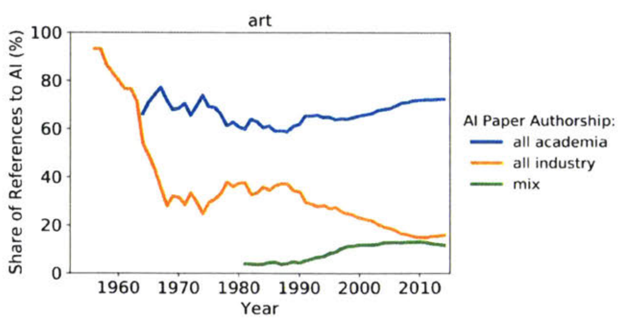


Figure E-12: The academic field's referencing behavior towards artificial intelligence papers according to the papers' authorship.

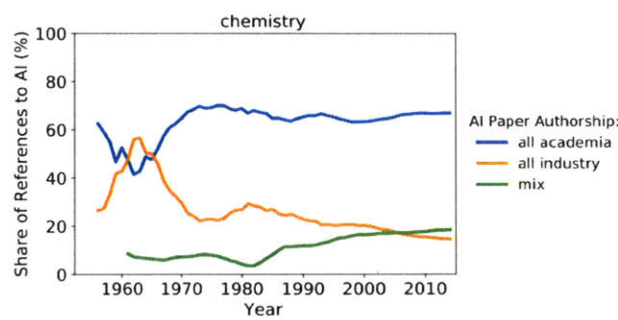


Figure E-13: The academic field's referencing behavior towards artificial intelligence papers according to the papers' authorship.

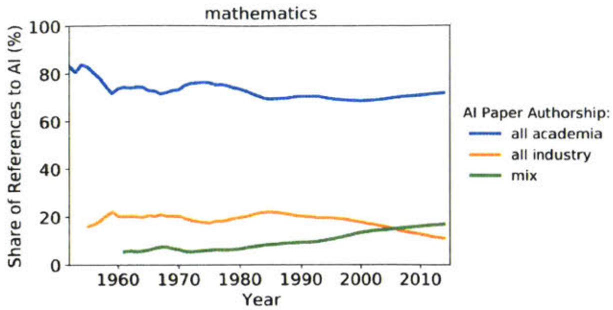


Figure E-14: The academic field's referencing behavior towards artificial intelligence papers according to the papers' authorship.

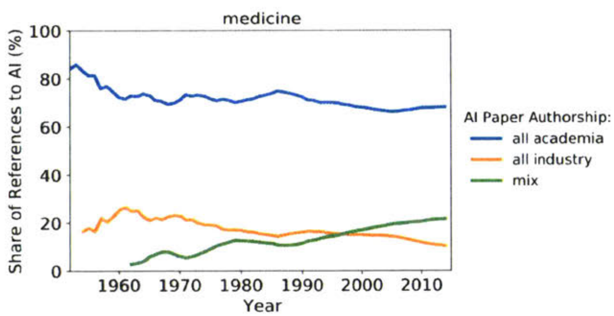


Figure E-15: The academic field's referencing behavior towards artificial intelligence papers according to the papers' authorship.

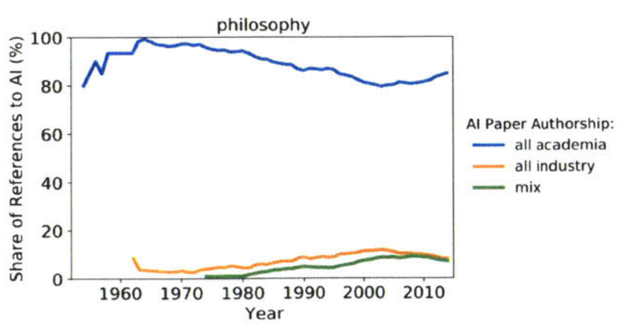


Figure E-16: The academic field's referencing behavior towards artificial intelligence papers according to the papers' authorship.

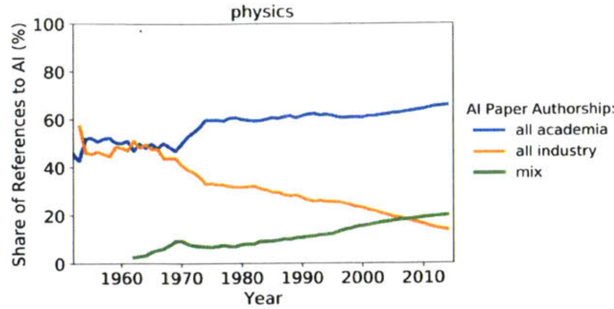


Figure E-17: The academic field's referencing behavior towards artificial intelligence papers according to the papers' authorship.

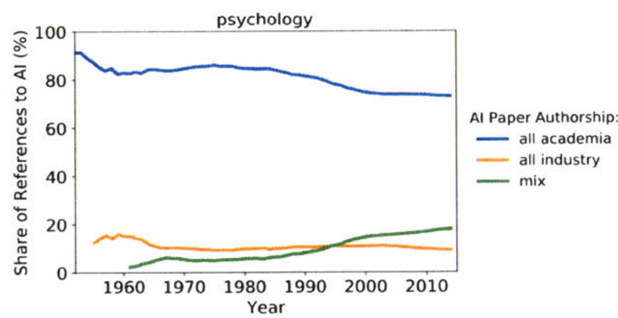


Figure E-18: The academic field's referencing behavior towards artificial intelligence papers according to the papers' authorship.

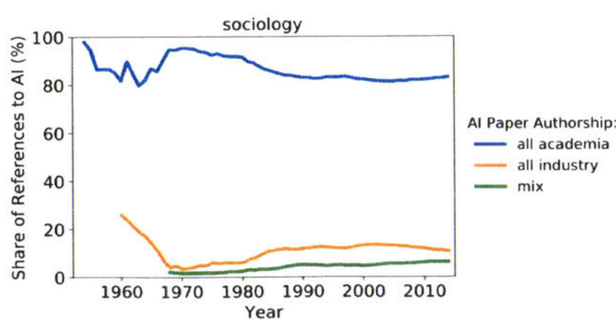


Figure E-19: The academic field's referencing behavior towards artificial intelligence papers according to the papers' authorship.

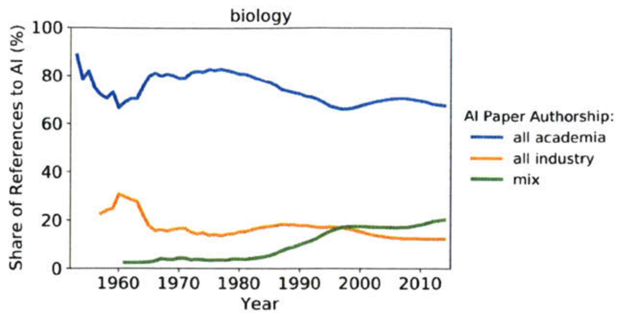


Figure E-20: The academic field's referencing behavior towards artificial intelligence papers according to the papers' authorship.

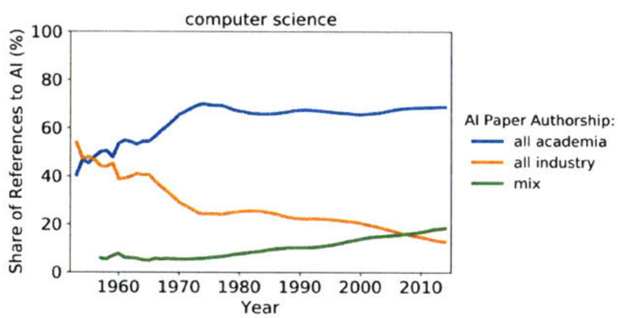


Figure E-21: The academic field's referencing behavior towards artificial intelligence papers according to the papers' authorship.

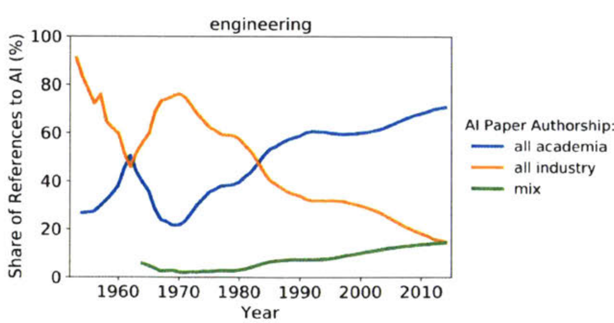


Figure E-22: The academic field's referencing behavior towards artificial intelligence papers according to the papers' authorship.

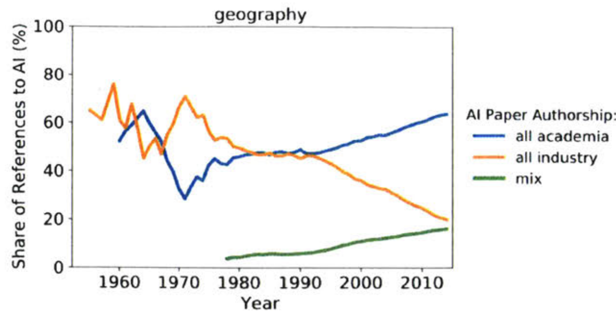


Figure E-23: The academic field's referencing behavior towards artificial intelligence papers according to the papers' authorship.

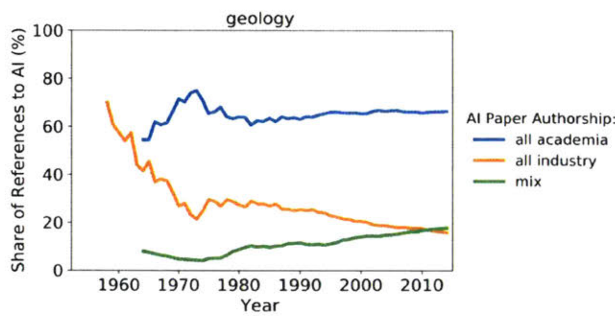


Figure E-24: The academic field's referencing behavior towards artificial intelligence papers according to the papers' authorship.

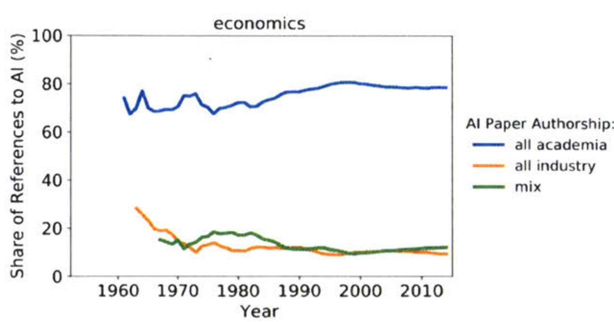


Figure E-25: The academic field's referencing behavior towards artificial intelligence papers according to the papers' authorship.

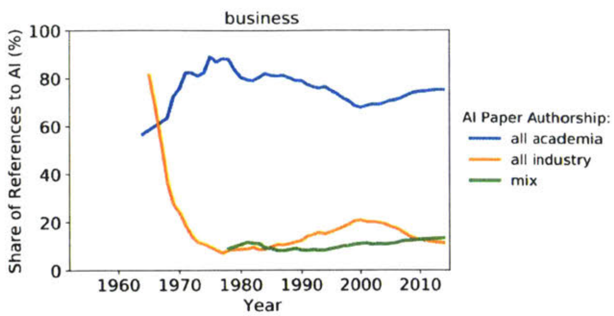


Figure E-26: The academic field's referencing behavior towards artificial intelligence papers according to the papers' authorship.

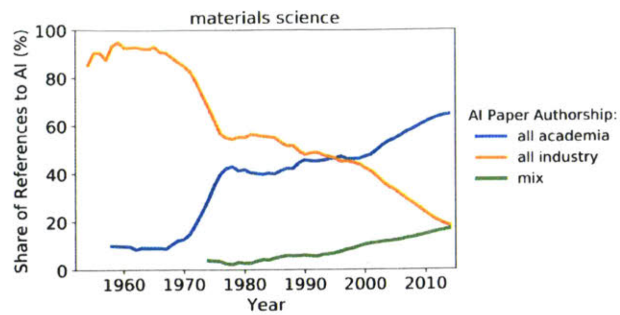


Figure E-28: The academic field's referencing behavior towards artificial intelligence papers according to the papers' authorship.

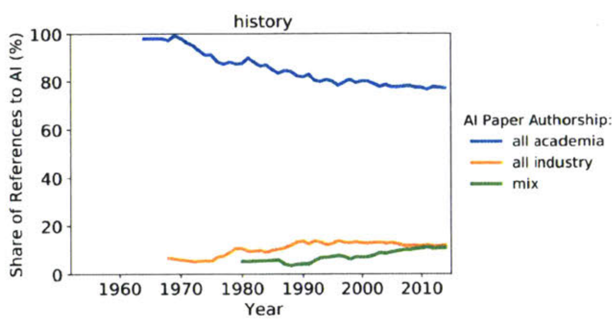


Figure E-27: The academic field's referencing behavior towards artificial intelligence papers according to the papers' authorship.

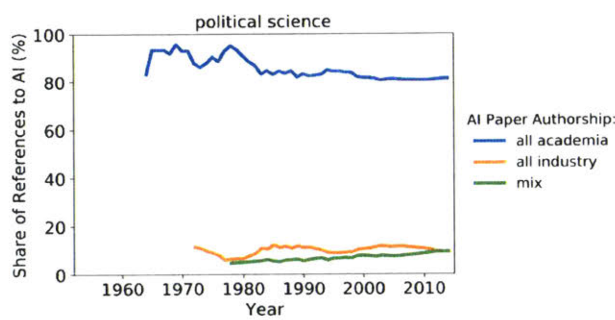


Figure E-29: The academic field's referencing behavior towards artificial intelligence papers according to the papers' authorship.

E.5 Bibliometric Diversity by Academic Field

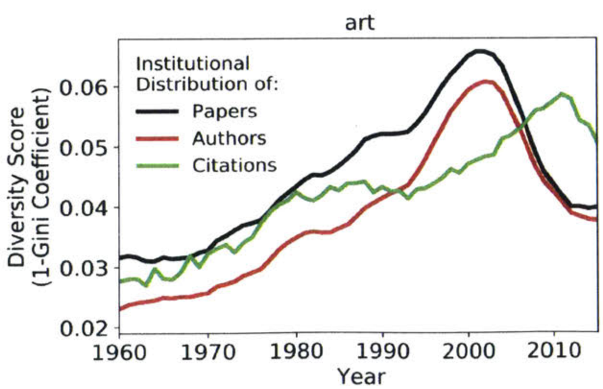


Figure E-30: The diversity of the annual distribution of all papers (black), authors (red), and all citations to papers (green) across research institutions according to the Gini coefficient.

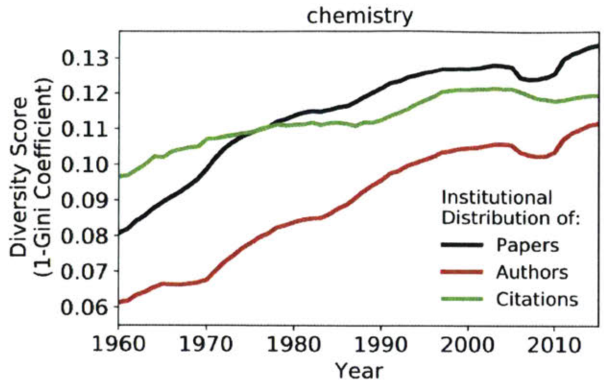


Figure E-31: The diversity of the annual distribution of all papers (black), authors (red), and all citations to papers (green) across research institutions according to the Gini coefficient.

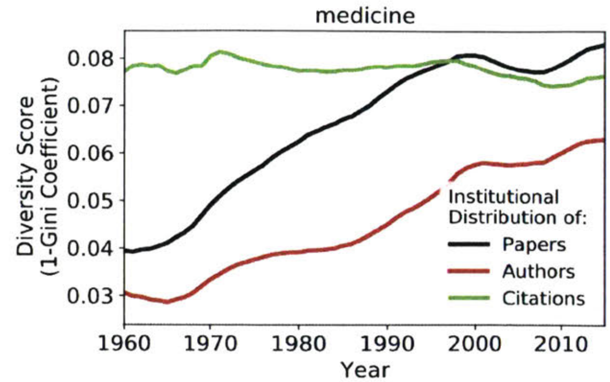


Figure E-33: The diversity of the annual distribution of all papers (black), authors (red), and all citations to papers (green) across research institutions according to the Gini coefficient.

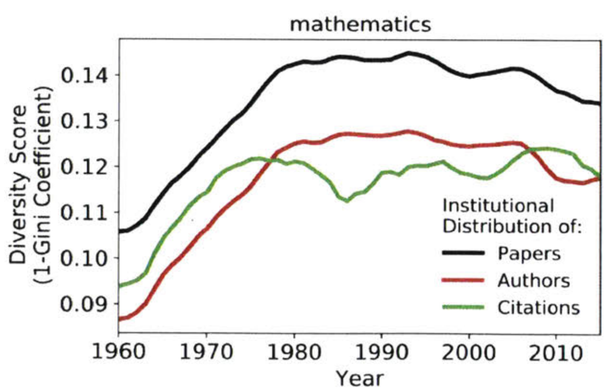


Figure E-32: The diversity of the annual distribution of all papers (black), authors (red), and all citations to papers (green) across research institutions according to the Gini coefficient.

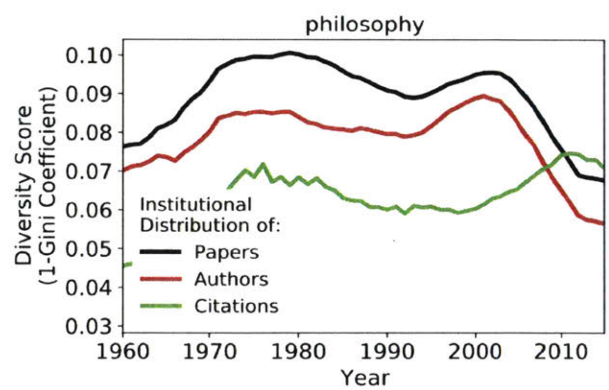
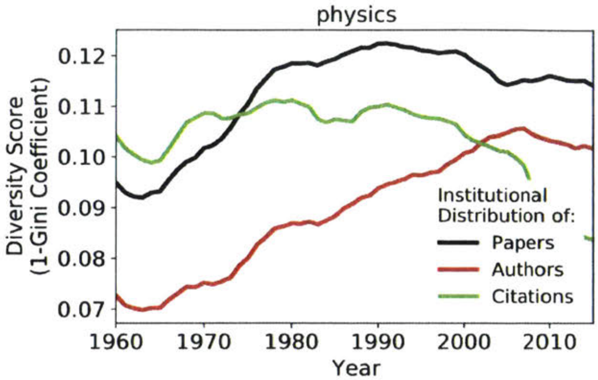


Figure E-34: The diversity of the annual distribution of all papers (black), authors (red), and all citations to papers (green) across research institutions according to the Gini coefficient.



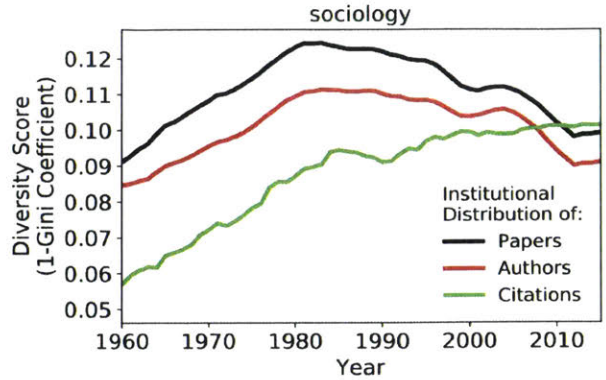
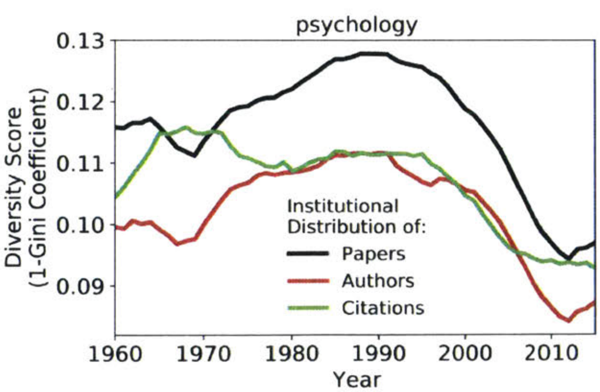


Figure E-35: The diversity of the annual distribution of all papers (black), authors (red), and all citations to papers (green) across research institutions according to the Gini coefficient.

Figure E-37: The diversity of the annual distribution of all papers (black), authors (red), and all citations to papers (green) across research institutions according to the Gini coefficient.



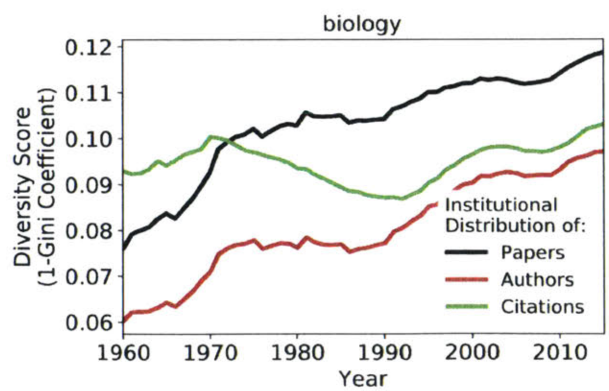


Figure E-36: The diversity of the annual distribution of all papers (black), authors (red), and all citations to papers (green) across research institutions according to the Gini coefficient.

Figure E-38: The diversity of the annual distribution of all papers (black), authors (red), and all citations to papers (green) across research institutions according to the Gini coefficient.

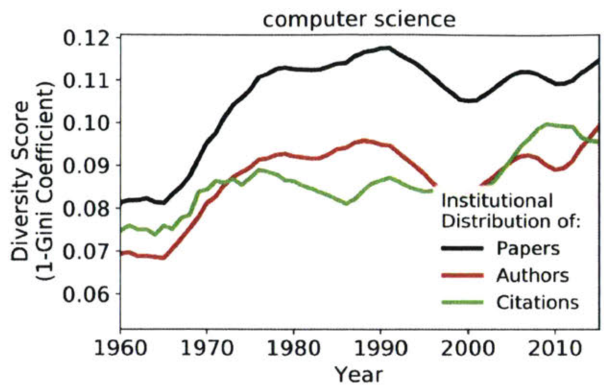


Figure E-39: The diversity of the annual distribution of all papers (black), authors (red), and all citations to papers (green) across research institutions according to the Gini coefficient.

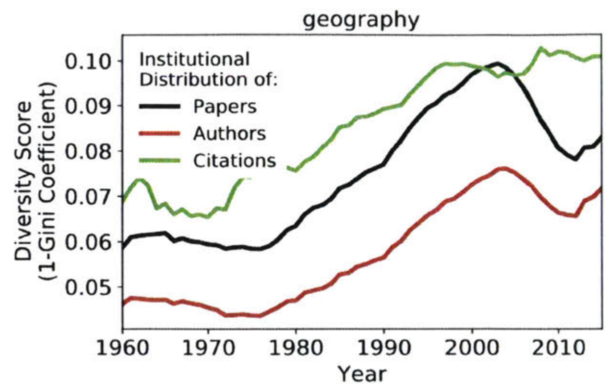


Figure E-41: The diversity of the annual distribution of all papers (black), authors (red), and all citations to papers (green) across research institutions according to the Gini coefficient.

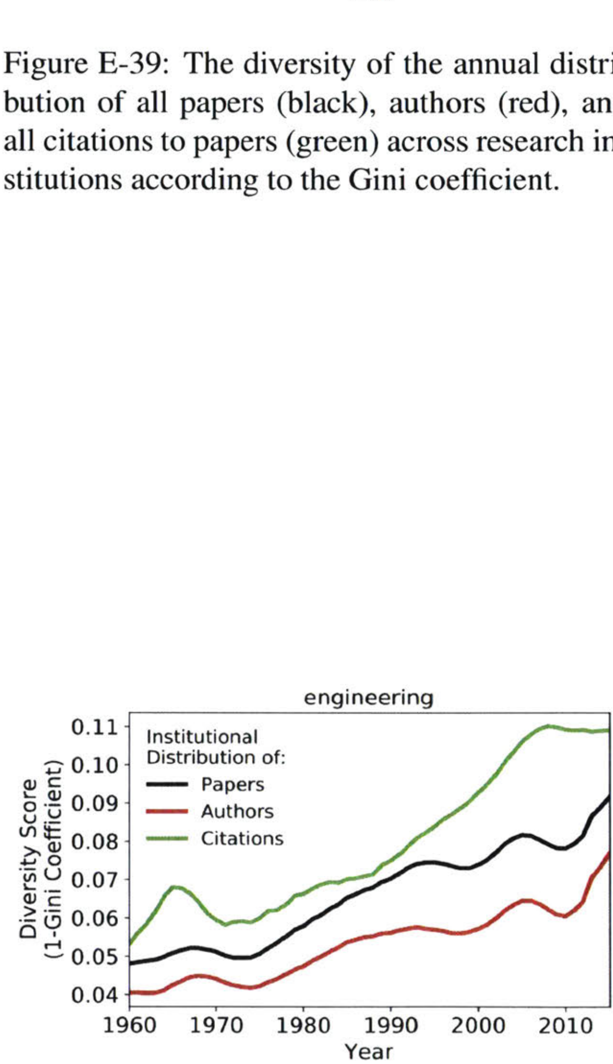


Figure E-40: The diversity of the annual distribution of all papers (black), authors (red), and all citations to papers (green) across research institutions according to the Gini coefficient.

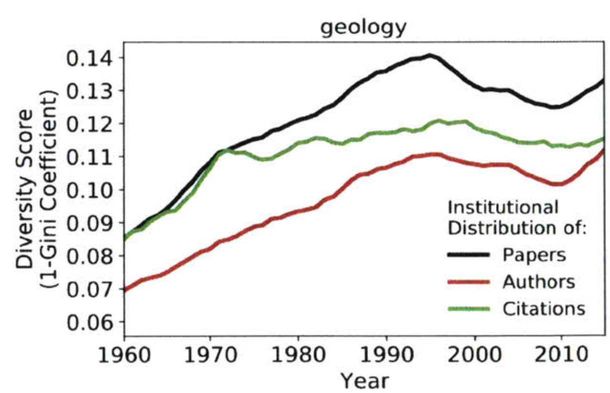


Figure E-42: The diversity of the annual distribution of all papers (black), authors (red), and all citations to papers (green) across research institutions according to the Gini coefficient.

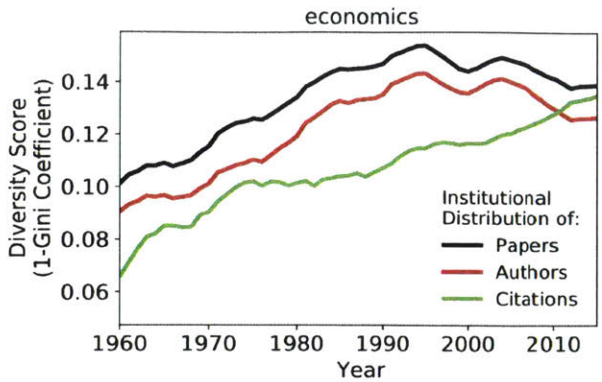


Figure E-43: The diversity of the annual distribution of all papers (black), authors (red), and all citations to papers (green) across research institutions according to the Gini coefficient.

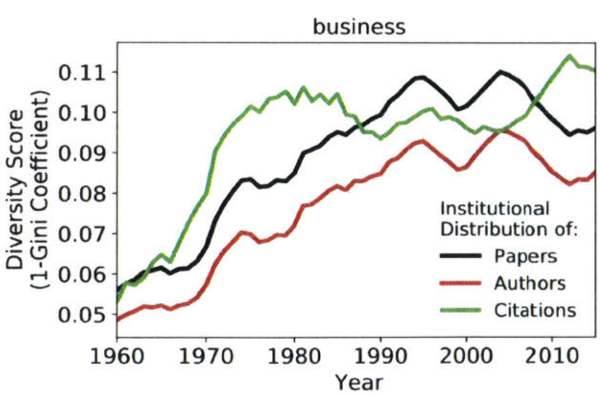


Figure E-44: The diversity of the annual distribution of all papers (black), authors (red), and all citations to papers (green) across research institutions according to the Gini coefficient.

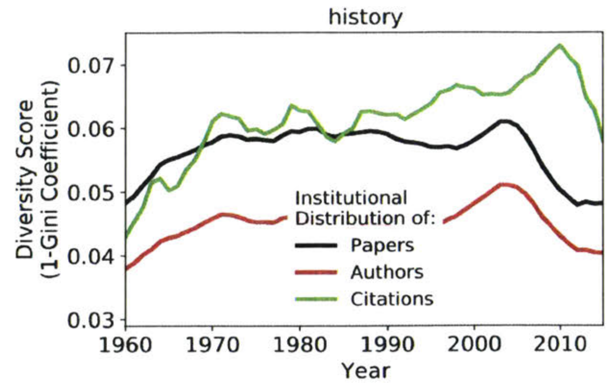


Figure E-45: The diversity of the annual distribution of all papers (black), authors (red), and all citations to papers (green) across research institutions according to the Gini coefficient.

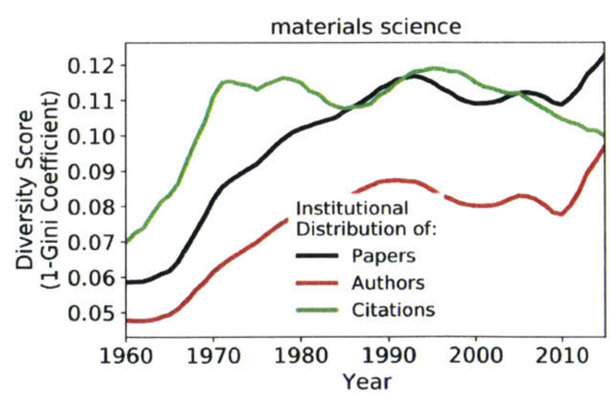


Figure E-46: The diversity of the annual distribution of all papers (black), authors (red), and all citations to papers (green) across research institutions according to the Gini coefficient.

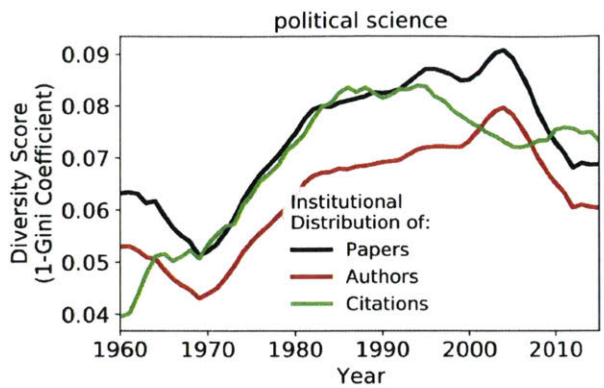


Figure E-47: The diversity of the annual distribution of all papers (black), authors (red), and all citations to papers (green) across research institutions according to the Gini coefficient.

E.6 Authorship by Field of Study Over Time

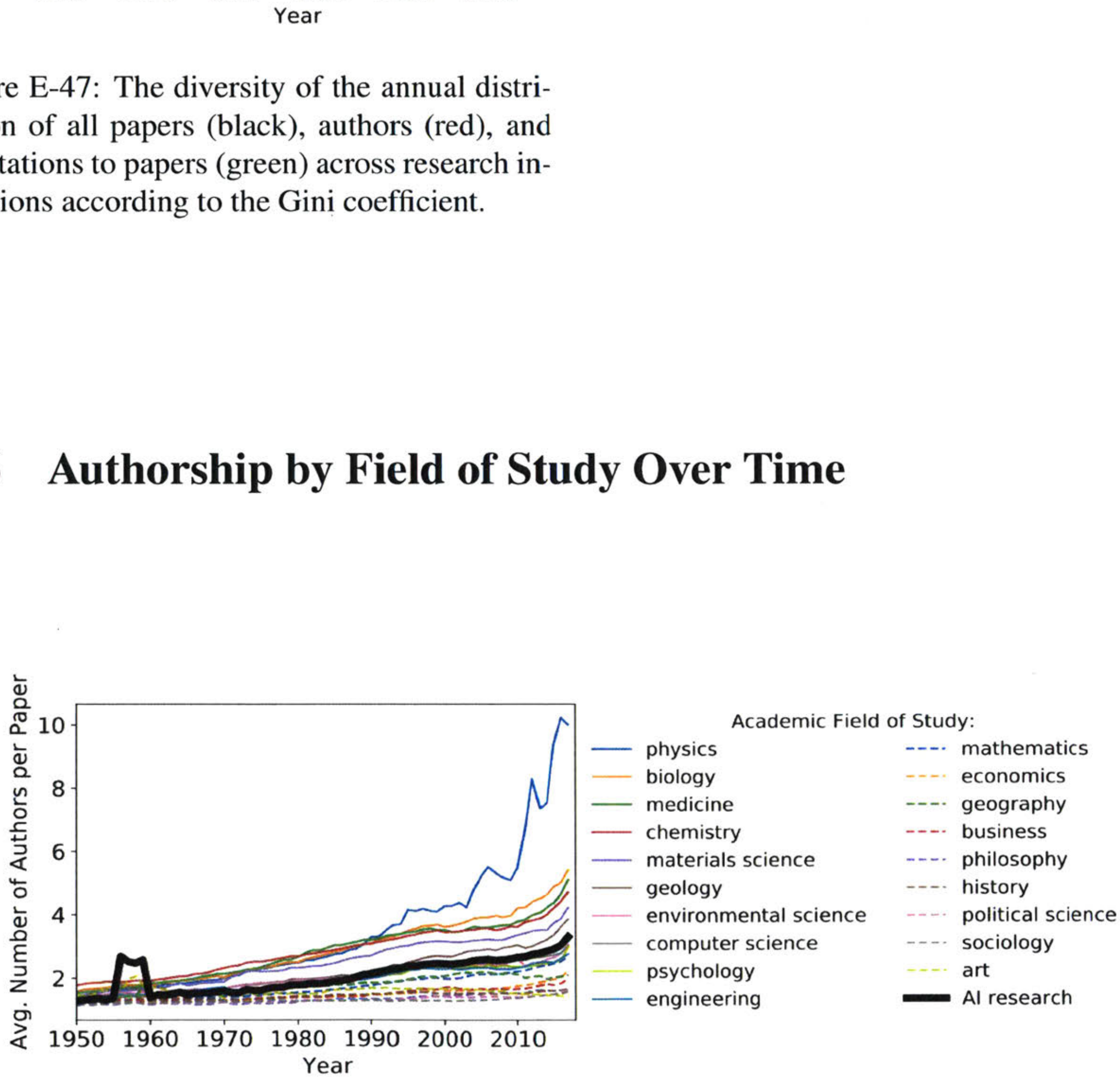


Figure E-48: For each Microsoft Academic Graph field of study, along with AI-related research as identified in the main text, we provide the average number of authors per paper published in each field in each year. The bump in AI research in the late 1950's is the result of a few publications with many co-authors combined with a reduced sample size in comparison to later years.

E.7 Authorship by AI Research Institution

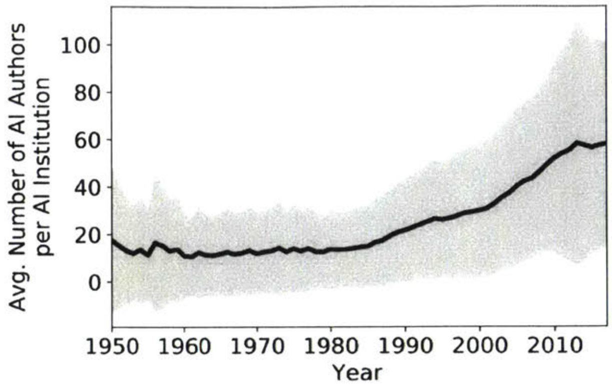


Figure E-49: In each year, we plot the average number of unique authors with at least one AI-related publication in that year across research institutions with at least one AI publication in that year. The grey area represents the 95% confidence interval for each year.

E.8 The Recent Rise of Chinese Institutions in AI Research

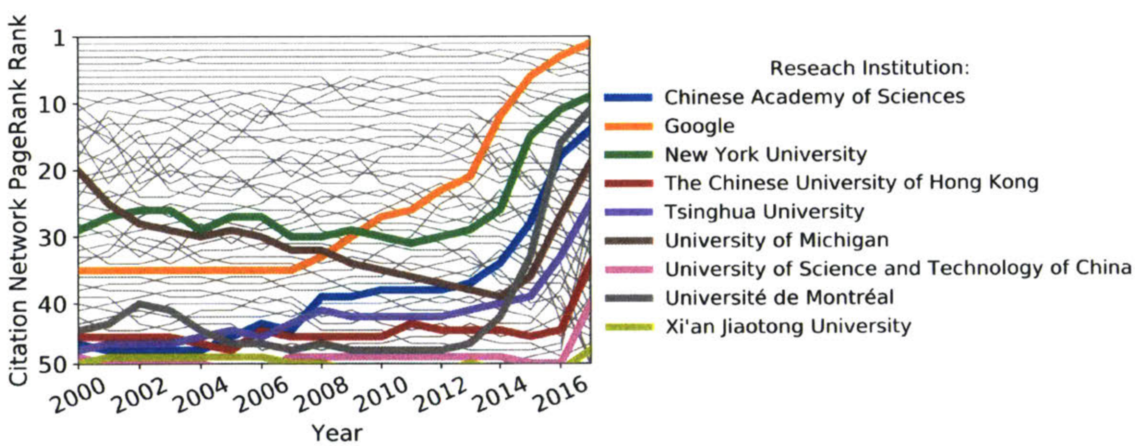


Figure E-50: Ranking the prominence of AI research institutions in recent years. Similar to Figure 4a in the main text, we calculate the PageRank of AI research institutions from the references made to AI papers published by other AI research institutions. Here, we rank-order the AI research institutions in each year since 2000 and highlight the institutions exhibiting the greatest increase in rank. In addition to Google's dramatic rise in PageRank rank, several academic institutions from around the world, but most notably in China, are rising in prominence within the AI research community. Gray lines represent AI research institutions that have fallen or remained constant in the AI prominence ranking in recent years.

THIS PAGE INTENTIONALLY LEFT BLANK

Bibliography

[1] World Migration Report 2015 - Migrants and Cities: New Partnerships to Manage Mobility. *International Organization for Migration*, 8, 2015.

[2] America's Shrinking Middle Class: A Close Look at Changes Within Metropolitan Areas. Technical report, Pew Research, May 2016.

[3] AmericaâĂŹs shrinking middle class: A close look at changes within metropolitan areas. Technical report, Pew Research, May 2016.

[4] The Road to Zero: A Vision of Achieving Zero Roadway Deaths by 2050. Technical report, National Safety Council and the RAND Corporation, 2018.

[5] Daron Acemoglu. Technical change, inequality, and the labor market. *Journal of Economic Literature*, 40(1):7–72, 2002.

[6] Daron Acemoglu. Labor-and capital-augmenting technical change. *Journal of the European Economic Association*, 1(1):1–37, 2003.

[7] Daron Acemoglu and David Autor. Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of labor economics*, volume 4, pages 1043–1171. Elsevier, 2011.

[8] Daron Acemoglu and Pascual Restrepo. Robots and jobs: Evidence from US labor markets. 2017.

[9] Philippe Aghion and Peter Howitt. A model of growth through creative destruction. Technical report, National Bureau of Economic Research, 1990.

[10] Ahmad Alabdulkareem*, Morgan R Frank*, Lijun Sun, Bedoor AlShebli, César Hidalgo, and Iyad Rahwan. Unpacking the polarization of workplace skills. *Science advances*, 4(7):eaao6030, 2018. (*authors contributed equally).

[11] Ali Alichi, Kory Kantenga, and Juan A Solé. Income Polarization in the United States. 2016.

[12] Katharine A Anderson. Skill networks and measures of complex human capital. *Proceedings of the National Academy of Sciences*, page 201706597, 2017.

[13] Dirk Antonczyk, Bernd Fitzenberger, and Ute Leuschner. Can a task-based approach explain the recent changes in the German wage structure? ZEW Discussion Papers 08-132, Mannheim, 2009.

[14] Elsa Arcaute, Carlos Molinero, Erez Hatna, Roberto Murcio, Camilo Vargas-Ruiz, A Paolo Masucci, and Michael Batty. Cities and regions in britain through hierarchical percolation. *Open Science*, 3(4):150691, 2016.

[15] Daniele Archibugi, Andrea Filippetti, and Marion Frenz. Economic crisis and innovation: Is destruction prevailing over accumulation? *Research Policy*, 42(2):303–314, 2013.

[16] Melanie Arntz, Terry Gregory, and Ulrich Zierahn. The risk of automation for jobs in oecd countries: A comparative analysis. *OECD Social, Employment, and Migration Working Papers*, (189):0_1, 2016.

[17] Caroline Ash, Barbara R Jasny, Leslie Roberts, Richard Stone, and Andrew M Sugden. Reimagining cities. *Science*, 319(5864):739–739, 2008.

[18] David Autor. The polarization of job opportunities in the US labor market: Implications for employment and earnings. *Center for American Progress and The Hamilton Project*, 2010.

[19] David Autor. Why are there still so many jobs? The history and future of workplace automation. *Journal of Economic Perspectives*, 29(3):3–30, 2015.

[20] David Autor. Work of the Past, Work of the Future. In *Richard T. Ely Lecture*. American Economic Association, January 2019.

[21] David Autor and David Dorn. The growth of low-skill service jobs and the polarization of the us labor market. *American Economic Review*, 103(5):1553–97, 2013.

[22] David Autor, David Dorn, and Gordon H Hanson. The China syndrome: Local labor market effects of import competition in the United States. *American Economic Review*, 103(6):2121–68, 2013.

[23] David H Autor, Lawrence F Katz, and Melissa S Kearney. Trends in US wage inequality: Revising the revisionists. *The Review of economics and statistics*, 90(2):300–323, 2008.

[24] David H Autor, Frank Levy, and Richard J Murnane. The skill content of recent technological change: An empirical exploration. *The Quarterly journal of economics*, 118(4):1279–1333, 2003.

[25] David H Autor and Brendan Price. The changing task composition of the US labor market: An update of Autor, Levy, and Murnane (2003). *Unpublished manuscript*, 2013.

[26] Ryan Avent. A world without work is coming–it could be utopia or it could be hell. *The Guardian*, 19, 2016.

[27] Robert L Axtell. Zipf distribution of us firm sizes. *Science*, 293(5536):1818–1820, 2001.

[28] Andreas Backhaus, Inmaculada Martinez-Zarzoso, and Chris Muris. Do climate variations explain bilateral migration? A gravity model analysis. *IZA Journal of Migration*, 4(1):3, 2015.

[29] Albert-Lászlő Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

[30] Mike Barlow. AI and medicine. *O'Reilly Media*, 2016.

[31] Eric Bartelsman, John Haltiwanger, and Stefano Scarpetta. Microeconomic evidence of creative destruction in industrial and developing countries. 2004.

[32] Marc Barthélemy. Spatial networks. *Physics Reports*, 499(1):1–101, 2011.

[33] Stefano Battiston, J Doyne Farmer, Andreas Flache, Diego Garlaschelli, Andrew G Haldane, Hans Heesterbeek, Cars Hommes, Carlo Jaeger, Robert May, and Marten Scheffer. Complexity theory and financial regulation. *Science*, 351(6275):818–819, 2016.

[34] Paul Beaudry, David A Green, and Benjamin M Sand. The great reversal in the demand for skill and cognitive tasks. *Journal of Labor Economics*, 34(S1):S199–S247, 2016.

[35] Michel Beine, Simone Bertoli, and Jesús Fernández-Huertas Moraga. A practitioner's guide to gravity models of international migration. *The World Economy*, 39(4):496–512, 2016.

[36] Mariano G Beiró, André Panisson, Michele Tizzoni, and Ciro Cattuto. Predicting human mobility through the assimilation of social media traces into mobility models. *EPJ Data Science*, 5(1):30, 2016.

[37] Eli Bekman, John Bound, and Stephen Machin. Implications of skill-biased technological change: international evidence. *The quarterly journal of economics*, 113(4):1245–1279, 1998.

[38] Alice Benessia and Silvio Funtowicz. Sustainability and techno-science: What do we want to sustain and for whom? *International Journal of Sustainable Development*, 18(4):329–348, 2015.

[39] Max Berest, Ralucca Gera, Zachary Lukens, N Martinez, and Ben McCaleb. Predicting network evolution through temporal Twitter snapshots for paris attacks of 2015. In *International Conference on Social Computing, Behavioral-Cultural Modeling, & Prediction and Behavior Representation in Modeling and Simulation (BRIMS)*, 2016.

[40] Ralf Bergholz, Klaus Timm, and Hubert Weisser. Autonomous vehicle arrangement and method for controlling an autonomous vehicle, November 21 2000. US Patent 6,151,539.

[41] James Bessen. How computer automation affects occupations: Technology, jobs, and skills. 2015.

[42] James Bessen. *Learning by doing: the real connection between innovation, wages, and wealth*. Yale University Press, 2015.

[43] James Bessen. AI and jobs: the role of demand. In *Economics of Artificial Intelligence*. University of Chicago Press, 2018.

[44] James E Bessen. How computer automation affects occupations: Technology, jobs, and skills. 2016.

[45] Luís M. A. Bettencourt, José Lobo, Deborah Strumsky, and Geoffrey B. West. Urban scaling and its deviations: Revealing the structure of wealth, innovation and crime across cities. *PLoS ONE*, 5(11):e13541, 2010.

[46] Luís M.A. Bettencourt, José Lobo, Dirk Helbing, Christian Kühnert, and Geoffrey B. West. Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences*, 104(17):7301–7306, 2007.

[47] Luís MA Bettencourt, Horacio Samaniego, and Hyejin Youn. Professional diversity and the productivity of cities. *Scientific reports*, 4, 2014.

[48] Govert E Bijwaard, Christian Schluter, and Jackline Wahba. The impact of labor market dynamics on the return migration of immigrants. *Review of Economics and Statistics*, 96(3):483–494, 2014.

[49] Sandra E. Black and Alexandra Spitz-Oener. Explaining Women's Success: Technological Change and the Skill Content of Women's Work. Technical report, National Bureau of Economic Research, 2010.

[50] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, 2008.

[51] David E Bloom, David Canning, and Günther Fink. Urbanization and the wealth of nations. *Science*, 319(5864):772–775, 2008.

[52] Jean-François Bonnefon, Azim Shariff, and Iyad Rahwan. The social dilemma of autonomous vehicles. *Science*, 352(6293):1573–1576, 2016.

[53] Katy Börner, Olga Scrivner, Mike Gallant, Shutian Ma, Xiaozhong Liu, Keith Chewning, Lingfei Wu, and James A. Evans. Skill discrepancies between research, education, and jobs reveal the critical need to supply soft skills for the data economy. *Proceedings of the National Academy of Sciences*, 115(50):12630–12637, 2018.

[54] Jeremy Bowles. The computerisation of european jobs–who will win and who will lose from the impact of new technology onto old areas of employment. *Bruegel blog*, 17, 2014.

[55] Jonathan Brogaard et al. High frequency trading and its impact on market quality. *Northwestern University Kellogg School of Management Working Paper*, 66, 2010.

[56] Erik Brynjolfsson and Andrew McAfee. *Race against the machine: How the digital revolution is accelerating innovation, driving productivity, and irreversibly transforming employment and the economy*. Brynjolfsson and McAfee, 2012.

[57] Erik Brynjolfsson and Andrew McAfee. *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. WW Norton & Company, 2014.

[58] Erik Brynjolfsson and Paul Milgrom. Complementarity in organizations. *The Handbook of Organizational Economics*, pages 11–55, 2013.

[59] Erik Brynjolfsson and Tom Mitchell. What can machine learning do? Workforce implications. *Science*, 358(6370):1530–1534, 2017.

[60] Erik Brynjolfsson, Tom Mitchell, and Daniel Rock. What can machines learn, and what does it mean for occupations and the economy? In *AEA Papers and Proceedings*, volume 108, pages 43–47, 2018.

[61] Erik Brynjolfsson, Daniel Rock, and Chad Syverson. Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics. Technical report, National Bureau of Economic Research, 2017.

[62] Carsten Brzeski and Inga Burk. Die roboter kommen. *Folgen der Automatisierung für den deutschen Arbeitsmarkt. INGDiBa Economic Research*, 2015.

[63] Joy Buolamwini and Timnit Gebru. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on Fairness, Accountability and Transparency*, pages 77–91, 2018.

[64] Joy Buolamwini and TED Talk. How IâĂŹm fighting bias in algorithms. *TED Talks*, 2016.

[65] Randy Burd, Kimberly Andrews Espy, Md Iqbal Hossain, Stephen Kobourov, Nirav Merchant, and Helen Purchase. GRAM: global research activity map. In *Proceedings of the 2018 International Conference on Advanced Visual Interfaces*, page 31. ACM, 2018.

[66] Beatriz Cabot, María Martell, Juan I Esteban, Sílvia Sauleda, Teresa Otero, Rafael Esteban, Jaime Guàrdia, and Jordi Gómez. Nucleotide and amino acid complexity of hepatitis c virus quasispecies in serum and liver. *Journal of virology*, 74(2):805–811, 2000.

[67] Bo Carlsson. *Technological systems and economic performance: the case of factory automation*, volume 5. Springer Science & Business Media, 2012.

[68] Raj Chetty, David Grusky, Maximilian Hell, Nathaniel Hendren, Robert Manduca, and Jimmy Narang. The Fading American Dream: Trends in Absolute Income Mobility Since 1940. Technical report, National Bureau of Economic Research, 2016.

[69] Federico Cingano. Trends in income inequality and its impact on economic growth. 2014.

[70] Ronald H Coase. The nature of the firm. *economica*, 4(16):386–405, 1937.

[71] Yehonatan Cohen, Danny Hendler, and Amir Rubin. Detection of Malicious Webmail Attachments Based on Propagation Patterns. *Knowledge-Based Systems*, 2017.

[72] Jacob W Crandall, Mayada Oudah, Fatimah Ishowo-Oloko, Sherief Abdallah, Jean-François Bonnefon, Manuel Cebrian, Azim Shariff, Michael A Goodrich, Iyad Rahwan, et al. Cooperating with machines. *Nature Communications*, 9(1):233, 2018.

[73] Mathias Czaika and Christopher R Parsons. The gravity of high-skilled migration policies. *Demography*, 54(2):603–630, 2017.

[74] Alison M Darcy, Alan K Louie, and Laura Weiss Roberts. Machine learning and the profession of medicine. *Jama*, 315(6):551–552, 2016.

[75] Wolfgang Dauth, Sebastian Findeisen, Jens Südekum, and Nicole Woessner. German robots-the impact of industrial robots on workers. *Institute for Employment Research*, 30, 2017.

[76] Thomas H Davenport and Julia Kirby. *Only humans need apply: winners and losers in the age of smart machines*. Harper Business, 2016.

[77] Mercedes Delgado, Michael E. Porter, and Scott Stern. Defining clusters of related industries. *Journal of Economic Geography*, 16(1):1–38, 2016.

[78] David Deming and Lisa B Kahn. Skill requirements across firms and labor markets: Evidence from job postings for professionals. *Journal of Labor Economics*, 36(S1):S337–S369, 2018.

[79] Timothy M Devinney and Jan Hohberger. The past is prologue: Moving on from Cultures Consequences. *Journal of International Business Studies*, 48(1):48–62, 2017.

[80] Christian L Dunis, Peter W Middleton, Andreas Karathanasopolous, and Konstantinos Theofilatos. *Artificial Intelligence in Financial Markets: Cutting Edge Applications for Risk Management, Portfolio Optimization and Economics*. Springer, 2016.

[81] Christian Dustmann, Johannes Ludsteck, and Uta Schanberg. Revisiting the German wage structure. *The Quarterly Journal of Economics*, 124(2):843–881, 2009.

[82] Nathan Eagle, Michael Macy, and Rob Claxton. Network diversity and economic development. *Science*, 328(5981):1029–1031, 2010.

[83] Avraham Ebenstein, Ann Harrison, Margaret McMillan, and Shannon Phillips. Estimating the impact of trade and offshoring on American workers using the current population surveys. *Review of Economics and Statistics*, 96(4):581–595, 2014.

[84] Suhendry Effendy and Roland HC Yap. Analysing trends in computer science research: A preliminary study using the Microsoft Academic Graph. In *Proceedings of the 26th International Conference on World Wide Web Companion*, pages 1245–1250. International World Wide Web Conferences Steering Committee, 2017.

[85] Stern Léa H. Tan Chenhao Erel, Isil and Michael S. Weisbach. Could Machine Learning Help Companies Select Better Board Directors? *Harvard Business Review*, 2018.

[86] Jingfang Fan, Jun Meng, Yosef Ashkenazy, Shlomo Havlin, and Hans Joachim Schellnhuber. Network analysis reveals strongly localized impacts of El Niño. *Proceedings of the National Academy of Sciences*, page 201701214, 2017.

[87] Chris Farrell. Will robots create economic utopia? *Bloomberg*, 2013.

[88] Dalibor Fiala and Gabriel Tutoky. Computer Science Papers in Web of Science: A Bibliometric Analysis. *Publications*, 5(4):23, 2017.

[89] Gary S Fields. Rural-urban migration, urban unemployment and underemployment, and job-search activity in ldcs. *Journal of development economics*, 2(2):165–187, 1975.

[90] C Jara Figueroa, Bogang Jun, Edward L Glaeser, and César Hidalgo. The Role of Industry, Occupation, and Location-Specific Knowledge in the Survival of New Firms. Technical report, National Bureau of Economic Research, 2018.

[91] Sergio Firpo, Nicole M. Fortin, and Thomas Lemieux. Occupational Tasks and Changes in the Wage Structure. IZA Discussion Papers 5542, Institute for the Study of Labor (IZA), February 2011.

[92] Richard Florida. *The rise of the creative class*. Basic books New York, 2004.

[93] Luciano Floridi. Distributed morality in an information society. *Science and engineering ethics*, 19(3):727–743, 2013.

[94] Morgan R. Frank, David Autor, James E. Bessen, Erik Brynjolfsson, Manuel Cebrian, David J. Deming, Maryann Feldman, Matthew Groh, José Lobo, Esteban Moro, Dashun Wang, Hyejin Youn, and Iyad Rahwan. Towards Understanding the Impact of AI on Labor. 2019. Accepted for publication at PNAS.

[95] Morgan R. Frank, Lijun Sun, Manuel Cebrian, Hyejin Youn, and Iyad Rahwan. Small cities face greater impact from automation. *Journal of The Royal Society Interface*, 15(139), 2018.

[96] Morgan R Frank, Dashun Wang, Manuel Cebrian, and Iyad Rahwan. The evolution of citation graphs in artificial intelligence research. *Nature Machine Intelligence*, 1(2):79–85, February 2019.

[97] Peter Frase. *Four futures: Life after capitalism*. Verso books, 2016.

[98] Carl Benedikt Frey and Michael A. Osborne. The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, 114:254 – 280, 2017.

[99] Jill Freyne, Lorcan Coyle, Barry Smyth, and Padraig Cunningham. Relative status of journal and conference publications in computer science. *Communications of the ACM*, 53(11):124–132, 2010.

[100] Jianxi Gao, Baruch Barzel, and Albert-László Barabási. Universal resilience patterns in complex networks. 2016.

[101] Andres J Garcia, Deepa K Pindolia, Kenneth K Lopiano, and Andrew J Tatem. Modeling internal migration flows in sub-saharan africa using census microdata. *Migration Studies*, 3(1):89–110, 2015.

[102] Christina Gathmann and Uta Schönberg. How general is human capital? a task-based approach. *Journal of Labor Economics*, 28(1):1–49, 2010.

[103] Laura K Gee, Jason J Jones, Christopher J Fariss, Moira Burke, and James H Fowler. The paradox of weak ties in 55 countries. *Journal of Economic Behavior & Organization*, 133:362–372, 2017.

[104] Alan S Gerber and Donald P Green. *Field experiments: Design, analysis, and interpretation.* WW Norton, 2012.

[105] Edward Glaeser. *Triumph of the city: How our greatest invention makes us richer, smarter, greener, healthier, and happier.* Penguin, 2011.

[106] Edward L Glaeser, Hedi D Kallal, Jose A Scheinkman, and Andrei Shleifer. Growth in cities. Technical report, National Bureau of Economic Research, 1991.

[107] Lawrence Goldin, Claudia. Katz. *The Race Between Education and Technology.* 2009.

[108] Mark S Granovetter. The strength of weak ties. *American Journal of Sociology*, 78(6):1360–1380, 1973.

[109] Omar A. Guerrero and Robert L. Axtell. Employment growth through labor flow networks. *PLOS ONE*, 8(5):1–12, 05 2013.

[110] Hrothgar John Habakkuk. *American and British Technology in the Nineteenth Century: The search for labour saving inventions.* Cambridge University Press, 1962.

[111] Ricardo Hausmann and César A. Hidalgo. The network structure of economic output. *Journal of Economic Growth*, 16(4):309–342, 2011.

[112] AB Yujia He. How China is preparing for an AI-powered Future. *Wilson Center, Washington, DC*, 2017.

[113] J Vernon Henderson et al. Urban development: Theory, fact, and illusion. *OUP Catalogue*, 1991.

[114] Joshua Seth Herbach and Nathaniel Fairfield. Detecting that an autonomous vehicle is in a stuck condition, March 31 2015. US Patent 8,996,224.

[115] C. A. Hidalgo, B. Klinger, A.-L. Barabási, and R. Hausmann. The Product Space Conditions the Development of Nations. *Science*, 317(5837):482–487, July 2007.

[116] César A Hidalgo, Pierre-Alexandre Balland, Ron Boschma, Mercedes Delgado, Maryann Feldman, Koen Frenken, Edward Glaeser, Canfei He, Dieter F Kogler, Andrea Morrison, et al. The principle of relatedness. In *International Conference on Complex Systems*, pages 451–457. Springer, 2018.

[117] César A. Hidalgo and Ricardo Hausmann. The building blocks of economic complexity. *PNAS*, 106(26):10570–10575, June 2009.

[118] Ahmed Hosny, Chintan Parmar, John Quackenbush, Lawrence H Schwartz, and Hugo JWL Aerts. Artificial intelligence in radiology. *Nature Reviews Cancer*, page 1, 2018.

[119] Sven E Hug and Martin P Brändle. The coverage of Microsoft Academic: Analyzing the publication output of a university. *Scientometrics*, 113(3):1551–1571, 2017.

[120] Toshie Ikenaga. Polarization of the Japanese labor market–adoption of ict and changes in tasks required. *Japanese Journal of Labor Studies*, 2009.

[121] Toshie Ikenaga and Ryo Kambayashi. Long-term trends in the polarization of the japanese labor market: the increase of non-routine task input and its valuation in the labor market. *Japanese Journal of Labor Studies*, 2009.

[122] Kosuke Imai, Luke Keele, and Dustin Tingley. A general approach to causal mediation analysis. *Psychological methods*, 15(4):309, 2010.

[123] Greg Ip. We survived spreadsheets, and we'll survive AI. *Wall Street Journal*, 2017.

[124] Andrew M Isserman. The location quotient approach to estimating regional economic impacts. *Journal of the American Institute of Planners*, 43(1):33–41, 1977.

[125] Kristin Jackson. The world's first motel rests upon its memories. *Seattle Times*, 25:1993, 1993.

[126] Heather Beth Johnson. *The American dream and the power of wealth: Choosing schools and inheriting inequality in the land of opportunity*. Routledge, 2014.

[127] Steven E Jones. *Against technology: From the Luddites to neo-Luddism*. Routledge, 2013.

[128] Robert F. Kennedy. Welcoming remarks to the opening session of the interfederal assembly of pax romana. *Department of Justice*, 1964.

[129] Alan C Kerckhoff. Education and social stratification processes in comparative perspective. *Sociology of Education*, pages 3–18, 2001.

[130] Hillel J Kieval. Pursuing the Golem of Prague: Jewish Culture and the Invention of a Tradition. *Modern Judaism*, 17(1):1–20, 1997.

[131] Andrei Kirilenko, Albert S Kyle, Mehrdad Samadi, and Tugkan Tuzun. The Flash Crash: High-frequency trading in an electronic market. *The Journal of Finance*, 72(3):967–998, 2017.

[132] Jon Kleinberg, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. Human decisions and machine predictions. *The quarterly journal of economics*, 133(1):237–293, 2017.

[133] Thomas Klier. From tail fins to hybrids: How detroit lost its dominance of the us auto market. *Economic Perspectives*, 33(2), 2009.

[134] Joel Klinger, Juan C Mateos-Garcia, and Konstantinos Stathoulopoulos. Deep Learning, Deep Change? Mapping the Development of the Artificial Intelligence General Purpose Technology. submitted, 2018.

[135] Rakesh Kochhar, Richard Fry, and Molly Rohal. The American Middle Class Is Losing Ground, 2015.

[136] Suzanne Kok and Bas ter Weel. Cities, tasks, and skills. *Journal of Regional Science*, 54(5):856–892, 2014.

[137] Frauke Kraas, Surinder Aggarwal, Martin Coy, and Günter Mertins. *Megacities: our global urban future*. Springer Science & Business Media, 2013.

[138] Uma Kumar, Vinod Kumar, and J N KAPUR. Normalized measures of entropy. *International Journal Of General System*, 12(1):55–69, 1986.

[139] Vincent Larivière, Benoit Macaluso, Philippe Mongeon, Kyle Siler, and Cassidy R. Sugimoto. Vanishing industries and the rising monopoly of universities in published research. *PLOS ONE*, 13(8):1–10, 08 2018.

[140] Jorge C Leitão, José María Miotto, Martin Gerlach, and Eduardo G Altmann. Is this scaling nonlinear? *Royal Society Open Science*, 3(7):150649, 2016.

[141] Wassily Leontief. Machines and man. *Scientific American*, 187(3):150–164, 1952.

[142] Frank MacCrory, George Westerman, Yousef Alhammadi, and Erik Brynjolfsson. Racing with and against the machine: changes in occupational skill composition in an era of rapid technological advance. 2014.

[143] Brigitte C Madrian and Lars Lefgren. A note on longitudinally matching Current Population Survey (CPS) respondents. 1999.

[144] Brigitte C Madrian and Lars John Lefgren. An approach to longitudinally matching current population survey (cps) respondents. *Journal of Economic and Social Measurement*, 26(1):31–62, 2000.

[145] Anne E Magurran, Maria Dornelas, Faye Moyes, Nicholas J Gotelli, and Brian McGill. Rapid biotic homogenization of marine fish assemblages. *Nature Communications*, 6:8405, 2015.

[146] Alvin P Makohon-Moore, Ming Zhang, Johannes G Reiter, Ivana Bozic, Benjamin Allen, Deepanjan Kundu, Krishnendu Chatterjee, Fay Wong, Yuchen Jiao, Zachary A Kohutek, et al. Limited heterogeneity of known driver gene mutations among the metastases of individual patients with pancreatic cancer. *Nature genetics*, 49(3):358, 2017.

[147] Mariana Mazzucato. Financing innovation: creative destruction vs. destructive creation. *Industrial and Corporate Change*, page dtt025, 2013.

[148] Pamela McCorduck. *Machines who think: A personal inquiry into the history and prospects of artificial intelligence*. AK Peters/CRC Press, 2009.

[149] Penny Mealy, J. Doyne Farmer, and Alexander Teytelboym. Interpreting economic complexity. *Science Advances*, 5(1), 2019.

[150] Guy Michaels, Ashwini Natraj, and John Van Reenen. Has ict polarized skill demand? evidence from eleven countries over 25 years. Working Paper 16138, National Bureau of Economic Research, June 2010.

[151] Alex P. Miller. Want Less-Biased Decisions? Use Algorithms. *Harvard Business Review*, 2018.

[152] Tom Mitchell and Erik Brynjolfsson. Track how technology is transforming work. *Nature*, 544(7650):290–292, 2017.

[153] Mark R Montgomery. The urban transformation of the developing world. *Science*, 319(5864):761–764, 2008.

[154] Allison C Morgan, Dimitrios Economou, Samuel F Way, and Aaron Clauset. Prestige drives epistemic inequality in the diffusion of scientific ideas. *EPJ Data Science*, 7:40, October 2018.

[155] Flaviano Morone, Gino Del Ferraro, and Hernán A. Makse. The k-core as a predictor of structural collapse in mutualistic ecosystems. *Nature Physics*, 2018.

[156] Dale T Mortensen and Christopher A Pissarides. Job creation and job destruction in the theory of unemployment. *The review of economic studies*, 61(3):397–415, 1994.

[157] Sendhil Mullainathan and Jann Spiess. Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, 31(2):87–106, 2017.

[158] Rachata Muneepeerakul, José Lobo, Shade T. Shutters, Andrés Goméz-Liévano, and Murad R. Qubbaj. Urban Economies and Occupation Space: Can They Get There from Here? *PLoS ONE*, 8(9):e73676, September 2013.

[159] Nikhil Naik, Scott Duke Kominers, Ramesh Raskar, Edward L Glaeser, and César A Hidalgo. Computer vision uncovers predictors of physical urban change. *Proceedings of the National Academy of Sciences*, 114(29):7571–7576, 2017.

[160] Erica Naone. The dark side of the technology utopia. *MIT Technology Review*, 2007.

[161] Frank Neffke and Martin Henning. Skill relatedness and firm diversification. *Strategic Management Journal*, 34(3):297–316, 2013.

[162] Frank Neffke, Martin Henning, and Ron Boschma. How do regions diversify over time? industry relatedness and the development of new growth paths in regions. *Economic Geography*, 87(3):237–265, 2011.

[163] Frank Neffke and Martin Svensson Henning. Revealed relatedness: Mapping industry space. *Papers in Evolutionary Economic Geography*, 8:19, 2008.

[164] Mark EJ Newman. Finding community structure in networks using the eigenvectors of matrices. *Physical review E*, 74(3):036104, 2006.

[165] Suphakit Niwattanakul, Jatsada Singthongchai, Ekkachai Naenudorn, and Supachanun Wanapu. Using of Jaccard coefficient for keywords similarity. In *Proceedings of the International MultiConference of Engineers and Computer Scientists*, volume 1, 2013.

[166] Erneson A Oliveira, José S Andrade Jr, and Hernán A Makse. Large cities are less green. *Scientific reports*, 4, 2014.

[167] Morten Olsen and David Hemous. The rise of the machines: Automation, horizontal innovation and income inequality. In *2014 Meeting Papers*, number 162. Society for Economic Dynamics, 2014.

[168] Felipe G Operti, Erneson A Oliveira, Humberto A Carmona, Javam C Machado, and José S Andrade. The light pollution as a surrogate for urban population of the us cities. *Physica A: Statistical Mechanics and its Applications*, 492:1088–1096, 2018.

[169] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The PageRank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab, 1999.

[170] Mica Pajarinen, Petri Rouvinen, and Anders Ekeland. Computerization threatens one-third of Finnish and Norwegian employment. *ETLA Brief*, (34), 2015.

[171] Wei Pan, Gourab Ghoshal, Coco Krumme, Manuel Cebrian, and Alex Pentland. Urban characteristics attributable to density-driven tie formation. *Nature communications*, 4, 2013.

[172] Manh Cuong Pham, Ralf Klamma, and Matthias Jarke. Development of computer science disciplines: a social network analysis approach. *Social Network Analysis and Mining*, 1(4):321–340, 2011.

[173] Thomas Edward Pilutti, Matt Y Rupp, Roger Arnold Trombley, Andrew Waldis, and Wilford Trent Yopp. Autonomous vehicle identification, January 24 2017. US Patent 9,552,735.

[174] Sadie Plant. *Zeros and ones*. Doubleday Books, 1997.

[175] Burton R Pollin. Philosophical and Literary Sources of Frankenstein. *Comparative Literature*, 17(2):97–108, 1965.

[176] Michael E Porter. *Clusters and the new economics of competition*, volume 76. Harvard Business Review Boston, 1998.

[177] Michael E Porter. Location, competition, and economic development: Local clusters in a global economy. *Economic Development Quarterly*, 14(1):15–34, 2000.

[178] Denise Pumain. Scaling laws and urban systems. *Santa Fe Working Paper*, 2004.

[179] Robert D Putnam. *Our kids: The American dream in crisis*. Simon and Schuster, 2016.

[180] John M Quigley. Urban diversity and economic growth. *The Journal of Economic Perspectives*, 12(2):127–138, 1998.

[181] Iyad Rahwan. Society-in-the-loop: programming the algorithmic social contract. *Ethics and Information Technology*, 20(1):5–14, 2018.

[182] Sean F. Reardon and Glenn Firebaugh. Measures of multigroup segregation. *Sociological Methodology*, 32:33–67, January 2002.

[183] Lauren A Rivera. When two bodies are (not) a problem: Gender and relationship status discrimination in academic hiring. *American Sociological Review*, 82(6):1111–1138, 2017.

[184] Céline Rozenblat and Denise Pumain. Firm linkages, innovation and the evolution of urban systems. In *Cities in Globalization: Practices, Policies, and Theories*, pages 130–156. Routledge, 2007.

[185] Hernán D Rozenfeld, Diego Rybski, José S Andrade, Michael Batty, H Eugene Stanley, and Hernán A Makse. Laws of population growth. *Proceedings of the National Academy of Sciences*, 105(48):18702–18707, 2008.

[186] Hernán D Rozenfeld, Diego Rybski, Xavier Gabaix, and Hernán A Makse. The area and population of cities: new insights from a different perspective on cities. *The American Economic Review*, 101(5):2205–2225, 2011.

[187] Mikail Rubinov and Olaf Sporns. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, 52(3):1059–1069, 2010.

[188] Stuart Russell, S Hauert, R Altman, and M Veloso. Ethics of artificial intelligence. *Nature*, 521(7553):415–416, 2015.

[189] Stuart J Russell and Peter Norvig. *Artificial intelligence: a modern approach*. Malaysia; Pearson Education Limited,, 2016.

[190] Serguei Saavedra, Rudolf P Rohr, Luis J Gilarranz, and Jordi Bascompte. How structurally stable are global socioeconomic systems? *Journal of the Royal Society Interface*, 11(100):20140693, 2014.

[191] Maximilian Schich, Chaoming Song, Yong-Yeol Ahn, Alexander Mirsky, Mauro Martino, Albert-László Barabási, and Dirk Helbing. A network framework of cultural history. *Science*, 345(6196):558–562, 2014.

[192] Shade T Shutters, Wes Herche, and Erin King. Anticipating Megacity Responses to Shocks: Using Urban Integration and Connectedness to Assess Resilience. *Small Wars Journal*, 26(3):21, 2016.

[193] Shade T. Shutters, José Lobo, Rachata Muneepeerakul, Deborah Strumsky, Charlotta Mellander, Matthias Brachert, Teresa Farinha, and Luis M. A. Bettencourt. Urban occupational structures as information networks: The effect on network density of increasing number of occupations. *PLOS ONE*, 13(5):1–14, 05 2018.

[194] Shade T Shutters, Rachata Muneepeerakul, and José Lobo. Quantifying urban economic resilience through labour force interdependence. *Palgrave Communications*, 1:15010, 2015.

[195] Shade T Shutters, Rachata Muneepeerakul, and José Lobo. Constrained pathways to a creative urban economy. *Urban Studies*, 53(16):3439–3454, 2016.

[196] Shade T Shutters, Rachata Muneepeerakul, and José Lobo. How hard is it for urban economies to become ÃćgreenÃć? *Environment and Planning B: Planning and Design*, 43(1):198–209, 2016.

[197] David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. Mastering the game of Go with deep neural networks and tree search. *nature*, 529(7587):484, 2016.

[198] Aaron Sim, Sophia N Yaliraki, Mauricio Barahona, and Michael PH Stumpf. Great cities look small. *Journal of The Royal Society Interface*, 12(109):20150315, 2015.

[199] Brian Simpson. Algorithms or advocacy: does the legal profession have a future in a digital world? *Information & Communications Technology Law*, 25(1):50–61, 2016.

[200] Roberta Sinatra, Pierre Deville, Michael Szell, Dashun Wang, and Albert-László Barabási. A century of physics. *Nature Physics*, 11(10):791, 2015.

[201] Arnab Sinha, Zhihong Shen, Yang Song, Hao Ma, Darrin Eide, Bo-june Paul Hsu, and Kuansan Wang. An overview of Microsoft Academic Service (MAS) and applications. In *Proceedings of the 24th international conference on world wide web*, pages 243–246. ACM, 2015.

[202] Adam Smith. An inquiry into the nature and causes of the wealth of nations (ed. rh campbell, as skinner, and wb todd). 1976.

[203] Annie Sneed. Moore's law keeps going, defying expectations. *Scientific American*, 19(05), 2015.

[204] Alexandra Spitz-Oener. Technical change, job tasks, and rising educational demands: Looking outside the wage structure. *Journal of Labor Economics*, 24(2):235–270, 2006.

[205] Olaf Sporns and Richard F Betzel. Modular brain networks. *Annual Review of Psychology*, 67:613–640, 2016.

[206] Oded Stark and David E Bloom. The new economics of labor migration. *The American Economic Review*, 75(2):173–178, 1985.

[207] Deborah Strumsky and José Lobo. Identifying the sources of technological novelty in the process of invention. *Research Policy*, 44(8):1445–1461, 2015.

[208] Deborah Strumsky and Jean-Claude Thill. Profiling us metropolitan regions by their social research networks and regional economic performance. *Journal of Regional Science*, 53(5):813–833, 2013.

[209] Leo Sveikauskas. The productivity of cities. *The Quarterly Journal of Economics*, pages 393–413, 1975.

[210] Mariarosaria Taddeo and Luciano Floridi. How AI can be a force for good. *Science*, 361(6404):751–752, 2018.

[211] Abraham Verghese, Nigam H Shah, and Robert A Harrington. What this computer needs is a physician: humanism and artificial intelligence. *Jama*, 319(1):19–20, 2018.

[212] Thomas L Vollrath. A theoretical evaluation of alternative trade intensity measures of revealed comparative advantage. *Weltwirtschaftliches Archiv*, 127(2):265–280, 1991.

[213] Stefanie Walter. Globalization and the demand-side of politics: How globalization shapes labor market risk perceptions and policy preferences. *Political Science Research and Methods*, 5(1):55–80, 2017.

[214] Andrew Weaver and Paul Osterman. Skill demands and mismatch in US manufacturing. *ILR Review*, 70(2):275–307, 2017.

[215] Hemantha W Wijesekera and Thomas M Dillon. Shannon entropy as an indicator of age for turbulent overturns in the oceanic thermocline. *Journal of Geophysical Research: Oceans*, 102(C2):3279–3291, 1997.

[216] Ian H Witten, Eibe Frank, Mark A Hall, and Christopher J Pal. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2016.

[217] David H Wolpert and William G Macready. No free lunch theorems for optimization. *IEEE transactions on evolutionary computation*, 1(1):67–82, 1997.

[218] Hyejin Youn, Luís MA Bettencourt, José Lobo, Deborah Strumsky, Horacio Samaniego, and Geoffrey B West. Scaling and universality in urban economic diversification. *Journal of The Royal Society Interface*, 13(114):20150937, 2016.

[219] Hyejin Youn, Deborah Strumsky, Luis MA Bettencourt, and José Lobo. Invention as a combinatorial process: evidence from US patents. *Journal of The Royal Society Interface*, 12(106):20150272, 2015.

[220] Andrew T Young and Maria Y Tackett. Globalization and the decline in labor shares: Exploring the relationship beyond trade and financial flows. *European Journal of Political Economy*, 2017.

[221] Harvey Yunis. *Plato: Phaedrus*. Cambridge University Press, 2011.