Systematic Inequalities in the Composition and Productivity of Computer Science Faculty

Samuel Frederick Way

University of Colorado at Boulder, samfway@gmail.com

Follow this and additional works at: https://scholar.colorado.edu/csci_gradetds

Part of the Computer Sciences Commons

Recommended Citation
Way, Samuel Frederick, "Systematic Inequalities in the Composition and Productivity of Computer Science Faculty" (2017). Computer Science Graduate Theses & Dissertations. 141.
https://scholar.colorado.edu/csci_gradetds/141

This Dissertation is brought to you for free and open access by Computer Science at CU Scholar. It has been accepted for inclusion in Computer Science Graduate Theses & Dissertations by an authorized administrator of CU Scholar. For more information, please contact cuscholaradmin@colorado.edu.
Systematic Inequalities in the Composition and Productivity of Computer Science Faculty

by

Samuel F. Way

B.S.C.E., University of Nebraska–Lincoln, 2010
B.S.E.E., University of Nebraska–Lincoln, 2010
M.S.E.E., University of Nebraska–Lincoln, 2012

A thesis submitted to the
Faculty of the Graduate School of the
University of Colorado in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Computer Science
2017
This thesis entitled:
Systematic Inequalities in the Composition and
Productivity of Computer Science Faculty
written by Samuel F. Way
has been approved for the Department of Computer Science

Prof. Aaron Clauset

Prof. Brian Keegan

Prof. Jordan Boyd-Graber

Prof. Kenneth Anderson

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.
Science exhibits many forms of imbalance, ranging from disparities in the representation of certain demographics in the scientific workforce to enormous variation in the quantity and quality of contributions that individuals make to the scientific literature. In this thesis, we investigate the drivers of such imbalances, seeking a greater understanding of both the factors that facilitate success in science and its potential sources of inequality or discrimination. Advances along either direction would inform policy decisions aiming to support scientific discoveries and the scientists who make them. Progress in these directions, however, is typically impeded by the complex nature of the processes that govern who works in science, where they work, and how productive they are. Specifically, interdependencies among these processes complicate any analysis attempting to isolate and quantify particular effects. Here, we use techniques from statistical modeling, machine learning, and causal inference to directly address these sources of complexity and explore the underrepresentation of women in science, the role of productivity in faculty hiring and retention, and how institutional prestige affects researchers’ success.

Computer science in many ways represents an ideal case study for investigating sources of imbalance in academia. Throughout the field’s history, women have been dramatically underrepresented, despite increasing participation in recent years. Research in computer science is also remarkably diverse, with varying scholastic traditions and rates of publication, and is incredibly well-documented, offering rich sources of data to investigate the drivers that sustain the field’s gender imbalance and disparities in research output. Our work therefore focuses on computer science in particular, however our findings have broad implications to the scientific community as a whole.
Dedication

To my family.
Acknowledgements

I would like to thank my dissertation committee, Jordan Boyd-Graber, Brian Keegan, Raf Frongillo, Ken Anderson, and Aaron Clauset, along with my preliminary committee members, Jim Martin and Qin Lv for their invaluable support and guidance. I am especially and forever grateful for the many hours of advise and countless opportunities that my advisor, Aaron, has given to me over the years. Working with Aaron has been immensely rewarding and perhaps just as importantly, a lot of fun. I also want to thank Dan Larremore, who through our collaborations and shenanigans together became like a second advisor to me and a dear friend. Finally, I would like to thank Liz Bradley for many thoughtful conversations, an open door, and goodies from the garden.

I also want to thank the entire Clauset Lab for always inspiring and challenging me to be a better scientist. I’m especially grateful to Abbie Jacobs for her friendship, mentoring, and for teaching me how to city, and to Allie Morgan for fun and fulfilling collaborations. I am also thankful for the continued support of the Biofrontiers Institute, especially Andrea Stith and Thomas Cech. I would like to thank my IQ Biology family, too, particularly Joey Azofeifa and Ryan Langendorf for many nights spent discussing science under the stars, over banjos and whiskey.

I want to thank all of my friends and wonderful family, especially my parents for their unwavering support throughout this and every endeavor. And, lastly, I would like to thank Laura, who for the last twelve years has stood by my side, encouraging me to be a better person, to have fun, and to smile. She is in every way, my everything, and I cannot begin to express how grateful I am for all of her sacrifices and support, her love, and her friendship.
Contents

Chapter

1 Background and Introduction 1

1.1 Early Sociology of Science 2

1.2 Processes Affecting the Composition of the Scientific Workforce 4

1.3 Processes Affecting the Productivity of the Scientific Workforce 7

2 Gender, Productivity, and Prestige in Computer Science Faculty Hiring Networks 12

2.1 Data and Features 15

2.2 A model of the faculty market 22

2.2.1 Model checking 24

2.3 Results 24

2.3.1 Market-level analysis 25

2.3.2 Institution-level analysis 27

2.3.3 Candidate-level analysis 29

2.3.4 Long-term forecast for gender parity 30

2.4 Discussion 31

3 The misleading narrative of the canonical faculty productivity trajectory 37

3.1 Results 40

3.1.1 General Trends in Productivity 40

3.1.2 Individual Productivity Trajectories 43


3.1.3 Transitions in Authorship Roles ............................................. 51

3.2 Discussion ................................................................. 52

4 The Effects of Departmental Prestige on Researcher Productivity and Prominence 57

4.1 Data ................................................................. 59

4.2 Results ................................................................. 61

4.2.1 Location’s Causal Effects .................................................. 61

4.2.2 Mechanisms Behind Employer Prestige .................................. 67

4.3 Discussion ................................................................. 73

5 Conclusions and Future Directions ............................................ 77

5.1 Conclusions ............................................................... 77

5.2 Future directions .......................................................... 79

5.2.1 The role of social networks in faculty hiring ......................... 80

5.2.2 A generative model for career productivity .......................... 81

5.2.3 Better measures of impact ............................................... 83

5.2.4 Hedgehogs and Foxes .................................................. 84

Bibliography ............................................................................. 86

Appendix

A Supplementary Material for Chapter 3 ...................................... 94

A.1 Collection of CV data ......................................................... 94

A.2 General trends in productivity data ....................................... 95

A.3 Modeling framework ........................................................ 101

A.3.1 Adjusting models instead of counts .................................. 101

A.3.2 Poisson Model .......................................................... 102
A.3.3 Negative Binomial Model .............................................. 103
A.3.4 Modeling outlook ....................................................... 105
A.4 Sensitivity to timing of publications ................................. 105
A.5 Model selection ............................................................. 107
A.6 Detection of alphabetized publication venues ..................... 109
A.7 Least squares fit of $f(t)$ .................................................. 110

B Supplementary Material for Chapter 4 ............................... 112
B.1 Measuring subfield similarity ......................................... 112
Women and men move up in the prestige rankings at similar rates (excluding self-hires.)

Median z-scores by gender and by whether a faculty moved up or down the ranking for their faculty position. We find no significant differences comparing men and women’s productivity scores in each of these categories. Median values are negative indicating that productivity scores are right-skewed due to prolific faculty.

For individuals graduating and being hired in separate geographic regions, men are significantly more likely to be moving up the ranks ($p = 0.01, \chi^2$).

For individuals with postdoctoral experience, men and women move up the ranks at similar rates ($p = 0.92, \chi^2$).

Network summary statistics used in model checking of uniform, step, and logistic choices of $f$. In each row, boldface indicates the model that best reproduces that characteristic of the observed network.
Figures

Figure

2.1 For the 2659 computer science faculty in our sample (collected in 2011), the distribution of years in which they were first hired as an assistant professor.

2.2 Reduction of modeling error as features are added to the model. Percent reductions are computed relative to the step function model as a baseline. Median percent reductions are reported for each model, and attributes producing a significant reduction in error ($p<0.05$, Mann–Whitney) are marked with braces and asterisks.

2.3 Three examples of model-based sampling of university-specific female hire distributions. Each green trajectory denotes the cumulative number of hires for a single simulation of the placement model at the indicated university. Running many simulations creates the distribution over final counts, shown on the right. The actual trajectory of hires made by the institution (within the data set) and the resulting final count are highlighted in black. UC Berkeley, Princeton, and Brigham Young represent examples of expected, female-skewed, and male-skewed hiring, as indicated by the location of the actual value within each sampled distribution.

2.4 Comparison of actual and expected female hiring over the top 50 institutions. Dots represent actual values minus expected values calculated from distributions samples as in Fig. 2.3. The shaded region denotes the 25th-75th percentiles, based on modeling outcomes. Six particular universities are annotated. Top 10 schools hire slightly above expectations while ranks 11–20 hire below expectations. This suggests that the efforts by the highly-ranked schools to rectify any gender imbalance may have impeded the efforts of lower-ranked schools hoping to do the same.

2.5 Mean placement error by year. Placement error is computed as the difference between the rank of the institution where the person was hired and the rank of the institution where they placed under simulation. Higher variance in female placement error is within fluctuations expected due to lower female representation in the data set. Adjusted for yearly representation in the data, error is neither systematically increasing nor decreasing in time.

2.6 Time series of the fraction of assistant professor hires since 1970 in our dataset that are women (green; with 95% confidence intervals around the mean), and the fraction of computer science doctoral recipients since 1970 that are women (black).
2.7 Gender ratio of assistant professors in computer science, by gender, and a projection for when gender parity will be reached. If the historical trend continues unaltered, gender parity will occur in approximately 2075. Shaded regions represent extrapolated 95% confidence intervals from an ordinary least squares regression.

3.1 Publications correlate with prestige of employing institutions. Dots indicate median number of publications per person per institution for researchers’ first ten years post-hire, adjusted for growth in publication rates over time (see Supplemental Fig. A.3 and Supplemental Text A.2) and ordered by institutional prestige, π. Effects of prestige are similar for private (black) and public (orange) institutions (p = 0.146, t-test; see text), increasing at a rate of nearly 2.7 publications per 10-rank increase in prestige. Shaded region denotes the 95% confidence interval for least squares regression.

3.2 Average publications follow conventional narrative across prestige. Three average productivity curves are shown, stratified by the prestige rank π of researchers’ employing institutions as indicated. Averages over researchers at all levels of institutional prestige follow similar productivity trajectories, in agreement with the conventional narrative, but at differing scales of output.

3.3 Example trajectory and piecewise model. Dots represent empirical annual publications. Orange line shows best fit of piecewise linear model [Eq. (3.1)] with slopes m1, m2, change point t∗, and intercept b annotated.

3.4 Distribution of individuals’ productivity trajectory parameters. Diverse trends in individual productivity fall into four quadrants based on their slopes m1 and m2 in the piecewise linear model Eq. (3.1). Subfigures show example publication trajectories to illustrate general characteristics of each quadrant. The shaded triangular region corresponds to the conventional narrative of early increase followed by gradual decline. Sensitivity of each individual’s quadrant to noise in publication years is represented by color and size: orange and gray dots represent trajectories with stable (74.6%) and unstable classifications (25.4%), respectively, and sizes are proportional to the fraction of stable trials; see Supplemental Text A.4. Counts and percentages tabulate the number of stable trajectories falling into each quadrant, relative to the total.

3.5 Heat map of researchers’ inferred change points. Each researcher’s inferred change-point parameter t∗ is plotted as a heat map, sorted by the length of their career in our dataset and restricted to individuals whose productivity trajectories are both stable under the addition of noise (see text) and better modeled by Eq. (3.1) than a straight line, determined by AIC (see Supplemental Text A.5).
3.6 Heat map of researchers’ most productive years. Each researcher’s most productive year (empirically; not model fit) is plotted as a heat map, sorted by the length of their career in our dataset. White box indicates researchers with fewer than 10 years of experience, whose most productive year is necessarily early. The marginal distribution (right) shows the empirically most productive year for all faculty in the dataset, separated by early career (first 10 years; grey) or later career (orange). The most common peak-productivity year is year 5, and only about half of senior faculty exhibit peak productivity in year 5 or earlier... 50

3.7 Early-career transitions in authorship roles. The average proportion of first-author (squares) and last-author papers (triangles) as fraction of total, as a function of career age, split by top-50 institutions (orange) and others (black). Shaded regions represent 95% confidence intervals for each proportion. Single-author publications are counted as first-author publications. On average, researchers at top-50 institutions transition more quickly into senior-authorship roles, compared to researchers at less prestigious institutions... 51

3.8 First-author publication rates. First-author publications as fraction of total in the first three years post-hire, and the three years thereafter, shown separately for researchers who placed at a top-50 institution (top) and researchers placing outside of the top-50 (bottom). Individual researcher data are plotted as points on top of a corresponding heatmap in which darker color denotes higher density by Gaussian kernel density estimation. Researchers at all levels of prestige tend to move out of first-authorship roles during this period, though researchers at top-50 institutions transition more completely by years 3–5 than others... 53

4.1 Early-career productivity is indistinguishable for similarly-placing but differently-trained peers: Average yearly productivity differences for 359 matched pairs of faculty. Positive values indicate higher productivity for graduates of more prestigious institutions, relative to matched faculty. Shaded region denotes 95% confidence interval for the mean. Year $y = -3$ displays a significant difference in favor of more prestigious graduates, and year $y = -1$ shows a significant difference in favor of less prestigious graduates... 63

4.2 Early-career prominence is indistinguishable for similarly-placing but differently-trained peers: Average yearly prominence (citation count) differences for 129 matched pairs of faculty. Positive values indicate greater prominence for graduates of more prestigious institutions, relative to matched faculty. Shaded region denotes 95% confidence interval for the mean. We find no significant differences in any of the years shown... 64

4.3 Early-career productivity advantages quickly emerge for prestigious faculty: Average yearly productivity differences for 194 matched pairs of faculty. Positive values indicate higher productivity for faculty at more prestigious institutions, relative to similarly-trained faculty. Shaded region denotes 95% confidence interval for the mean. Productivity differences appear after only one year following initial placement... 65
4.4 Early-career prominence advantages conferred to prestigious faculty: Average yearly prominence differences for 97 matched pairs of faculty. Positive values indicate greater prominence for faculty at more prestigious institutions, relative to similarly-trained faculty. Shaded region denotes 95% confidence interval for the mean. Significant differences emerge three years following initial placement and continue to rise in the years thereafter. .................................................. 66

4.5 Pre-hire productivity correlates with prestige: Total publications in the five years pre-hire versus the prestige of the individual’s initial appointment. Black line denotes ordinary least squares regression, with slope indicating that new faculty are on average 0.28 papers more productive for every 10-rank increase in prestige. .... 69

4.6 Pre-hire prominence correlates with prestige: Total citations received in the five years pre-hire versus the prestige of the individual’s initial appointment. Black line denotes ordinary least squares regression, with slope indicating that new faculty receive on average 18.59 more citations for every 10-rank increase in prestige. .... 69

4.7 Faculty leave their appointments at similar rates across the prestige hierarchy: Binned by departmental prestige, faculty leave their initial appointments, either for other institutions or out of academia altogether, at similar rates. Error bars indicate 95% confidence intervals around the means. Dotted purple line indicates the total fraction of department faculty from all 205 institutions. ......................... 72

4.8 Faculty who left academia tended to be less productive: Post-hire productivities as z-scores (relative to others in the same department) for all junior faculty in the follow-up sample. Purple dots highlight individuals who were no longer employed as tenure-track faculty at one of the in-sample institutions in 2016. ................. 73

4.9 Faculty relocation often reverses initial rank differences. Faculty who relocate tend to undo the rank change incurred by their initial placement, returning to an institution similar in rank to where they received their doctorate. This reversal of rank-changes is similar for individuals placing in the top-50 (purple) and otherwise (black). ..... 74

5.1 Random walk models capture aggregate productivity trends: Using normally-distributed step sizes, linear attenuation, and a strict minimum threshold for productivity, random walk models recapitulate aggregate trends for career productivity. ............. 82

A.1 DBLP coverage improves for more recent publications. Fraction of all publications found in DBLP data compared to publication lists extracted from CVs of corresponding researchers, separated by year. Regression of these fractions reveals that DBLP coverage improves by approximately 1.06% each year. Shaded region denotes the 95% confidence interval for the regression. ......................... 96

A.2 Individual productivity has increased over time. After adjusting for DBLP coverage (see Supplemental Text A.2), evaluation of average individual performances in select indicator years reveals that researchers have become more productive over time, growing at a rate of approximately one additional paper per decade. Shaded regions denote the 95% confidence interval for each regression. ................. 97
A.3 Annual publication rates have grown steadily. For faculty in this study, per-person annual publications have increased over time at a rate of approximately one additional paper every 10 years. This rate of growth affects researchers at all levels of prestige rank $\pi$ [19]. Slopes represent least-squares linear regressions, with shaded regions denoting corresponding 95% confidence intervals.

A.4 Publications correlate with prestige of employing institutions. Dots indicate median number of publications per person per institution for all years post-hire, adjusted for growth in publication rates over time and ordered by institutional prestige. Effects of prestige are statistically indistinguishable for private (black) and public (orange) institutions. Shaded region denotes the 95% confidence interval for least squares regression.

A.5 Production imbalance among faculty. Lorenz curves of adjusted production by in-sample faculty, stratified by decade of first hire (see legend), show that approximately 20% of faculty account for half of all publications in the dataset. The Gini coefficient $G$ for each curve is noted in the legend; diagonal line indicates equal production by all faculty.

A.6 Example model fits for noise-free and noise-added publication trajectories. Example publication trajectories (gray, dotted) with piecewise linear fits (black) and 200 fits to noise-added trajectories (orange). Trajectories are categorized as stable whenever 75% or more of fits fall within a single quadrant, as indicated in inset diagrams.

A.7 Trajectory shapes are robust to perturbations in publication years. Applying the piecewise linear model to 200 noise-added publication trajectories for each researcher, the median fraction of trials resulting in sign changes of model parameters $m_1$ or $m_2$ compared to noise-free fits is 0.045.

B.1 Subfield similarities for all faculty: Distribution of Jensen-Shannon divergence scores for all possible pairs of faculty in our dataset. The shaded (purple) region denotes scores within the 90th percentile of similarity (see text), and only faculty pairs within this range were included in our matched pair analyses.
Chapter 1

Background and Introduction

The science of science is a new and actively growing field of study that encompasses all forms of social science research seeking a greater understanding of the progression of science and the scientific workforce. Subfields in this area include the history, psychology, sociology, and even philosophy of science, which collectively explore the underlying mechanisms that drive science forward, with the goal of identifying opportunities to make scientific research more efficient and its associated communities more inclusive. The sociology of science in particular seeks a greater understanding of the fundamental social processes within science: (1) the exchange of scientists and processes that shape the composition of the scientific workforce; and (2), exchanges amongst scientists, most notably scientific collaboration and publication, processes that determine the productivity of the scientific workforce. This dissertation details contributions to each of these two key areas.

As an additional focus, the included studies explore potential drivers contributing to the persistent underrepresentation of certain demographics, primarily women, in STEM fields. Computer science, as detailed in future chapters, is in many ways an ideal case study for addressing these issues and is therefore the focus of the presented work. In short, women are dramatically underrepresented at all levels of the field and account for only 16%, or one in every six of its tenure-track faculty. Further, computer science, from the viewpoint of individual faculty is an extremely well-documented field with publicly-available records of publication, citation, and training histories available for most researchers, allowing for detailed studies of individual-level performances.

Our investigations into the underrepresentation of women in computer science reflect a much
larger theme that pervades the entire history of the sociology of science and played a major role in precipitating the field’s formation. That theme is one of imbalance and applies to many facets of science including who participates, how productive they are, and how much recognition they receive for their work. The goal of this dissertation is therefore to disentangle the complex processes that give rise to imbalance in each of these facets and ask, to what degree do these imbalances potentially reflect or contribute to inequality or unfair treatment of certain researchers. This distinction is a subtle one but is nevertheless exceedingly important: imbalance itself is not evidence of inequality, and, similarly, the absence of imbalance is not necessarily evidence of equality. We proceed with this distinction in mind and look for opportunities to inform future scientific policies that might enhance the efficiency of the scientific workforce.

The desire to make the scientific community more welcoming of women and other underrepresented groups stems from a variety of anticipated benefits, not the least of which is a culture of fairness and acceptance. Modern studies have shown that science increasingly relies on the collaboration of individuals working as teams [115, 114, 80] and that diverse teams are often more creative, more productive when tackling complex problems as a result of leveraging varied approaches and perspectives [83]. Promoting diversity within the scientific workforce therefore stands to benefit not just those who work within science but society as a whole through the products and insights that grow out of the creation of scientific knowledge.

In the sections that follow, we introduce key works from the history of the sociology of science. By highlighting these past studies, we provide context for how contributions made by this dissertation advance the state of the field and inform policy decisions moving forward. We begin by discussing the earliest works in the sociology of science before turning to foundational studies in each of the two aforementioned key areas of research.

1.1 Early Sociology of Science

The earliest works in the sociology of science (and many thereafter) sought to describe and understand extreme imbalances observed in the production of scholarly articles and the awarding
of credit or recognition to their authors. The late Robert Merton is credited by many as the founder of the field and made a number of great contributions, including several detailing processes of accumulative advantage in science. In his seminal 1968 work [72], Merton coined “The Matthew Effect” – a concept and terminology that has since spread to all of sociology and to other fields – to describe how established scientists are more likely than their less-established colleagues to receive disproportionate amounts of credit for their work, further strengthening their reputations as scientists and thus the likelihood that they will continue to be successful and receive even more credit in the future. The phrase, as Merton described, comes from the gospel of Matthew, which says, “for unto every one that hath shall be given, and he shall have abundance: but from him that hath not shall be taken away even that which he hath.” Others have since summarized Merton’s observation as “the rich get richer” phenomenon.

Merton’s theory stemmed in large part from his graduate student (and eventual wife) Harriet Zuckerman’s interviews with Nobel prize recipients[120], in which several laureates suggested that “the man who’s best known gets more credit, an inordinate amount of credit.” Zuckerman found that this effect persists regardless of where the most famous collaborator appears in the author list [122], and disproportionately affects women more than men. In fact, Zuckerman’s lack of credit in the discovery of the Matthew Effect has itself become a canonical example for another form of bias in science: The Matilda Effect [94], which describes the practice of misallocating credit for scientific discoveries made by women to their male colleagues.

While Merton is credited for creating the sociology of science domain, his work built upon earlier studies, which noted the imbalance in productivity and creativity of scientists. Of note, Alfred Lotka, the now-famous bio-mathematician, developed the inverse square law of scientific productivity to describe the relationship that he saw between scientists and their scholarly production. The law is lauded as the first law of scientometrics – a sub-discipline within the sociology of science that focuses on quantitative analyses of scientific knowledge production – and states that the number of scientists who have written \(N\) papers is equal to \(1/N^2\) of the number of scientists who have written one paper [68]. This work has since inspired many studies in the fields of sci-
entometrics and bibliometrics and also serves as one of the earliest observations of a heavy-tailed distribution in the social sciences.

In terms of creativity imbalance, early works by Harvey Lehman [56] noted that, disproportionately, important scientific discoveries are made by young scientists. Lehman’s work represents an early effort to distinguish between quantity and quality of researchers’ scholarly output, a distinction that will be discussed in more detail in following sections. Lehman’s observation was later criticized by a number of researchers [31, 25] who pointed out that the bulk of scientific discoveries “are indeed made by young scientists because most scientists are young, not because age has a causal influence on scientific creativity” [25]. This scientific exchange serves to underscore the importance of proper controls and statistical rigor in the sociology of science. Due to the complex nature of human interaction and its inherent biases, researchers must be extremely careful not to draw false conclusions based on what can often be very compelling correlations.

1.2 Processes Affecting the Composition of the Scientific Workforce

Most studies examining imbalance in the composition of the scientific workforce have focused on processes contributing to the underrepresentation women in science, particularly in STEM fields. Ethnic minorities are generally regarded as a “statistical rarity” [66], which has historically precluded their analysis. It is thought, however, that women and ethnic minorities face similar obstacles in their careers, such that progress made to improve the academic environment for one group will, hopefully, have a similarly positive impact on the other. In this section we discuss several influential works investigating the recruitment and hiring of women into scientific careers, highlighting key insights and takeaways that have influenced and inspired the contributions made by this dissertation.

The underrepresentation of women in science is frequently explained using the metaphor of

---

1 We acknowledge that this attitude, that the underrepresentation of these individuals in science must at least partially fix itself before it can be studied in detail, sets a dangerous precedent. In our future work, we look to apply the analyses developed in this dissertation to explicitly study individuals under the NSF’s classification of underrepresented minorities (URMs) [105].
a “leaky pipeline.” The metaphor suggests that there is no single cause for the relatively small number of women in the professoriate, but, instead, the academic career pipeline contains a series of leaks, occurring primarily at career transitions (e.g. from undergraduate student to graduate student, graduate student to junior faculty, etc.) and ultimately giving rise to the large imbalance observed at its endpoint. Definitions of the pipeline vary, but many suggest it begins as early as K–12 [110], and, crucially, the metaphor has been leveraged to develop and reform education and departmental policies for several decades now [11, 73]. Over the years, sociology of science researchers have investigated many of the pipeline’s perceived leaks in efforts to understand and mend issues leading to women’s departures.

In 2005, Clark Blickenstaff [11] published an exhaustive review article in which he detailed the most commonly-stated explanations for why women leak out of the scientific pipeline, providing a summary of the literature associated with each. Blickenstaff’s list includes nine common explanations, which can be summarized as conceptions either regarding the inherent abilities and aptitudes of women (including biological differences and disparities in self-assessment) or societal factors (including workplaces biases and stereotypes). After reviewing the research behind each of the nine explanations, Blickenstaff concluded that the factors which represent normatively justifiable, universalistic characteristics are unsupported by research, and, as a result, the pipeline analogy dangerously purports a situation in which many women naturally depart or simply opt out of scientific careers. Instead, he argued, these factors represent a cascade or layers of sex-based filters that remove women from the scientific workforce.

Blickenstaff’s sentiments towards the pipeline metaphor and its (in-)ability to inform effective educational policies are underscored by anemic growth in the numbers of underrepresented groups in science, this despite the millions of dollars invested into various attempts to broaden their participation [81, 12]. The ineffectiveness of the policies and interventions designed to correct the imbalance in the scientific workforce point to a fundamental lack of understanding of the processes that drive women and other underrepresented minorities out of science. Reflecting on past failures, we reach two conclusions: first, we must rethink our approaches to studying the social processes
in science and adopt strategies that reflect both the complexity of the underlying problems and their interrelatedness; second, we must make concerted effort to quantify the amount of impact that each complicating factor contributes to imbalance. Together, these facets will elicit a greater understanding of not only how to fix issues in science but whether fixing them will result in the magnitude of change we hope to accomplish and on what timescales.

One particular leak in the academic pipeline that has received a great deal of attention in recent years is the transition between graduate school into the professoriate and corresponding processes surrounding academic faculty hiring. Much of this attention originates from a handful of recent studies reporting that the STEM pipeline no longer leaks [15, 74]. Upon publication, these studies made national headlines [8] and were widely circulated on social media, but the analyses employed by these studies and the conclusions drawn from them raise concerns amongst sociology of science researchers.

In 2015, a paper by Williams and Ceci [113] concluded that not only has leaking slowed at the faculty hiring transition, preferences expressed in national hiring experiments reveal that women hold a two-to-one advantage in considerations for STEM tenure-track faculty positions. This advantage was calculated through a series of hiring experiments in which study participants were asked to rank and rate curricula vitae belonging to equally-qualified candidates who differ only by gender. Overwhelmingly, the authors found, study participants indicated that they would hire the female applicant over the male.

While Williams and Ceci’s results suggest a positive environment for women in science, several aspects surrounding the study’s experimental design cast doubt on the validity of its conclusions. The primary concern regarding this study is whether or not the outcomes of its experiments are in any way representative of actual hiring outcomes. First, evaluating individuals’ curricula vitae represents only one aspect of faculty hiring, alongside interviews and formal presentations. Second, by creating hypothetical applicants and providing research descriptions rather than actual lists of publications, the study’s designers may have fundamentally changed how participants reviewed the applicants. Third, the experiments represent, at best, an over-simplified simulation of faculty hiring
by asking participants to choose between three artificial applicants, one of whom possessed slightly-weaker qualifications than the competing male-female duo of applicants with virtually identical qualifications. Real-world hiring situations, by contrast, can involve dozens of candidates who are often similarly qualified but in vastly different areas of study, complicating the search committee’s final decision.

We note these concerns in order to highlight the difficulty – and, perhaps in some cases, practical impossibility – of crafting traditional sociological experiments that faithfully replicate complex social processes and decisions like academic faculty hiring. Moreover, we caution that the practice of drawing broad-brush conclusions from such experiments risks perpetuating the very problems that studies like this one seek to address. Instead, looking forward, we find reason to advocate for the analysis of actual outcomes of social processes and, where possible, to look for natural experiments that allow for causal inferences to be drawn from real-world data. Such opportunities have become more common in recent years and will continue in the future as digital records detailing the outcomes of processes like faculty hiring, retention, and promotion appear online, opening the door for data-driven, quantitative analyses.

1.3 Processes Affecting the Productivity of the Scientific Workforce

In the previous section, we highlighted key studies in the research of social processes affecting the composition of the scientific workforce, noting important lessons to be learned from each. These studies shed light on the factors that influence who, ultimately, works in science and where they find employment. We now shift our focus away from processes that determine where people end up working in science and, instead, examine those that affect how productive researchers are once they get there.

Scholarly productivity or the creation and subsequent dissemination of scientific knowledge is, without question, the single most important aspect of a scientist’s career. The goal of being a scientist is, after all, to contribute to the larger body of scientific knowledge, building on previous studies in order to refine and expand our understanding of the world around us. And, like any
career, scientists need to be evaluated in order to determine who should be hired, promoted, and recognized for their achievements. Publications, which serve as the outputs of science, thus provide a natural means for assessing an individual’s effectiveness as a scientist, and, indeed, researchers’ curricula vitae serve as portfolios of their work and are dutifully inspected at every important stage in the academic career trajectory.

In the ideal case, evaluations of a researcher’s productivity should provide an objective assessment, measuring the importance of their contributions to science. In practice, objectively quantifying the quality of a researcher is an immensely difficult task. A great variety of measures have been proposed [53], but despite these efforts, there is little consensus regarding how to properly evaluate scientists. Generally speaking, attempts fall into two categories: those measuring the quantity or sheer volume of a researcher’s contributions, and, separately, those measuring the quality or impact of a researcher’s contributions. For measures built up from quantity, differences amongst various approaches generally stem from decisions surrounding what counts. For quality-based metrics, differences stem from decisions of what matters.

Deciding which of these metrics is “best” is a futile task. The “best” metric depends entirely on the context to which it is being applied. That is, for a given task, the best or most appropriate metric is the one that best solves the task at hand. Few tend to disagree, however, that quality should matter more than quantity: a smaller number of highly influential papers is more important than a larger number of non-influential papers. But there is a problem with quality measures like those built up from citation counts. The problem is that it often takes time (and, occasionally, a lot of time [109]) for the impact of a study to become apparent, for the corresponding article to accrue citations and become central to a field’s identity.

The latency with which importance is realized, then, does not lend itself well to assessments that require a timely resolution. Central to the focus of this dissertation, faculty hiring, retention, and promotion are all processes whose timelines limit the usefulness of many quality-based metrics. As a result, while quality measures are important in the long-term, these decisions are more likely to be based on the quantity of a researcher’s work, particularly outside of elite institutions [26].
For this reason, the rest of this section and indeed the remainder of this dissertation will look at productivity in terms of the volume of articles produced by a researcher and not on some measure of recognition that these articles accrue over time.

Many studies in the past have examined imbalances in the production of scholarly articles. Again, we note that some of the oldest works in the sociology of science were in fact focused on this very subject [68]. One particularly common theme in this area of research, though, is the production imbalance between male and female researchers in science, specifically the observation that men tend to publish more articles than women.

In 1984, Cole and Zuckerman published a study in which they describe the gap in scholarly production between men and women as “the productivity puzzle” [24]. This phrase was subsequently adopted throughout the research community, and, for years, researchers attempted to solve this puzzle and explain sex differences in research productivity. A noteworthy answer was published in 1998 by Xie and Shauman, whose analysis built on many previous studies in this area and demonstrated that, after accounting for known factors that affect men and women differentially, productivity differences become negligible.

In their study, Xie and Shauman evaluated productivity as the raw count of publications produced. They elected to evaluate short-term differences in the productivity between men and women, noting that women’s relatively recent and gradually increasing participation in science means that they are at a distinct disadvantage, judging by long-term measurements of productivity. As is common in this type of study, the authors utilized a multivariate regression framework to assess the role that gender and other factors played in shaping individual productivity. Unlike other studies, however, the authors leveraged survey responses gathered for men and women in the sample, allowing them to incorporate information about the researchers’ personal attributes like access to research resources and their marital status, in addition to institutional attributes.

What Xie and Shauman found was that, not only had productivity differences diminished since previous studies, those differences went away almost entirely when controlling for personal and environmental factors. Unfortunately, the authors lacked longitudinal data that would have
allowed them to study productivity’s effect on their subjects’ promotion and retention, but, they noted that productivity differences lessen as rank increases, an effect that the authors acknowledge could be due to selectivity or filtering. Finally, the authors conclude that, as a result of their work, we now know that differences in productivity are accounted for by differences in personal characteristics, structural positions, and facilitating resources. What is not clear, the authors state, is why men and women differ systematically along these axes to begin with. In this dissertation, we hope to elucidate how women’s past productivity, their environment, and other factors affect their long-term productivity and therefore their longevity in the scientific workforce.

In addition to investigations of sex differences in research output, several sociology of science studies have focused on another important factor that drives scholarly productivity: the researcher’s academic environment. In the Xie and Shauman study discussed above, the authors found that differences in institutional properties and access to resources were at least partly responsible for differences in productivity between male and female researchers. However, exactly which attributes and how they contribute to this discrepancy was not discussed. Allison and Long published a study in 1990 that investigated these effects in more detail [2], specifically examining how an institution’s prestige affects the productivity of its researchers.

Allison and Long’s study was motivated by previous observations [10, 23] that researchers at top-ranked departments tend to be more productive than researchers and lower-ranked departments. The question that these the authors wondered was: are these prolific faculty productive because they work at a prestigious institution, or do they work at a prestigious institution because they are good, productive researchers. The authors introduced these two explanations as the “departmental effects hypothesis” and the “selection hypothesis,” noting that the two theories are not mutually exclusive. Under the departmental effects hypothesis, researchers who find employment in top-ranked schools enjoy a variety of benefits, including first-class facilities and intellectual stimulation with outstanding peers. The selection hypothesis, on the other hand, suggests that researchers at these institutions have been recruited by top departments because, correctly, they anticipate the researcher will be successful.
To investigate these two hypotheses, Allison and Long analyzed the productivity trajectories of 179 scientists who changed from one academic institution to another between 1961 and 1975. For large changes in department rank, the authors found that researchers who moved up the rank became both more productive and more highly cited, and, similarly, researchers who moved down became less productive and less well-cited. The number of subjects in this study is rather small, but the results lent support to the authors’ departmental effects hypothesis, they felt. In regards to the selection hypothesis, Allison and Long suggested that, should the selection hypothesis be in place, one might expect to see an upward trend in the productivity trajectories of scientists just before relocating to a prestigious department. The authors noted no such trends in the data.

Allison and Long’s study presents one of the most compelling pieces of evidence to date regarding the impact of institutional prestige on individual researchers’ productivity and, specifically, that being in a top-ranked department makes researchers more productive. The authors’ data limited their ability to make stronger claims, and the study’s design did not attempt to infer causality beyond providing encouraging evidence for the departmental effects hypothesis. The study itself illustrates a common issue that has historically plagued sociology of science research: limited data. In recent years, publication records are available through a variety of online sources, and data on the education and academic appointment histories [19] has also been assembled. In later chapters of this dissertation, we combine these rich sources of data and revisit Allison and Long’s hypotheses.

We begin by investigating faculty hiring in computer science and its effects on the composition of the field’s workforce. As will be discussed, the gender ratio in computer science is very much unbalanced, with women making up only a very small fraction of the professoriate. The goal of our study then will be to understand the extent to which gender discrimination shapes hiring decisions and potentially contributes to women’s persistent underrepresentation in the field. As part of this study, we will also explore how pre-hire productivity impacts placement outcomes and thus the composition of the workforce. The interplay between these two facets of science, the composition and productivity of its researchers, will be a constant theme throughout this and future chapters.
Chapter 2

Gender, Productivity, and Prestige in Computer Science Faculty Hiring Networks

Portions of this chapter are adapted from:


This work was presented in April 2016 at the International Conference on World Wide Web in Montréal, Québec.

Women continue to be dramatically underrepresented in computer science, receiving only 18% of bachelors’ degrees and 20% of doctorates in 2011[1] and are estimated to hold fewer than 20% of technical positions in the computing industry[2]. Women are especially underrepresented in the professoriate, making up only 15% of tenured or tenure-track faculty in computer science departments [19]. Understanding the causes of gender imbalance in faculty hiring would illuminate the underlying social processes that shape academic disciplines, and facilitate efforts both to support equal opportunities and to address the many non-meritocratic differences in male and female faculty experiences [35, 49, 97]. These differences include disparities in tenure rates, competency evaluations, remuneration, allocation of research facilities, and grant competitions. Rectifying these differences and improving the gender balance in computer science would serve not only to advance...

social justice but would also promote the sort of diversity in skills and research approaches that has been found to improve group performance [83], particularly in innovation-focused industries [51].

Much of the past research on gender imbalance among faculty has focused on the “leaky pipeline,” the name given to the observation that women leave science, technology, engineering and mathematics (STEM) fields at greater rates than men at every stage of an academic career, from grade school to full professor [45]. At the faculty hiring stage of the pipeline, several experimental studies have aimed to identify the causes of gender imbalance [77, 15, 113]. However, these have yielded inconsistent, even contradictory findings, and little past work has focused specifically on computer science.

Essentially, faculty hiring is a community-based competitive process of subjective expert evaluations under conflicting and evolving preferences; that is to say, it’s complicated. These features, along with the non-independent nature of hiring outcomes, make it difficult to reliably assess the presence and source of real biases. Here we investigate the role of gender in faculty hiring in computer science using a novel network model of the hiring process itself, across institutions and time. We then use this model to study the hiring histories of individual institutions and the experiences of individual faculty. We train this model using comprehensive data on the hiring outcomes, scholarly productivity, and gender of 2659 tenured or tenure-track faculty across all 205 computer science Ph.D.-granting departments in the United States and Canada [19].

Many studies have found evidence of gender bias in academia. For instance, male faculty in the life sciences tend to train fewer female graduate students and postdocs, relative to female representation in the pool of trainees [96]. This tendency is more pronounced at elite institutions, which tend to produce the majority of future faculty [19]. Women often perceive greater barriers to becoming faculty than do men [106], which may discourage them from seeking faculty jobs at all. Both grant proposal and peer review success rates can be higher for men than for women, because of implicit biases in the evaluations of the competence of women [49, 108]. Technical disciplines, including computer science, often have a normative expectation of intellectual brilliance, and in these fields women are less likely than men to seek doctoral degrees [58]. Experiments using name
and gender variations on resumes have found that both male and female faculty members tend to rate male applicants as more competent, more hireable, and worthy of more mentoring than female applicants [77]. Taken together, it appears reasonable to expect strong and pervasive evidence of gender bias in faculty hiring outcomes across computer science.

Other studies have argued that the evidence of bias is lacking, even if it may have existed in the past. For instance, a review of 30 years of research on the leaky pipeline found that while gender differences were substantial prior to the 1990s in STEM fields, the gap has since closed [74]. A separate review article surveyed literature on mathematical abilities in children, attitudes toward math-intensive fields, and access to, persistence in, and remuneration for faculty, concluding that no evidence of systematic gender bias exists today [15]. One recent study controversially claimed to find a 2-to-1 preference for female applicants over male applicants in STEM tenure-track faculty positions, based on a hypothetical hiring scenario [113]. However, the experimental design did not include applicant publications, presentations, or reference letters, and thus it is unclear the degree to which these results reflect real preferences, aspirations, or political correctness. Even if the evidence is real, identifying its cause remains difficult. For instance, some studies argue that the critical variable underlying female underrepresentation is not gender itself but differences in personality [27] and structural position [117]; better access to resources for hiring, reviewing, and publishing [24, 117, 16]; or the lower likelihood of workplace sexual harassment [48], that happen to correlate with being male.

The role of gender in shaping outcomes in faculty hiring is difficult to assess, in part, because the hiring process itself is complicated and opaque. In real faculty searches, applicants will vary along dimensions of gender, productivity, subfield, doctoral prestige, postdoctoral experience, references, and more; applicants apply to many, but not all searches; and both applicants and institutions have internal, often undeclared preferences. Our aim in this paper is not to model all of these complexities. Instead, we adopt the more narrow goal of estimating the effective role of measurable factors like gender, productivity, and institutional prestige on observed faculty hiring outcomes. We do this by formulating a network model of the yearly matching process of applicants
to faculty openings, which we parameterize to allow us to quantify the impact of different features of faculty applicants. This approach allows us to investigate gender balance in the hiring histories of individual institutions and in individual faculty placement.

We begin by describing the faculty hiring and scholarly productivity data sets and the statistical features we derive from them. We then formulate a network model for faculty placement, check its accuracy in reproducing patterns found in the real hiring network, and use it to test a variety of hypotheses about the model features. Finally, we discuss our results in the context of other findings on gender inequality and highlight strengths and weaknesses of our analysis, before concluding.

Figure 2.1: For the 2659 computer science faculty in our sample (collected in 2011), the distribution of years in which they were first hired as an assistant professor.

2.1 Data and Features

The primary data set that we used is a comprehensive, hand-curated list of the education and academic appointment histories of tenure-track or tenured computer science faculty [19]. This data set covers the 205 departmental or school-level academic units on the Computer Research Association’s authoritative Forsythe List of Ph.D.-granting departments in computing-related disciplines.
in the United States and Canada. For each of these units, the data set provides a complete list of regular faculty from the 2011–2012 academic year, and for each of the 5032 faculty listed, it provides partial or complete information on their education and academic appointments, obtained from public online sources, mainly resumés and homepages.

Within this group, we selected the 2659 faculty who both received their Ph.D. from and held their first assistant professorship at one of these institutions, and for whom the year of that hire is known and occurred in 1970–2011. Figure 2.1 shows the distribution of these hire dates. The first requirement ensured that we modeled the relatively closed North American faculty market; roughly 87% of computing faculty received their Ph.D. from one of the Forsythe institutions, and past analysis has shown that Canada and the United States are not distinct job markets in computer science [19]. A number of faculty were removed in this step because the location of their first assistant professorship was not known; these were mainly senior faculty. The second requirement allowed us to extract a yearly time series of applicants and openings, and thus use a more realistic model of faculty hiring over time. Of the included faculty, women made up 16.1%, which was not significantly different from the fraction in the discarded set \( p = 0.92, \chi^2 \), and the changes in institutional rank (see next subsection) were not significantly different between men and women in the discarded set \( p = 0.325, \text{Mann-Whitney} \). Thus, our inclusion criteria are unlikely to bias our subsequent results.

We modeled the hiring process using a parametric model of edge formation in the faculty hiring network, in which the probability that a particular applicant is matched to a particular job opening depends on features of both applicant and opening. These features were (i) an applicant’s gender, (ii) the prestige of the hiring institution, (iii) an applicant’s scholarly productivity, (iv) an applicant’s postdoctoral training, (v) the prestige difference between doctoral and hiring institution, and (vi) whether those institutions are in the same or different geographic regions. For each, we describe the way the feature was constructed and provide some simple statistics describing their relationship to gender.

\(^3\)http://archive.cra.org/reports/forsythe.html
Institutional prestige. From the education and appointment data, we constructed a faculty hiring network, a directed multigraph where each node is an institution and each Ph.D. graduate from an institution $u$ who began as an assistant professor at $v$ is represented by a single directed edge $(u, v)$. Each node in this network is annotated with its institution’s prestige rank [19], which is also given in the primary data set.

The prestige rank of an institution quantifies its ability to place its graduates as faculty at other prestigious institutions. Formally, $\text{rank}(u)$ is the mean rank of $u$ across all orderings that have the minimum number of “violating” arcs, i.e., an upward-pointing arc $(u, v)$, where $\text{rank}(v)$ is better than $\text{rank}(u)$. Such a ranking is called a minimum violation ranking (MVR) and is a common way to measure prestige in social systems [30, 44]. The prestige ranking we used was obtained by sampling the MVRs for the full faculty hiring network, and it represents a hierarchy on the institutions in which only 12% of edges violate the ranking, i.e., only 12% of individuals were hired at an institution more prestigious than their doctorate institution. This ranking correlates with the popular but widely criticized [6] computer science ranking by U.S. News & World Reports ($r^2 = 0.80$), but it has the advantages of covering the complete Forsythe list and being based on the collective hiring decisions of the departments themselves.

We constructed two features using these ranks: the rank difference $\Delta \text{rank}(u, v)$ between the applicant’s doctoral institution $u$ and the hiring institution $v$, and the $\text{rank}(v)$ of the hiring institution alone.

Comparing female and male faculty in our sample, we found no significant difference in the ranks of the doctoral institutions ($p = 0.41$, Mann–Whitney) or the hiring institutions ($p = 0.12$, Mann–Whitney). The distribution of the rank differences quantifies the degree to which applicants tend to move up or down the ranking when they take a faculty position (see Table 2.1). We found no significant difference in the rank differences between men and women, both including ($p = 0.33$, Mann–Whitney) and excluding “self-hires” ($p = 0.11$, Mann–Whitney), i.e., cases in which a university hires one of its own graduates. We did find a significant difference in the rates of self-hires, with 9.4% of women being self-hired compared to 6.1% of men ($p=0.02$, $\chi^2$). Altogether,
men and women are trained and hired at similar rates across prestige rankings.

<table>
<thead>
<tr>
<th></th>
<th>down</th>
<th>up</th>
</tr>
</thead>
<tbody>
<tr>
<td>men</td>
<td>1877 (79.3%)</td>
<td>491 (20.7%)</td>
</tr>
<tr>
<td>women</td>
<td>357 (81.0%)</td>
<td>84 (19.0%)</td>
</tr>
</tbody>
</table>

Table 2.1: Women and men move up in the prestige rankings at similar rates (excluding self-hires.)

**Scholarly productivity.** Publication records are an important factor in the evaluation of faculty candidates. For each applicant we assigned a feature that captures their scholarly productivity, controlling for subfield variations, prior to being hired into their first assistant professorship.

To construct this feature we first collected a complete publication profile for each faculty from DBLP, an online bibliographic database\(^4\) that, in late 2015, indexed over 3.1 million publications written by over 1.6 million authors, mainly computer scientists, using manual name disambiguation as necessary. Through this procedure, we obtained publication records, including titles and publication dates, for 2528 (95.1%) faculty in our sample. The few individuals for whom we could not identify a DBLP profile were assumed to have no publications.

Publication records in DBLP include journal articles, conference papers (which, in computer science, are peer reviewed), as well as workshop papers (which often are not). The perceived value of different publication types, particular venues, or position in the author list varies by subfield, and we did not attempt to account for these differences here. Instead, we used the number of publications that each faculty had published by one year after starting their assistant professorship, but normalized to control for publication rate variability across subfields. To construct this normalization, we first aggregated the text contained in all the paper titles of a particular faculty’s DBLP profile, a technique that is common in semantic analysis of short texts\(^{[10]}\). We then applied Latent Dirichlet Allocation\(^{[10]}\) to obtain 10 topics or subfield distributions over words, which together captured the total variation in words across all publication records. As a side effect, we also inferred for each faculty a probability distribution over subfields that characterizes their individual

---

\(^4\) [http://dblp.uni-trier.de/](http://dblp.uni-trier.de/)
publication record. To verify that these distributions were reasonable, we manually inspected the most common words in each topic and found good agreement with classic subfields in computer science. Similarly, we verified that the inferred topic distributions for a set of well-known computer scientists aligned with their known specialities.

For each subfield, we computed a distribution over paper counts, weighted by each faculty’s inferred emphasis on that subfield. For each faculty, we computed a single composite $z$-score for their overall productivity by taking a weighted average of $z$-scores over subfield distributions, with weights given by the faculty’s subfield probability distribution. The result is a feature that represents each person’s relative productivity, controlled for their own distribution of work across subfields and the norms within those subfields.

Productivity scores do not differ between men and women. This is true even when we consider only men and women who moved up the ranks and, separately, men and women who moved down ($p > 0.05$, Mann–Whitney). Median productivity scores for men and women in each of these categories are reported in Table 2.2. We did find that individuals with postdoctoral experience have significantly higher productivity scores than individuals without postdoctoral experience ($p < 0.01$, Mann–Whitney). This was true for men and women, separately and together. This is not surprising, as postdoctoral training allows more time to write papers prior to going on the faculty job market. As we note below, separate treatment of productivity and postdoctoral training allowed us to assess whether or not there is intrinsic value in postdoc experience beyond providing additional time to publish papers.

We note that the productivity scores of men and women do differ when we restrict our analysis to include men and women hired after 2002 (the median start year for women). Among these individuals, men are significantly more productive than women ($p = 0.03$, Mann–Whitney). This finding supports the existence of a productivity gap in recent years, despite the previously mentioned studies, which suggest that such gaps have narrowed or closed over time in other disciplines [107, 117].
Table 2.2: Median z-scores by gender and by whether a faculty moved up or down the ranking for their faculty position. We find no significant differences comparing men and women’s productivity scores in each of these categories. Median values are negative indicating that productivity scores are right-skewed due to prolific faculty.

<table>
<thead>
<tr>
<th></th>
<th>down</th>
<th>up</th>
<th>all</th>
</tr>
</thead>
<tbody>
<tr>
<td>men</td>
<td>-0.322</td>
<td>-0.207</td>
<td>-0.327</td>
</tr>
<tr>
<td>women</td>
<td>-0.331</td>
<td>-0.215</td>
<td>-0.329</td>
</tr>
</tbody>
</table>

Geography and postdoctoral training. Geography and postdoctoral training were captured in two binary features. For the former, we assigned a value of 1 if the pair \((u, v)\) spanned two institutions in the same geographic region (U.S. Census regions plus Canada), and a 0 otherwise. For the latter, we assigned a value of 1 if a person had any postdoctoral experience recorded in our primary data set, and a 0 otherwise.

We found no difference in the percentages of men and women graduating and being hired in the same geographic region \((p = 0.12, \chi^2)\). Of the people falling into this category, we next asked whether movement up or down in the ranks was linked to gender, and we found no evidence to suggest that these variables were related \((p=0.72, \chi^2)\). We did find, however, that for individuals who changed geographic regions, men were significantly more likely than women to have moved up in rank \((p = 0.01, \chi^2)\). These results are presented in Table 2.3. Additionally, conditioned on moving up the ranks, men changed geographic regions significantly more than women \((p = 0.03, \chi^2)\), with 67.8% of men changing regions compared to only 48.7% of women.

Table 2.3: For individuals graduating and being hired in separate geographic regions, men are significantly more likely to be moving up the ranks \((p = 0.01, \chi^2)\).

<table>
<thead>
<tr>
<th></th>
<th>down</th>
<th>up</th>
</tr>
</thead>
<tbody>
<tr>
<td>men</td>
<td>1150 (85.7%)</td>
<td>192 (14.3%)</td>
</tr>
<tr>
<td>women</td>
<td>220 (92.1%)</td>
<td>19 (7.9%)</td>
</tr>
</tbody>
</table>

We found that, in general, women were significantly more likely than men to have postdoctoral experience. 24.1% of women in the dataset completed at least one postdoc compared to only 19.3%
of the men \((p=0.03, \chi^2)\). Having postdoctoral experience, though, did not make women any more
or less likely to move up the ranks than men \((p=0.92, \chi^2)\), as displayed in Table 2.4.

<table>
<thead>
<tr>
<th></th>
<th>down</th>
<th>up</th>
</tr>
</thead>
<tbody>
<tr>
<td>men</td>
<td>347 (86.3%)</td>
<td>55 (13.7%)</td>
</tr>
<tr>
<td>women</td>
<td>80 (86.0%)</td>
<td>13 (14.0%)</td>
</tr>
</tbody>
</table>

Table 2.4: For individuals with postdoctoral experience, men and women move up the ranks at
similar rates \((p = 0.92, \chi^2)\).

Finally, we note that the role of postdoctoral experience appears to have changed in recent
years. Comparing individuals whose first assistant professorship began either before or after 2002,
postdoctoral training rates were significantly higher following 2002, 28.1% compared to only 15.5%
before 2002 \((p<0.01, \chi^2)\). Men and women received postdoctoral training at similar rates post-2002,
29.5% for women and 27.7% for men \((p = 0.68, \chi^2)\), but the men who did were significantly more
productive than the women \((p<0.01, \text{Mann–Whitney})\). We also note that after 2002 women with
postdoctoral training were not significantly more or less productive than men without postdoctoral
training \((p=0.44, \text{Mann–Whitney})\), suggesting that women faced additional obstacles which limited
their productivity.

<table>
<thead>
<tr>
<th></th>
<th>observed</th>
<th>uniform</th>
<th>step</th>
<th>logistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean geodesic path length</td>
<td>2.23</td>
<td>2.05 ± 0.01</td>
<td>2.07 ± 0.01</td>
<td>2.16 ± 0.01</td>
</tr>
<tr>
<td>mean local clustering coefficient</td>
<td>0.25</td>
<td>0.34 ± 0.01</td>
<td>0.38 ± 0.01</td>
<td>0.22 ± 0.01</td>
</tr>
<tr>
<td>% reciprocated hires</td>
<td>18.95</td>
<td>14.52 ± 0.81</td>
<td>4.17 ± 0.34</td>
<td>13.93 ± 0.69</td>
</tr>
<tr>
<td>% reciprocating institutions</td>
<td>14.25</td>
<td>13.23 ± 0.77</td>
<td>1.72 ± 0.21</td>
<td>9.86 ± 0.61</td>
</tr>
<tr>
<td>% self-hires</td>
<td>6.62</td>
<td>0.93 ± 0.18</td>
<td>3.74 ± 0.27</td>
<td>1.95 ± 0.25</td>
</tr>
<tr>
<td>% placements within same region</td>
<td>40.54</td>
<td>21.27 ± 0.77</td>
<td>24.48 ± 0.76</td>
<td>29.15 ± 0.75</td>
</tr>
</tbody>
</table>

Table 2.5: Network summary statistics used in model checking of uniform, step, and logistic choices
of \(f\). In each row, boldface indicates the model that best reproduces that characteristic of the
observed network.
2.2 A model of the faculty market

Faculty hiring is a complicated process, and the particular outcome of a faculty search can depend on a surprising variety of factors. Here, we aim to pare down this complexity to formulate a reasonably simple but still useful model of the faculty market as a whole in order to estimate the influence of different features on hiring outcomes in computer science. Our approach uses a data-driven statistical model of the observed outcomes and their features, which is distinct from models of strategic interactions among departments [52].

We note two key properties of the faculty market: (i) assistant professor hires are made in rounds, generally once per year, and (ii) these hires are not independent of each other. This second property comes from the fact that two institutions cannot hire the same applicant. A faculty hiring network (where each directed edge (u, v) represents the hiring a graduate of node u as an assistant professor at node v) is thus the accumulation of yearly sets of such non-independent hiring edges.

We model this network assembly process by modeling the annual matching of candidates to openings in each year of the data. Systematic information on unsuccessful applicants and unfilled openings is not generally available for any year, and for this reason we make the simplifying assumption that matchings are made among the observed candidates and openings (the positions that were filled) in each year. This is not an unreasonable assumption: in practice, only a small fraction of faculty openings go unfilled each year, meaning that the set of successful applicants is a reasonable approximation of the top candidates across all searches. Thus, for each year t, we first break the observed hiring edges \( \{(u_i, v)\}_t \), where i indexes across all candidates, into two “stub” sets, one for the candidates \( \{u_i\}_t \) and one for the openings \( \{v\}_t \). We then generate a matching \( M_t \) on these stubs using a probabilistic model \( f \) that is parameterized by the pair-level features described in the previous section.

Regardless of the reasons why, in practice, hiring committees prefer applicants trained at more prestigious departments about 80% of the time [19]. We model this and other preferences of
a typical hiring committee via a logistic function for the pairwise probabilistic model:

\[
f(\vec{x}[u_i, v], \vec{w}) \propto \left(1 + e^{-\vec{x}[u_i, v] \cdot \vec{w}}\right)^{-1},
\]

where \(\vec{x}[u_i, v]\) is a vector of features of the candidate-opening pair \(u_i, v\), and \(\vec{w}\) is the global set of weights on those features that we learn from the data.

This choice of \(f\) allows us to automatically capture two important special cases: if \(f\) is independent of \(\vec{x}\), then rank and other features play no role and the matching is equivalent to the popular configuration random graph model \([76]\); when \(f\) is a step function on rank, and independent of other features, then hires are chosen uniformly at random from those trained at more prestigious departments, which is equivalent to the MVR ranking method used in \([19]\). The step function is the simplest \(f\) that depends on some of our features, and we use it as a baseline model later in order quantify the improvement from incorporating additional model features.

Applicants may also prefer openings at highly ranked departments, desiring the prestige and resources associated with these institutions. We model this preference by filling the openings \(\{v\}_t\) sequentially, choosing an unfilled opening to fill with probability proportional to \(1/\text{rank}(v)\) (where more highly ranked departments have smaller rank scores). Through this sequential matching process, our model fills each opening in a given year \(t\) from the available candidates in that year. Applying this process for each year \(t\) from 1970 to 2011, the model assembles a full faculty hiring network. It is worth noting that this model is loosely similar to the popular exponential random graph model \([92]\); however, in our formulation, edge formation is ordered and not independent, which requires a slightly different treatment.

We score the quality of our model by measuring its total error with respect to the observed placements, where total error is defined as the mean squared error (MSE) in the placements plus an L1 regularization term to prevent the model from overfitting. Mathematically,

\[
\text{err} = \frac{1}{m} \sum_{i=1}^{m} \left[\text{observed}(u_i) - \text{model}(u_i)\right]^2 + \lambda \sum_k |\vec{w}_k|,
\]

where \(\text{observed}(u_i)\) is the observed placement rank of candidate \(i\) and \(\text{model}(u_i)\) is the simulated placement rank. Using the MSE allows the model to receive partial credit for matching an applicant
to an opening with rank similar to the observed rank, rather than, for example, receiving credit only if the applicant matches to the observed opening (which simply counts the number of correct placements). To estimate the model’s parameters $\vec{w}$, we use a standard implementation of a direct search optimization algorithm (Nelder-Mead).

2.2.1 Model checking

As a first step, we check that synthetic faculty hiring networks produced by our model have similar structural patterns to the observed network. We do this for each of three choices of $f$, the logistic function of Eq. (2.1) using all six features, as well as its two special cases, a uniform function and a step function. Using standard network summary statistics [78], such as the mean geodesic path length and the mean local clustering coefficient, as well as hiring-specific statistics on reciprocal hiring, self-hiring, and within-region placement, we compare the observed and simulated networks. Table 2.5 summarizes the results of this exercise.

In general, we find very good agreement between the statistical properties of the real network and those generated by each of our models, with the logistic model performing best overall. Each of our models underestimates the rates of reciprocal hiring and self-hiring. This suggests that additional factors not present in our model likely influence these types of hires, perhaps related to the pre-existing social and professional connections associated with such hires.

Finally, we verify that the feature weights learned by our model are consistent under cross-validation in which sets of five randomly selected years of data are set aside for testing. Feature weights are largely stable across runs with only minor fluctuations that do not have a significant impact on modeling error.

2.3 Results

In the following sections, we examine gender’s role in university faculty hiring at three levels by investigating (i) system-wide effects, (ii) hiring results for individual institutions, and (iii) hiring results for individual candidates. We conclude by forecasting when computer science will reach
gender parity, should women’s presence in the field continue to grow at the current rate.

2.3.1 Market-level analysis

We trained a series of placement models by incorporating, one at a time, the attributes described in the previous section. The order in which attributes were added to the model was determined greedily: each remaining attribute was added separately to the previous model, and the attribute producing the greatest reduction of error was built into the subsequent model. Gender was incorporated last in order to determine if it significantly improved modeling results beyond the effects of all other variables. Figure 2.2 shows the extent to which modeling error decreased as attributes were incrementally incorporated.

![Figure 2.2: Reduction of modeling error as features are added to the model. Percent reductions are computed relative to the step function model as a baseline. Median percent reductions are reported for each model, and attributes producing a significant reduction in error ($p<0.05$, Mann–Whitney) are marked with braces and asterisks.](image_url)

The list of attributes added to the model, in decreasing order of error reduction, was (i) rank difference between doctoral and hiring institutions, (ii) scholarly productivity, (iii) rank of hiring
institution, (iv) postdoctoral training, and (v) whether doctoral and hiring institutions were in the same geographic region. It is perhaps unsurprising that rank difference and productivity yield the largest improvements in modeling results as these attributes are known to play key roles in faculty hiring. Incorporating the rank of the hiring institution also significantly improves modeling results ($p < 0.01$, Mann–Whitney). Based on the sign of the inferred coefficient, this suggests that the most prestigious universities are more selective in their hires and potentially value prestige more than lower-ranked universities.

Neither postdoctoral experience nor geographic information alone produced a significant change in modeling error. Together, however, these features accounted for a small but significant improvement. Because the productivity score had already been greedily added to the model prior to postdoctoral training, this result implies that postdoctoral training, in general, is only nominally useful beyond the extent to which it offers a trainee additional time to publish more papers and to thereby increase his or her productivity score. Geographic information, similarly, has little effect on modeling error. On its own, this finding suggests that issues of mobility do not strongly and systematically affect the placement of all faculty. We noted in Sec. 2, however, that men who moved up in the ranks are more likely than women who moved up to have changed geographic regions. Together, these findings suggest that mobility may play a small but real role in placement differences for some groups of men and women.

Finally, the addition of gender into the placement model did not significantly improve modeling results. We found this to be true both when we computed placement error for all faculty, and for women, separately. That the incorporation of gender does not significantly improve global error suggests that gender in and of itself does not systematically affect all hires beyond potential indirect effects encoded in other features, such as productivity. This finding echoes historical work [22], which suggests that gender discrimination within science is not evenly distributed and warns that ignoring this non-uniformity risks promoting inequality.

That being said the weight assigned to gender was nevertheless non-zero, indicating that a subtle difference does exist. To convert this difference into more tangible terms, we calculated the
number of additional papers a female candidate would need to publish in order to achieve the same job placement as an otherwise equivalent male candidate. Across subfields, on average, women must publish approximately one additional paper—a roughly 10% increase in productivity—in order to compete on even footing with men.

### 2.3.2 Institution-level analysis

For faculty hiring to be free of uniform and systematic gender bias does not suggest that inequality cannot exist at the level of individual institutions. In this section, we explore this possibility directly by comparing the observed hiring at each institution with the distribution of outcomes drawn from our generative model of faculty placement. Using all features listed in previous sections, we simulated 1000 complete hiring histories, requiring as before that universities compete for candidates during each year of the process. For each simulation, we tracked the number of male and female hires by year and by institution, resulting in an evaluation of the gender balance of each department, taking into account the number of women on the job market when the department was hiring and the likelihood that those candidates would have been hired by the institution. The result is a set of institution-specific assessments that accommodate the non-independence of hires while controlling for placement likelihoods of candidates.

Figure 2.3: Three examples of model-based sampling of university-specific female hire distributions. Each green trajectory denotes the cumulative number of hires for a single simulation of the placement model at the indicated university. Running many simulations creates the distribution over final counts, shown on the right. The actual trajectory of hires made by the institution (within the data set) and the resulting final count are highlighted in black. UC Berkeley, Princeton, and Brigham Young represent examples of expected, female-skewed, and male-skewed hiring, as indicated by the location of the actual value within each sampled distribution.
In comparing each institution’s actual number of female hires to the expected number under simulation, we find that most institutions perform very closely to their expected values. There are, however, institutions that exceed or fall short of the model’s expectations. Figure 2.3 highlights universities in each of these three categories.

By comparing the results of many institutions, we asked whether female hiring patterns change as a function of rank. Figure 2.4 illustrates the difference between actual and expected counts of women at the top 50 universities, sorted by rank. We note that top-ranked institutions (ranks 1–10) tend to hire more women than expected, while slightly lower-ranked institutions (ranks 11–20) typically hire fewer. This pattern may suggest that efforts made by top institutions to rectify instances of gender imbalance in their own departments could come at the expense of impeding similar efforts by lower-ranked institutions.

Figure 2.4: Comparison of actual and expected female hiring over the top 50 institutions. Dots represent actual values minus expected values calculated from distributions samples as in Fig. 2.3. The shaded region denotes the 25th-75th percentiles, based on modeling outcomes. Six particular universities are annotated. Top 10 schools hire slightly above expectations while ranks 11–20 hire below expectations. This suggests that the efforts by the highly-ranked schools to rectify any gender imbalance may have impeded the efforts of lower-ranked schools hoping to do the same.
2.3.3 Candidate-level analysis

Having analyzed faculty hiring at the system level and at the level of individual institutions in previous sections, we now investigate the placement of individual faculty. The complete simulations of the faculty market used in the institution-level analyses were re-analyzed for each individual faculty. Specifically, for each individual, we compiled a list of simulated placements and their frequencies, constituting a distribution of plausible outcomes for that person. By comparing the ranks of the institutions in an individual’s list of plausible outcomes to that of their hiring institution, we obtained a distribution representing the amounts by which each person has over- or under-performed relative to their simulated outcomes. We separated these individuals by gender, and found that men and women meet or exceed model expectations at similar rates, though women are more likely to exceed expectations ($p < 0.01$, Mann–Whitney). For under-performing individuals, however, men tend to fall short of their expectations by significantly larger amounts ($p<0.01$, Mann–Whitney).

![Figure 2.5: Mean placement error by year. Placement error is computed as the difference between the rank of the institution where the person was hired and the rank of the institution where they placed under simulation. Higher variance in female placement error is within fluctuations expected due to lower female representation in the data set. Adjusted for yearly representation in the data, error is neither systematically increasing nor decreasing in time.](image-url)
We also find that individuals with postdoctoral training are more likely to outperform model expectations than those without this experience \( (p < 0.01, \chi^2) \). This result is true for men and women, both separately and together, although women tend to exceed their expectations by larger amounts \( (p < 0.01, \text{Mann–Whitney}) \). This implies that in the past, postdoctoral experience may have provided a strategic advantage to women looking to move up the ranks of the prestige rankings. With more men receiving postdoctoral training in recent years, however, it appears that what was once a competitive strategy may now be the norm.

Grouping individuals together by hiring year, we investigated how placement error is distributed over time. This allows us to assess the degree to which faculty hiring appears to have changed over the timeframe spanned by the dataset. Like the previous analysis, this is equivalent to looking at the average amount by which men and women over- or under-perform, collectively, in each hiring year. For instance, a pattern of women tending to under-perform early in the time period, and to over-perform later in the time period would be consistent with improved conditions for female faculty today. Instead, we see noisy, but relatively flat functions for the placement errors for both women and men (Fig. 2.5), with the difference in fluctuations by gender attributable to the difference in sample size. This pattern indicates that model errors in either direction are equally likely for men and for women, and for both recent hires and hires from several decades ago.

2.3.4 Long-term forecast for gender parity

Over the four decades spanned by our data, the proportions of received doctoral degrees and assistant professor positions held in computer science by women have both steadily increased, from around 5% to roughly 20% (Fig. 2.6). However, the share of new faculty positions held by women is on average about 1% lower than the share of doctorates, which reflects the well-documented leakiness of the academic training pipeline [45]. While not a large number in magnitude, a 1% gap is a substantial proportional difference (about 7–20%) given that the gender ratio is so heavily skewed toward men.

Nevertheless, the long-term trend in computer science is toward gender parity. To estimate
when women and men will hold equal shares of new faculty positions, we fitted a simple linear model to the historical trend and extrapolated it into the future (Fig. 2.7). Under this model, the share of positions held by women increases by 0.43% per year on average, meaning that it will take roughly 60 years from 2012 to reach parity at the assistant professor level, with a 95% confidence interval of 30–100 years. Full gender parity across all levels of faculty should then occur 30–40 years later, when the first gender-parity cohort of assistant professors begins to retire.

2.4 Discussion

Here, we used a unique data set on the hiring of assistant professors in computer science from 1970–2011 to measure the importance of six features of candidates on observed hiring outcomes. Among these, doctoral prestige and scholarly productivity play an outsized role, while gender alone does not appear to be a significant factor in the typical hiring decision. At face value, these findings are consistent with a system that is not overtly biased by a candidate’s gender.

However, we also found evidence of (i) unexpectedly gender imbalanced hiring patterns at individual institutions, (ii) significant differences between genders in rates and the effects of publish-
Figure 2.7: Gender ratio of assistant professors in computer science, by gender, and a projection for when gender parity will be reached. If the historical trend continues unaltered, gender parity will occur in approximately 2075. Shaded regions represent extrapolated 95% confidence intervals from an ordinary least squares regression.

...ing and postdoctoral training, (iii) differences between men and women who move up the prestige ranking, and (iv) evidence of that higher ranked institutions’ success at hiring female faculty may be limiting similar efforts at marginally less highly ranked institutions. The apparent conflict between these two sets of findings about the same faculty market shows that the role of gender in faculty hiring is subtle and generally not well characterized by simple statistics or broad generalizations. Overall, our results suggest that the actual faculty hiring market in computer science is neither extremely dire for women [77] nor extremely favorable [113].

Under our model, the inclusion of candidate gender did not significantly improve its ability to correctly place faculty overall. There are at least three plausible interpretations of this behavior. First, gender could be an irrelevant feature in faculty hiring. This interpretation is implausible because we also found that gender correlates with postdoctoral training, productivity, and geographic mobility, especially in the past 10 years. Second, the effect of gender may not be included realistically in the model. Evidently, a uniform penalty or advantage based on gender does not help reduce placement error rates, and so the gender feature received a weight near zero. Or third,
the primary effects of gender on placement are already incorporated into the model through other features that correlate with gender.

This latter interpretation is particularly plausible. For assistant professors who started since 2002, productivity scores correlate with gender, with men being on average more productive than women with the same amount of training ($p < 0.01$, Mann-Whitney). Moreover, the productivity of women with postdoctoral training is not significantly different from men without it ($p = 0.44$, Mann-Whitney), and under our model, women need to be about 10% more productive, on average, in order to place at equal rates as men. That is, productivity already encodes gender-based differences, making a separate gender variable in the model redundant. The origin of this productivity gap seems unlikely to be related to inherent differences in talent or effort, and may instead be related to differential access to resources and mentoring [16], greater rates of hostile work environments or sexual harassment [18], differences in self-perceptions [17], or other gender-correlated factors. Additional research is needed to investigate these possibilities.

Our findings that support the existence of a gender-driven productivity gap in recent years are at odds with several studies indicating that such gaps have narrowed over time or perhaps closed altogether in other disciplines [107, 117]. These studies, however, examine the total number of publications and citations accumulated over one’s entire career whereas we focus on an individual’s publication record up until one year after being hired. Differences in productivity at this stage have been noted previously [64] and are most relevant to our study of faculty hiring, as these differences likely influence hiring as well as tenure decisions and thus the individuals observed in our dataset. Indeed, we find that women are overrepresented on the low end of our productivity measure and publish fewer papers per year on average for the first several years of employment. A better understanding of the causes behind this lag in productivity would inform faculty evaluation procedures and tenure policies, potentially improving retention of women at this career stage.

The productivity gap also suggests that postdoctoral training has been one way for women to compete on an equal basis with men in the faculty market. For faculty who started prior to 2002, the rate of postdoctoral training was indeed higher among women than men, which may reflect a
compensatory adaptation to a biased system [116]. Since 2002, however, these rates have equalized, meaning that in a typical faculty search today, men are likely to appear more productive, on average, than women. Institutional self-hiring, i.e., becoming faculty at one’s doctoral institution, may reflect a separate kind of compensatory adaptation. Across 40 years, women have been hired by their doctoral institutions at a greater rate than men, and this difference has grown significantly since 2002. Determining the extent to which these patterns reflect strategic responses to a changing market would shed new light on the underlying market structure.

The long-term trend in the gender ratio in computer science faculty hiring is toward parity. The pace, however, is glacial, and we estimate that it will take roughly 60 years to reach. There are two main reasons to want to accelerate this trend: (i) social justice and the provision of equal opportunities [34, 29], and (ii) increased scientific innovation, creativity, and productivity [5, 83, 14, 51]. Achieving parity sooner, however, is likely to require novel and concerted efforts, as the faculty gender ratio correlates strongly with the doctoral gender ratio (Fig. 2.6), suggesting that relatively little has changed, fundamentally, over the past 40 years.

For an individual computer science department aiming to improve its faculty gender balance, the non-independence of hires poses a thorny problem. We observe a rank-dependent pattern indicating that more highly ranked departments tend to have better than expected rates of female faculty hiring and retention (Fig. 2.4), potentially at the expense of those departments ranked just below, e.g., ranks 1–10 vs. 11–19, and ranks 20–25 vs. 26–40. Even if all departments wished to hire more female faculty, the more highly ranked institutions will tend to have a competitive advantage in attracting any candidates. Thus, if many departments are competing to hire a small number of female candidates, the lower-ranked departments will tend to lose out. Broadening the pool of female candidates is one solution to this problem, which a recent experimental study showed has a direct improvement on the gender ratio among faculty hires [101].

Because the hiring network data set is a snapshot of regular faculty in the United States and Canada in the 2011–2012 academic year, it necessarily omits any information about faculty who left or retired from computer science prior to 2012, who were hired since 2012, or who were
hired at the associate or full professor level during our study period, e.g., faculty who spent time in industry or who did their assistant professorship outside of computer science or outside the U.S. and Canada. As a result, hiring and retention are confounded in our analysis, and the current gender imbalance at some departments may be smaller than what we estimate. Were information on these missing individuals to become available, our model could be used to study questions about the leaky pipeline, e.g., do certain institutions or groups of institutions contribute more or less to women leaving the pipeline, or to compare the dynamics of the new-hire market and the senior-hire market. Another limitation of this data set is that it does not include information on other faculty variables, such as their ethnicity, which can be particularly skewed, e.g., with African American faculty \[1\], socio-economic background, or nationality. These represent important directions for future research.

The productivity feature developed here could potentially be improved. For simplicity, we assigned all publications equal weight in our analysis, which favors quantity over quality. A better feature, however, would combine a candidate’s scholarly record with an estimate of its scholarly quality and the author’s level of contribution. However, such an extension would be highly non-trivial, in part because quality is difficult to measure accurately and automatically, across subfields. In fact, reliably assessing publication quality is hard even for humans, particularly when that contribution is interdisciplinary \[55\]. An automated tool for doing so would have value both for the scientometrics and text mining communities as well as hiring committees.

In our model, we used a logistic function to score potential matchings between candidates and hiring institutions. Allowing this function to take a more complex form could improve the model’s accuracy, either through the incorporation of interaction terms or by adopting a richer functional form in place of Eq. \((2.1)\). Additionally, other loss functions might improve modeling results beyond MSE, which assumes uniform prestige for all faculty in a department and that institutions are similar only by ranking. Though we do not explore these possibilities here, such modifications could enrich future analyses in this area and offer a source of flexibility for adapting our modeling framework to suit other applications.
Faculty hiring networks provide a powerful new tool for understanding the dynamics of academic disciplines, and for investigating the role of different factors in shaping academic careers. The computer science hiring network reveals substantial evidence that gender inequality is present, subtle, and non-uniform. For predicting faculty placement, doctoral prestige and relative productivity appear to be the most important variables. However, the correlation between productivity and gender raises the questions of why, how the gap can be closed, and how our assessments can be informed by its underlying causes. Although the details are different, the computing industry has an equally large gender imbalance. Employing a similar approach to industrial hiring networks and productivity may shed new light on its underlying causes and the means to address it.

To conclude, in this chapter, we explored gender’s role in shaping hiring outcomes for tenure-track faculty in computer science, and through this process, we identified a number of interesting and important directions for future research. One such direction aims to achieve a greater understanding of how a researcher’s scholarly productivity develops over time and potentially in response to their academic environment, giving rise to the observed discrepancies in productivity. In the following chapter, we address this question directly by evaluating the merits of a long-held presumption about the career productivity lifecycle or trajectory of scientists. Crucially, this presumption suggests a null model that, if valid, would provide a valuable template for assessing how well an individual researcher’s productivity is advancing as a function of career age. That template could then in turn be used to benchmark individuals’ relative performances (e.g. compared to their departmental peers) and study how productivity predicts placement, retention, and promotion decisions in computer science.
Chapter 3

The misleading narrative of the canonical faculty productivity trajectory

Portions of this chapter are adapted from:

S. F. Way, A. C. Morgan, A. Clauset, and D. B. Larremore. The misleading narrative of the canonical faculty productivity trajectory, accepted to the International Conference on Web and Social Media. A journal version is currently under review.

This work was presented in June 2016 at the International Conference on Computational Social Science in Evanston, Illinois, and in May 2017 at the International Conference on Web and Social Media in Montréal, Québec.

Scholarly publications serve as the primary mode of communication through which scientific knowledge is developed, discussed, and disseminated. The amount that an individual researcher contributes to this dialogue—their scholarly productivity—thus serves as an important measure of the rate at which they contribute units of knowledge to the field, and this measure is known to influence the placement of graduates into faculty jobs [111], the likelihood of being granted tenure [26, 69], and the ability to secure funding for future research [103].

The trajectory of productivity over the course of a researcher’s lifetime has been studied for at least 60 years, with the common observation being that a researcher’s productivity rises rapidly to a peak and then slowly declines [57, 31, 33, 47, 7], which has inspired the construction of mechanistic models with a similar profile [7, 99, 33, 59]. These models have included factors like cognitive decline with age, career age, finite supplies of human capital, knowledge advantages
conferred by recent education, as well as skill deficits among the young, among others, and have been supported by the observation that productivity curves are not well described by even fourth-degree polynomial models [7]. Indeed, every study we found to date proposes or confirms a “rise and decline,” “curvilinear,” or “peak and tapering” productivity trajectory, regardless of whether researchers are binned by chronological age [57, 31, 25, 33, 47, 99, 59], career age [7, 99], or (only for young researchers) years since first publication [104]. In fact, this conventional narrative of the life course is not restricted to academia, with similar trajectories observed in criminal behavior and artistic production in 1800s France [86] and even productivity of food acquisition by hunter-gatherers [50].

While these past studies have firmly established that the conventional academic productivity narrative is equally descriptive across fields and time, their analyses are based on averages over hundreds or thousands of individuals [57, 31, 25, 33, 47, 99, 59, 7, 104, 86, 50]. This raises two crucial and previously unanswered questions: is this average trajectory representative of individual faculty, and how much diversity is hidden by a focus on a central tendency over a population? To answer these questions, we combine and study two comprehensive datasets that span forty years of productivity for nearly every tenure-track professor in a North American Ph.D.-granting computer science department. By introducing a simple mathematical description of the shape of a scientist’s productivity over time, we map individuals’ publication histories to a low-dimensional parameter space, revealing enormous diversity in the publication trends of individual faculty and showing that only a minority follow the conventional narrative of productivity. In fact, even among the conventional trajectories, individuals exhibit large fluctuations in their productivity around the average trend. Together, these results reveal that productivity patterns are both more diverse and less predictable than previously thought, and that population averages provide a dramatically inaccurate picture of intellectual contributions over time.

Moreover, while we show that the distribution of productivity trajectories resists natural categorization, it is nevertheless possible to explore covariates that are associated with different regions of its parameter space. The literature on such associations has avoided detailed trajectories
and instead focused on the complicated relationship between prestige, productivity, and hiring. Past studies have found that researchers trained at prestigious institutions are likely to remain productive [13, 40], regardless of where they place as faculty [28]. Other results link the prestige of the doctorate and the advisor to early-career productivity but not long-term productivity [90], which is at odds with others [65, 18] who found that early-career productivity predicts long-term productivity. Disagreement about hiring exists as well, with multiple studies finding that doctoral prestige, and not productivity drives the initial placement of faculty [41, 123, 63] while recent work based on comprehensive data in multiple fields suggests that prestige alone is insufficient to fully explain faculty placement [19, 111]. This, too, is complicated by hypotheses of mutual causality, where departments both select for and facilitate high productivity [2]. Unfortunately, while such studies shed light on a complicated system, they tend to restrict their analyses to unusual scientists, such as Nobel laureates or faculty at elite departments, rather than typical researchers. In contrast, the data analyzed here are comprehensive, covering faculty across the prestige hierarchy, which enables us to move beyond total productivity to study publication trajectories in light of prestige, hiring, and past productivity alike.

This study exploits and combines two large datasets related to faculty productivity. The first is a comprehensive, hand-curated collection of education and academic appointment histories for tenure-track and tenured computer science faculty [19]. This dataset spans all 205 departmental or school-level academic units on the Computing Research Association’s Forsythe List of Ph.D.-granting departments in computing-related disciplines in the United States and Canada [1]. For each department, the dataset provides a complete list of regular faculty for the 2011–2012 academic year, and for each of the 5032 faculty in this collection, it provides partial or complete information on their education and academic appointments, obtained from public online sources, mainly résumés and homepages. Of these, we selected the 2583 faculty who both received their Ph.D. from and held their first assistant professorship at one of these institutions, and for whom the year of that hire is known and occurred in 1970–2011. The first requirement ensured that we modeled the relatively

---

1 See http://archive.cra.org/reports/forsythe.html
closed North American faculty market; roughly 87% of computing faculty received their Ph.D. from one of the Forsythe institutions, and past analysis has shown that Canada and the United States are not distinct job markets in computer science [19]. A number of faculty were removed in this step because the location of their first assistant professorship was not known; these were mainly senior faculty.

The second dataset, constructed around the first, is a complete publication history as listed on DBLP, an online database that provides open bibliographic information for most journals and conference proceedings relevant to computing research, using manual name disambiguation as necessary. For each paper in a faculty’s publication history, we recorded the paper’s title, author list (preserving author order), publication type (journal or conference paper), and year of publication. By following this procedure, we collected data for 200,476 publications which covered 2453 (95.0%) faculty in our sample. Of those, we manually collected records of all peer-reviewed conference and journal publication histories from the publicly available curricula vitae (CVs) of 109 faculty, a randomly selected 10% of the 1091 faculty with career lengths between 10 and 25 years, providing a benchmark dataset to evaluate the accuracy of DBLP data (see Supplemental Text A.1).

Our combined dataset consists of the career trajectories of these 2453 tenure-track faculty as of 2011–2012: each professor’s publicly-accessible metadata, their time-stamped Ph.D. and employment history, and the annotated time series of their publications. Briefly, we note that this dataset does not include information on faculty who have retired or left academia prior to 2011. Implications of these data limitations for the conclusions that can be drawn from our analyses are explored in the Discussion section.

3.1 Results

3.1.1 General Trends in Productivity

Two broad trends characterize scholarly productivity in academic computer science. First, publication rates have been increasing over the past 45 years, and second, higher publication rates

\[\text{See http://dblp.uni-trier.de}\]
are correlated with higher prestige. These two observations are intertwined and underpin a number of subsequent analyses, so we explore them briefly in more depth.

Past studies have found that researchers at more prestigious institutions tend to be more productive [28, 123, 84, 63, 90, 2]. Our data corroborate this finding, but we also find that the typical productivity advantage associated with greater prestige holds regardless of whether an institution is public or private, for both early-career publications (first ten years; Fig. 3.1) and lifetime publications (Supplemental Fig. A.4). Regressing the median number of time-adjusted publications (see below) among faculty in a department against departmental prestige indicates that the relationships between prestige and productivity are statistically indistinguishable for public and private institutions, with expected increases in the first decade of a career of roughly 2.7 publications for every 10-rank increase in prestige. In fact, when comparing public and private institutions, neither the prestige-productivity slope nor productivity overall is significantly different
(p = 0.150, 0.148, respectively, two-tailed t-test), contradicting the conventional wisdom that private universities enjoy a productivity advantage over public ones. The conventional wisdom is likely skewed by a focus on elite departments, as 8 of the top 10 computer science departments are private [19], but in fact, private institutions are distributed evenly across all ranks. Expanding this analysis to include lifetime publications increases the prestige-publication slope to 3.28 publications per 10-rank increase in prestige but does not alter the non-significance of public/private status (p = 0.714, 0.346, two-tailed t-test).

Past studies have also found that publication rates have increased over time [32, 54]. However, prior to investigating whether changes in publication rates apply to computer science, we used the manually collected CV data to probe the extent of DBLP’s coverage. Indeed, the fraction of publications indexed by DBLP is non-uniform over time, increasing linearly from around 55% in the 1980s to over 85% by 2011 ($R^2$ = 0.685 $p < 0.001$, two-tailed t-test; see Supplemental Figure A.1 and Supplemental Text A.2). Because DBLP’s coverage of the published literature varies over time, in the analyses that follow we use data from hand-collected faculty CVs whenever possible, and otherwise apply a statistical correction to DBLP’s data in order to account for its lower coverage.

Knowing already that there are substantial differences in productivity by prestige, we separated universities arbitrarily into high, medium, and low prestige groups, and investigated whether the growth of publication rate varies by prestige. We find that the average number of publications per person produced in each calendar year has been increasing at all three strata of prestige at rates between 0.90 and 1.13 publications per decade, for 45 years (Supplemental Figure A.3). Because we have used data from hand-collected faculty CVs to adjust DBLP-derived paper counts for DBLP’s steadily improving coverage over time, these estimated growth rates represent a real increase in publication rates over this 40 year period. Moreover, the observed steady increase in productivity is not uniform across prestige, and the difference between production growth rates between higher and lower prestige departments have widened slightly but significantly over time ($p < 0.05$, two-tailed t-test). In other words, prestigious and non-prestigious institutions have contributed to the overall growth at different rates. Not only are there small but significant differences in productivity
by prestige (Figure 3.1) but those differences are slowly growing (Supplemental Figure A.3).

In order to investigate the productivity patterns of individual researchers, and test the conventional narrative of rapidly rising productivity followed by a gradual decline, for the remainder of this paper we focus on time series of individual productivity. However, due to both the observed growth in productivity, and the variability in DBLP coverage, it would be misleading to directly compare a 1975 publication with a 2011 publication. Thus, hereafter we use “adjusted” publication counts, which corrects the raw DBLP counts to account for both the changing DBLP coverage and the increasing mean publication rate over time (Supplemental Text A.2). All publication counts are hence reported as 2011-equivalent counts.

3.1.2 Individual Productivity Trajectories

Examining the productivity trajectories of individual researchers, we find that they too exhibit substantial and significant differences in their publication rates. Early studies of scholarly productivity noted profound imbalance in the number of articles published by individual researchers [68, 98]. Cole [21] and Reskin [89] in the 1970s noted that about 50% of all scholarly articles were produced by about 15% of the scientific workforce. Our data reflect similar levels of imbalance, with approximately half of all contributions in the dataset authored by only 20% of all faculty. Stratifying by decade, however, the Gini coefficients for productivity imbalance have been declining, from 0.62 in the 1970s to 0.40 in the 2000s (See Supplemental Figure A.5). This trend persists when researchers are restricted to only the publications within the first five years of their careers.

There are several possible explanations for the trend of decreasing inequality in individual productivity. For instance, the lower end of the productivity distribution could have become relatively more productive over time, perhaps as more institutions shifted focus from teaching to research. Or, it may reflect a strengthening selective filter on highly productive faculty, perhaps as community expectations for continual productivity rose. It may also reflect non-uniform errors in the DBLP data, although the correction for DBLP coverage should account for these (Supplemental Text A.2). We leave the investigation of these possibilities for future work.
Figure 3.2: Average publications follow conventional narrative across prestige. Three average productivity curves are shown, stratified by the prestige rank $\pi$ of researchers’ employing institutions as indicated. Averages over researchers at all levels of institutional prestige follow similar productivity trajectories, in agreement with the conventional narrative, but at differing scales of output.

Instead, we focus on testing the conventional productivity narrative that has been described in various disciplines and at many points in time [57, 81, 25, 33, 47, 99, 59, 7, 114, 86, 50]: productivity climbs to a peak and then gradually declines over the course of the researcher’s career. Across computer science faculty, we find that the average number of publications per year over a faculty career is highly stereotyped (Figure 3.2), with a rising productivity that peaks after around 5 years, declines slowly for another 5 years, and then remains roughly constant for any remaining years. Although departmental prestige correlates with productivity in several ways (Figures 3.1 and A.3), it does not alter this stereotypical pattern, which appears essentially unchanged across departments with different levels of prestige, except for a roughly constant shift up as prestige increases (Figure 3.2).

The suggestion that productivity grows in the early years of a career has intuitive appeal. Professors settle into their research environments, begin training graduate students, and build their cases for promotion and tenure. Similarly, many reasons have been suggested for why productivity might decrease after promotion, including increased service and non-research commitments, declin-
ing cognitive abilities, and increased levels of distraction from outside work due to health issues and childcare obligations [37]. Although an average over faculty appears to confirm the stereotyped trajectory of rapid growth, peak, and slow decline, it does not reveal whether this average is representative of the many individual trajectories it averages over, nor does it show how much diversity there might be around the average, and whether that diversity correlates with other factors of interest.

Figure 3.3: Example trajectory and piecewise model. Dots represent empirical annual publications. Orange line shows best fit of piecewise linear model \([\text{Eq. (3.1)}]\) with slopes \(m_1, m_2\), change point \(t^*\), and intercept \(b\) annotated.

To characterize the productivity pattern within an individual career, we fit a simple stereotypical model of productivity over time to the number of papers published per year,

\[
f(t) = \begin{cases} 
  b + m_1 t & 0 \leq t \leq t^* \\
  b + m_1 t^* + m_2 (t - t^*) & t > t^* 
\end{cases}
\]  

(3.1)
a piecewise linear function in which \(t^*\) is the change point between the two lines, \(m_1\) and \(m_2\) are the rates of change in productivity before and after the change point, respectively, and \(b\) is the initial productivity (Figure 3.3). By fitting these four parameters to each individual’s publication trajectory, we map that trajectory into a low-dimensional description of its overall pattern (fitting done by least squares; see Supplemental Text A.7 for optimal numerical methods and A.3 for
detailed discussion of statistical models).

To address the possibility that a researcher’s best-fit parameters may be sensitive to small changes in the years of their publications, we repeatedly re-fit model parameters to productivity trajectories with small amounts of noise (see Supplemental Text A.4). This revealed that a majority (74.6%) of trajectories are robustly characterized under the model, each consistently falling into the same region of parameter space for over 75% of resampled trajectories. We refer to these trajectories as “stable” in subsequent analyses. Due to the sensitivity of the remaining 25.4% of trajectories’ parameters, all analyses and discussions of model parameters hereafter refer to stable trajectories unless otherwise noted.

Figure 3.4: Distribution of individuals’ productivity trajectory parameters. Diverse trends in individual productivity fall into four quadrants based on their slopes $m_1$ and $m_2$ in the piecewise linear model Eq. (3.1). Subfigures show example publication trajectories to illustrate general characteristics of each quadrant. The shaded triangular region corresponds to the conventional narrative of early increase followed by gradual decline. Sensitivity of each individual’s quadrant to noise in publication years is represented by color and size: orange and gray dots represent trajectories with stable (74.6%) and unstable classifications (25.4%), respectively, and sizes are proportional to the fraction of stable trials; see Supplemental Text A.4. Counts and percentages tabulate the number of stable trajectories falling into each quadrant, relative to the total.
The narrative of “early growth in productivity, followed by a slow decline” implies four conditions on the inferred parameters: while the conditions of growth \((m_1 > 0)\) and decline \((m_2 < 0)\) are straightforward, we interpret “early growth” to mean that inferred peak productivity comes within the first decade after hiring \((t^* \leq 10)\) and “slow” to mean that the slope of decline is smaller in magnitude than the slope of growth \(|m_2| < m_1\). After fitting individual trajectory models to all 2543 faculty in our sample, we find that for faculty who have been employed for 10–25 years \((N = 1091)\), only 32.1% follow the stereotypical trajectory. Even dropping the aforementioned restriction on \(t^*\) increases the fraction meeting the stereotype to only 32.7%. In other words, the average trajectory, which has been held up as established fact for more than 50 years, describes the behavior of only a minority of researchers, and a large majority of researchers follow qualitatively different trajectories.

Publication trajectories can be divided into four general classes based on the signs of the two slope parameters, \(m_1\) and \(m_2\), corresponding to the quadrants shown in Figure 3.4. Individual trajectory shapes exhibit enormous diversity, spanning all four quadrants. Even among faculty whose publication rates grew and then declined (lower right quadrant, 62.1%), the conventional narrative only includes the 32.7% of individuals whose rate of growth exceeds their rate of decline \((m_1 > |m_2|)\); shaded region, Figure 3.4. Additionally, the distribution of trajectory parameters extracted from DBLP data was statistically indistinguishable from the distribution derived from hand-collected CV data \((p = 0.95, \chi^2)\), confirming that the dispersion shown in Figure 3.4 represents the true diversity of careers.

The cloud of faculty trajectory parameters shown in Figure 3.4 does not naturally separate into coherent clusters. In their absence, what are the covariates that predict which region of the plot an individual is likely to occupy? First, early-career growth rate of yearly publications \(m_1\) is significantly higher for researchers at top-50 institutions \((p < 0.001, \text{one-tailed Mann-Whitney})\), growing at 1.21 additional papers per year in the top 50 compared with 0.75 for others. Perhaps as a result—what goes up must come down—the slope after the point of change, \(m_2\), is significantly more negative for researchers at higher-ranked institutions, compared to those at lower-ranked
institutions \((p < 0.001, \text{one-tailed Mann-Whitney})\). Additionally, researchers who received their doctorates at top-ranked institutions exhibit faster early-career growth than those who trained at lower-ranked institutions \((p<0.05, \text{one-tailed Mann-Whitney})\).

Second, the early-career initial productivity \(b\) is significantly higher both for those who are faculty in \((p < 0.001, \text{one-tailed Mann-Whitney})\) and those who graduated from \((p < 0.001, \text{one-tailed Mann-Whitney})\) top-50 departments. Unsurprisingly, we also find that researchers who have postdoctoral experience start out significantly more productive \((p < 0.05, \text{Mann-Whitney})\). These findings regarding \(m_1\) and \(b\) combine to suggest that current academic environment correlates with—and perhaps influences—productivity, while prior academic environment does not. Finally, faculty at top-ranked departments are statistically no more or less likely to be found within this triangular region, a result robust to alternatives cutoffs for “top ranked” institutions.

The relationship between trajectories and gender is more complicated. First, trajectories of male and female researchers were similarly distributed across the four quadrants \((p=0.52, \chi^2)\), and gender was uncorrelated with the likelihood of meeting the four criteria of the canonical narrative \((p = 0.93, \chi^2)\). Further, within this canonical subset, the women’s initial productivity grew at a rate indistinguishable from the men’s \((p = 0.14, \text{Mann-Whitney})\) and peaked in similar years \((p = 0.78, \text{Kolmogorov-Smirnov})\). Women’s initial productivity, however, was 0.26 publications lower than the men’s \((p=0.039, \text{Mann-Whitney})\), and this difference exists in spite of the fact that men and women in this subset trained and were hired at similarly-ranked institutions \((p > 0.05, \text{Kolmogorov-Smirnov})\), and completed postdoctoral training at similar rates \((p=0.45, \chi^2)\).

The change point within a career may indicate regime shifts in productivity, regardless of which type of trajectory an individual may follow. While the change-point parameter \(t^*\) does not correlate with the other parameters of \(f(t)\), its distribution reveals that for most faculty, the inferred change point in productivity rates occurs at approximately year 5, with the median at 5 years and the mode at 4 years. However, the piecewise \(f(t)\) may be overparameterized for individuals with near-linear publication histories, running the risk of slightly overfitting their trajectories and making their fitted \(t^*\) values uninterpretable as productivity change points.
Figure 3.5: Heat map of researchers’ inferred change points. Each researcher’s inferred change-point parameter $t^*$ is plotted as a heat map, sorted by the length of their career in our dataset and restricted to individuals whose productivity trajectories are both stable under the addition of noise (see text) and better modeled by Eq. (3.1) than a straight line, determined by AIC (see Supplemental Text A.5).

To correct for the possibility of overfitting, we performed model selection for stable trajectories, asking whether AIC with finite-size correction favored a straight line or the more complex $f(t)$ (see Supplemental Text A.5). This process conservatively selected only 30.2% ($N = 329$) of researchers who are more confidently modeled by the piecewise function, yet their distribution of productivity change points is largely unaffected. Figure 3.5 translates each selected faculty member’s career length and inferred change point into an ordered pair, creating a heat map of career change points. Shown in the accompanying marginal distribution, the modal value for $t^*$ is year five, closely preceding tenure decisions at most institutions. Nevertheless, there is still enormous diversity in career transitions, and the average remains misleading as the descriptor of a majority
of individuals.

Figure 3.6: Heat map of researchers’ most productive years. Each researcher’s most productive year (empirically; not model fit) is plotted as a heat map, sorted by the length of their career in our dataset. White box indicates researchers with fewer than 10 years of experience, whose most productive year is necessarily early. The marginal distribution (right) shows the empirically most productive year for all faculty in the dataset, separated by early career (first 10 years; grey) or later career (orange). The most common peak-productivity year is year 5, and only about half of senior faculty exhibit peak productivity in year 5 or earlier.

In fact, the trends and diversity observed in $t^*$ distributions remain true even when models are avoided entirely. A direct empirical examination of all DBLP and CV publication time series reveals that a computer science professor’s productivity is also most likely to peak in the fifth year, yet peak productivity can nevertheless occur in any year of a professor’s career (Figure 3.6). While the marginal distribution shows that 41.9% of faculty have their peak productivity within the first 6 years, with the modal peak year in year 5, there is substantial variance. Note, for example, that individuals along the bottom of Figure 3.6 published the most in their first year as faculty, while individuals along the diagonal published the most in their most recent recorded year as faculty.
3.1.3 Transitions in Authorship Roles

Finally, other transitions exist that are not quantifiable in publication counts alone, yet these are surprisingly well synchronized with the transitions noted above. As faculty train graduate students, their roles ordinarily shift from lead researcher to senior advisor or principal investigator, and this transition is commonly reflected in a shift from first author to last author. While common, this first/last convention is not universal. For example, papers in theoretical computer science typically order authors alphabetically, so the relative position of these researchers in the author list will not exhibit any consistent pattern over a career. To investigate career-stage transitions in author position, we first identified the set of journals or conferences that list authors alphabetically by computing whether each venue’s authors are alphabetized significantly more often than is expected by chance (\(\alpha = 0.05\)) and exceeding twice the expected rate (see Supplemental Text A.6). These conditions selected 11.9% of publication venues, accounting for 15.3% of all papers in the dataset.
which we manually verified includes all top theoretical computer science conferences and excludes all top machine learning and data mining conferences. We then discarded these alphabetically biased venues from the following analysis. The remaining data show clear evidence of a progressive shift toward last-authorship position over time, with the relative first/last proportion reaching stability around year eight (Figure 3.7). Interestingly, the onset of this change is earlier among top-50 faculty, and their average proportion of last-author papers is significantly higher than those of other faculty, consistent with a hypothesis that faculty at top-50 institutions tend to begin working with students earlier, and have larger or more productive research groups.

As with the aggregate trend in productivity over a faculty career (Figure 3.2), the transition from first- to last-author publications (Figure 3.7) is based on averaging across many faculty and thus may not reflect the pattern of any particular individual. To characterize individual performances, we compared the fraction of first-authored papers in the first three years post-hire to the same fraction in the second three years, for faculty with careers longer than six years ($N = 2036$). A substantial drop in this fraction across these two periods would be consistent with the average trend reflecting individual patterns. For this analysis, we treated single-author papers as first-author publications. Overall 70.1% of researchers undergo this transition, publishing a larger fraction of first-author publications in the first three years of their faculty career than in the second three years. These fractions are consistent for faculty at top-50 institutions (70.7%) and those at other institutions (69.7%), but individuals at top-ranked institutions appear to make the transition more quickly and completely by the end of the six-year period (Figure 3.8). In spite of these trends, there remains enormous diversity among first/last author transitions, reinforcing the notion that averages may be poor descriptors of many individuals.

3.2 Discussion

The conventional narrative of faculty productivity over a career is pervasive, with repeated findings reinforcing a canonical trajectory where productivity rises rapidly to a peak early in one’s career and then declines slowly [57, 31, 25, 33, 47, 99, 50, 7, 104, 86, 50]. This narrative shapes
Researchers at top-50 institutions ($\pi < 50$)

Researchers at institutions below the top 50 ($\pi \geq 50$)

Figure 3.8: First-author publication rates. First-author publications as fraction of total in the first three years post-hire, and the three years thereafter, shown separately for researchers who placed at a top-50 institution (top) and researchers placing outside of the top-50 (bottom). Individual researcher data are plotted as points on top of a corresponding heatmap in which darker color denotes higher density by Gaussian kernel density estimation. Researchers at all levels of prestige tend to move out of first-authorship roles during this period, though researchers at top-50 institutions transition more completely by years 3–5 than others.

Expectations of faculty across career stages, and publication counts have been shown to impact both
tenure decisions \[26, 69\], and the ability to secure funding for future research \[103\]. In this study, we showed that the conventional narrative, while intuitive, and certainly applicable to averages of many professors, is a remarkably inaccurate description of most professors’ trajectories. By applying a simple piecewise-linear model to a comprehensive dataset of academic appointment histories and publication records, we found that only about one third of tenured or tenure-track computer science faculty resemble the average, regardless of their department’s prestige.

While diverse, some aspects of a trajectory are nevertheless partially predictable. For example, although the diversity of trajectories remains unaffected, productivity does tend to scale with prestige: researchers who graduated from or were hired by top-ranked institutions are significantly more productive at the onset of their careers, and, furthermore, productivity of high-prestige faculty tends to grow at faster rates and achieve higher peaks than researchers employed by other institutions. Together, these results support previously suggested hypotheses that top-ranked universities both select for and facilitate productivity \[2\]. In fact, our results suggest that researchers at top-ranked institutions transition into leadership roles more quickly than others, further implicating facilitation effects in addition to selection.

The relationship between productivity trajectories and gender is complicated and requires careful study. Gender has been shown to correlate with differences in productivity across fields \[64, 118, 38\], but these relationships are complicated by prestige \[19\] and have also changed over time \[111\]. Other work has uncovered differences in collaboration patterns between subfields \[119\], as well as productivity differences that depend on both student and advisor genders \[85\]. Here, we found that men and women follow the canonical productivity narrative at equal rates. However, among those who do, we found significant differences in initial and peak productivities between men and women. Given the complications revealed in past studies, the extent to which these differences reflect inequalities, past or present, and contribute to women’s underrepresentation in computer science is an important topic of research and warrants thorough exploration in future work.

Within the space of career trajectories, there is a noticeable tendency toward peak productivities and shifts in publication rates around 5 years after beginning as faculty. This is surely
not a coincidence, given the fundamental role of tenure as a change point within the typical academic career, after which the total number of hours worked does not substantially change, but the time devoted to service tends to dramatically increase, with concomitant decreases in research and grant-writing [60]. However, our data cannot yet say how, from a mechanistic perspective, the existence of tenure requirements drives faculty to change or shape their productivity before or after promotion. If anything, the results in this paper make clear that there are numerous ways in which computer scientists meet promotion requirements, not all of which necessarily involve publishing a large number of papers. Indeed, in parallel with career shapes more broadly, there remains enormous diversity in the distributions of productivity peaks and change points. This diversity in overall production, combined with the observation that an individual’s highest impact work is equally likely to be any of his or her publications [100], implies there are fundamental limits to predicting scientific careers [20].

Computer science is, itself, a multifaceted field, and previous studies of the DBLP dataset revealed that productivity rates differ by subfield [111]. This observation, coupled with the zoo of fluctuating trajectories revealed here, may suggest that year-to-year differences in individual trajectories are related to which subfields a researcher studies. Past work has revealed a first-mover advantage associated with entry into a rapidly growing field [79], so changes to individual research interests may contribute to noisy trajectories, particularly if they coincide with concentrated growth of popular new subfields.

Larger and higher-resolution data sets may improve our ability to identify expanding new subfields and other factors that could explain or predict trajectories. Although DBLP has the advantage of covering computer science journals and peer-reviewed conferences alike, we found that its coverage of those venues was incomplete in predictable ways. By manually collecting CV data for 10% of the scattered trajectories shown in Figure 3.4 we adjusted DBLP data for missing publications and established the rate at which publishing rates have grown since 1970. Trajectories derived from DBLP data and benchmark CV data were statistically indistinguishable from each other. Investigations of productivity trajectories outside computer science will lack the
field-specific DBLP database and may require additional calibration, name disambiguation, and data deduplication.

The misleading narrative of the canonical productivity trajectory is not likely to be unique to computer science. Although future studies will reveal whether this claim is correct, the enormous diversity revealed here seems to demand a reevaluation of the conventional narrative of careers across academia. Other studies that investigate the impact of this pervasive narrative on decisions of promotion, retention, and funding would be particularly valuable. Expectations, whether perceived or enforced through tenure decisions, might give rise to some of our results. If these expectations vary from field to field, it is possible that while diversity remains a feature that spans academia, some types of trajectories may be more common in certain fields. Regardless of whether future work combining productivity and hiring data across fields reveals this to be true, models of faculty productivity will need to be revisited and revised.

Nevertheless, our results here once again suggest a complex relationship exists between the prestige of a researcher’s academic environment and her productivity. The collective productivity of a department, too, certainly affects its prestige or perceived quality in turn, creating a rich set of interdependencies, which complicate any analyses hoping to isolate the effects of specific mechanisms driving researcher productivity. The inability to disentangle the effects of such mechanisms has historically limited efforts to identify and prioritize opportunities to improve conditions for researchers at all levels of the prestige hierarchy. In the following chapter, we use a series of matched pair experiments to isolate and measure causal effects of researcher location on productivity and prominence. Additionally, we examine how early-career productivity in turn affects future location by predicting retention and relocation of faculty.
Chapter 4

The Effects of Departmental Prestige on Researcher Productivity and Prominence

The prestige of an academic institution is intimately linked to the prominence or perceived quality of its research faculty. Elite institutions house brilliant researchers who publish their findings in top journals and present at top conferences \[61, 42\]. Researchers at elite institutions produce the majority of scientific studies found in the literature \[23, 40\], and receive the greatest number of citations \[28, 43, 75\]. These institutions also produce and employ the vast majority of new researchers entering the professoriate \[70, 82, 19\], whose doctoral prestige and productivity decide the prestige of their initial appointment \[111\]. Awards for scholarship, too, including the prestigious Nobel prize, are frequently awarded to researchers at elite institutions \[121, 3, 95\]. In the field of computer science, the top 20 North American universities either trained or employed approximately 80\% of recipients of the prestigious Turing award\(^1\), the John von Neumann Medal\(^2\), and the Knuth Prize\(^3\).

Past research has outlined a number of correlations between an individual’s institutional affiliation and their success in research. In these studies, “success” is typically quantified in one of two ways: either (i) in terms of quantity or the number of contributions made to the scientific literature, or (ii) in terms of quality or the amount of impact her publications have had on influencing the progression of science, often measured by counting citations. Here, we refer to these performance

---

1 See http://amturing.acm.org; rankings from \[19\]
2 See http://www.ieee.org/about/awards/medals/vonneumann.html
3 See http://www.sigact.org/Prizes/Knuth/
measures as “productivity” and “prominence,” respectively, and success along either dimension has been shown to positively correlate with the prestige of the researcher’s affiliation.

Research faculty carry at least two academic affiliations, corresponding to their doctoral institution and their current employer, and previous research suggests that each plays an important role in driving an individual’s productivity and prominence \[28 \ 72 \ 40 \ 123 \ 91 \ 65 \ 37 \ 2\]. Intuitively, a person’s doctoral institution teaches them how to conduct research and how to publish their findings. By way of their doctoral advisors, graduate students are introduced to research practices, communities, and collaborators that affect what, where, and with whom the individual publishes in the future. Indeed, past work suggests graduates’ productivity mirrors their advisors’ \[62\]. Similarly, a person’s current, employing institution is thought to affect their access to facilities, quality graduate students, funding, and other resources important for producing research \[37\]. Changes in researcher location have also been found to correlate with changes in research performance, with increases in prestige corresponding to increases in productivity and prominence \[67 \ 2\]. Furthermore, social theory on conformity would suggest that research faculty are subject to group pressure, and therefore may attempt to match the publication efforts of their peers \[88 \ 67\].

The complex interdependencies between faculty’s past and present environments and their success as researchers pose significant challenges to administrators of science hoping to provide the necessary conditions for faculty to flourish at all level of the prestige hierarchy in academia. Specifically, they face a problem of endogeneity: if prestigious institutions both graduate and hire productive faculty, how can we determine if their success is due to selection at the hiring stage or facilitation post-hire? This raises the question, can the effects of training at an elite institution and placing into one be isolated? A second major complicating factor is one of selection bias. Past work tends to focus on the aggregate performance of entire departments, the majority of whom are tenured faculty, rather than junior faculty whose productivity may help or hinder their chances of being retained by their department. In order to test hypotheses of facilitation, one must therefore consider that there exists a second selection event, tenure evaluation, whose outcomes may artificially signal facilitation of success.
In this study, we utilize comprehensive data on the academic employment and education histories for all tenure-track faculty in computer science, and perform matched-pair experiments to isolate and infer the causal effects of training at and, separately, placing into an elite department. Our analyses focus where relevant on junior, un-tenured faculty to study both how research performance is affected by location and how, in turn, research performance predicts future location following tenure evaluations. The effects of location on junior faculty are especially interesting given that such individuals are continuing to learn how to succeed as a researcher while also recruiting students, securing funding, and fulfilling service obligations, all activities that potentially benefit from mentorship and other forms of departmental support. To this end, our findings shed new light on the mechanisms through which researcher productivity and prominence are selected for or facilitated by prestigious institutions.

4.1 Data

Our study makes use of several large, complementary datasets, which together provide comprehensive information on the education and academic employment histories of all tenure-track faculty in the field of computer science, along with each researcher’s scholarly publication and citation histories. The first dataset is a hand-curated collection of profiles for faculty at the 205 departmental or school-level academic units on the Computing Research Association’s Forsythe List of Ph.D.-granting departments for computing-related disciplines in the United States and Canada. Collected during the 2011–2012 academic year, this dataset provides partial or complete information on the education and academic appointment histories of 5032 regular faculty, assembled from publicly-available sources.

From this larger set of researchers, we selected the 2583 tenured or tenure-track faculty who both received their Ph.D. from and were hired to their first assistant faculty position by one of the 205 in-sample institutions in 1970–2011. Past work has established that approximately 87% of computing faculty have trained at and are employed by one of these institutions [19]. A number

---

4 See http://archive.cra.org/reports/forsythe.html
of faculty were removed in this step because the location or exact year of their first assistant professorship was not known; these were mainly senior faculty. Each faculty’s perceived gender was coded by data collectors as “male” or “female,” based on available images and the individual’s name. We believe the perceptions of the data collectors likely reflect those of the larger scientific community, yet we make no claims about whether these perceived genders align with researchers’ self-identifications.

To complement this first dataset, we constructed a complete record of the publication histories for the faculty included in our sample by linking faculty profiles together with author pages on DBLP, an online database that provides open bibliographic information for most journals and conference proceedings relevant to computing research, using manual name disambiguation as necessary. DBLP provides, for each publication on an author’s profile, a record of the paper’s title, its authors, publication venue, publication type (journal or conference paper; discarding pre-prints and other non-peer-reviewed formats), and year of publication. Following this procedure, we collected data for 200,476 publications which covered 2453 (95.0%) faculty in our sample.

In previous work, using these same data, we inferred subfield information for each individual faculty according to the titles of their publications. Additionally, we previously validated the DBLP dataset by manually collecting CVs for 10% of the included faculty, which we used to estimate DBLP’s increasing coverage and rising productivity levels over time. In this study, we reuse these previously inferred subfield characterizations and, according to rates calculated from the CV dataset, adjust all publication counts into 2011-equivalent levels.

Extending these two datasets, we also recorded citation histories as listed on faculty’s Google Scholar profile pages. Google Scholar provides an extensive record of citations in computer science, though it often lists multiple entries for the same paper, due in part to typos in paper titles and author names, or multiple versions retrieved from pre-print servers and publisher websites. With this in mind, we collected the raw number of citations amassed by any of an author’s papers.

---

5 See http://dblp.uni-trier.de
6 See http://scholar.google.com
through each calendar year, without removing self-citations or performing additional filtering. In total, we recorded Google Scholar citation information for 1586 (61.4%) faculty in our sample, who were collectively responsible over 7.4 million citations between 1970 and 2011. Faculty excluded from this subset were once again mainly senior faculty.

Finally, in order to investigate how productivity and prominence affect faculty retention and relocation, we updated all profiles belonging to faculty who were assistant professors at the time of our 2011 sample. Specifically, we successfully updated 555 of the 595 (93.3%) then-junior faculty profiles to include locations and job titles as of November in the 2016–2017 academic year, excluding individuals for whom neither DBLP nor Google Scholar information was available. Data for this updated sample were retrieved by three data collectors, who re-collected a random 10% of the records updated by each of their peers, allowing us to calculate inter-rater reliability. On average, only 7.5% of updated records differed between the collectors, and these conflicts were resolved manually in the finalized dataset. Of the 555 updated faculty profiles, 474 (85.4%) individuals remain tenure-track faculty at one of our in-sample institutions, with the other 81 (14.6%) having relocated to positions predominately outside of academia. We provide further, in-depth analyses of the retention and relocation of these select faculty in later sections.

4.2 Results

4.2.1 Location’s Causal Effects

Tenure-track faculty hold at least two academic affiliations over the course of their career, corresponding to their doctoral institution and their initial employing institution, and previous studies suggest that each contributes to researchers’ success in publishing [28, 72, 40, 123, 91, 65, 37, 2]. Here, we isolate the causal effects of each institution using matched pair experiments to determine which location(s) drive early-career performances, treating placement as a natural experiment that separates individuals who train at or place into more or less prestigious institutions. The answer to which affiliation(s) drive researchers’ success will inform our subsequent analyses.
into the underlying mechanisms of scholarly productivity and prominence, the identification of which will shape future policies of academic departments looking to enhance the scientific output of its researchers. We begin by investigating potentially continued advantages of training at elite institutions before shifting our focus to researchers' initial appointment locations.

**Prestige of doctoral institutions.** In order to determine whether a continued advantage is conferred to individuals who train at more prestigious institutions than their departmental peers, we constructed several matched pair experiments wherein individuals were matched according to a number of individual- and institution-level attributes. Specifically, these attributes were (i) the prestige of the hiring institution, (ii) the year of initial placement, (iii) the researcher’s gender, (iv) their inferred subfield, and (v) whether they received postdoctoral training. Matches were constructed by caliper matching, requiring that matched individuals be of the same gender, have both received (or both not received) postdoctoral training, and that their inferred subfield distributions were in the 90th percentile of similarity, measured by Jensen-Shannon Divergence (see Appendix B.1). Caliper widths for the remaining features were selected to be narrow enough such that differences in the matched individuals’ outcomes were not simply explained by differences in these features. That is, our thresholds for similarity along each dimension ensure that the more productive individuals are statistically indistinguishable from the less productive individuals.

Having matched individuals who placed similarly in the prestige hierarchy, we first tested whether the total productivity in the first five years post-hire (omitting the year of placement, \( y=0 \)) is significantly higher for the individual who trained at the more prestigious location. If continued advantage is afforded to graduates from more prestigious institutions, we would expect the matched pairs’ productivity to be biased in favor of the individuals with more prestigious degrees. Apply all matching criteria listed above, we formed a total of \( N = 359 \) matched faculty pairs and found no significant differences in their post-hire productivities (\( p = 0.59, t\)-test). In fact, only 52.1% of trials exhibited an advantage of any size (\( p = 0.23, \) one-tailed binomial test). Our results were similar for men and women, separately, and were unchanged by the exclusion of the subfield or postdoctoral experience criteria, or by restricting the analysis to researchers hired after
Figure 4.1: Early-career productivity is indistinguishable for similarly-placing but differently-trained peers: Average yearly productivity differences for 359 matched pairs of faculty. Positive values indicate higher productivity for graduates of more prestigious institutions, relative to matched faculty. Shaded region denotes 95% confidence interval for the mean. Year $y = -3$ displays a significant difference in favor of more prestigious graduates, and year $y = -1$ shows a significant difference in favor of less prestigious graduates.

Examining the yearly productivity of matched faculty (Figure 4.1), we found no significant differences in any of the five years post-hire ($p > 0.05$, $t$-test). There were significant differences one ($p < 0.05$, $t$-test) and three years ($p < 0.01$, $t$-test) prior to placement, however, the effects were reversed for these years, with more prestigious graduates being more productive in year $y = -3$, and less prestigious graduates being more productive in year $y = -1$. These differences offset each other in our evaluations of the entire five-year period.

We next considered the continued effect of doctoral institution on researchers’ prominence, measured as the total number of new citations acquired in each of a researchers’ first five years post-hire. Matching once again on the five previously mentioned criteria, we formed 129 pairs of matched researchers and found, as with productivity, that researchers’ doctoral institution’s have no continued causal effect on their prominence ($p = 0.06$, $t$-test), with 56.6% of trials exhibiting an advantage of any size ($p = 0.08$, binomial test).

Examining the yearly citation counts of matched faculty (Figure 4.2), we found no significant
differences in any of the five years before or after the individuals placed as faculty \((p > 0.05, t\text{-test})\). Given that citations continue to accumulate over time, however, the trend suggests that this advantage might continue to grow over time and possibly become significant later in individuals’ careers.

**Prestige of initial appointments.** Next, we investigated whether individuals who place at more prestigious institutions become more productive or more prominent than their similarly-qualified but lower-placing peers. As in our analysis of the prestige of researchers’ doctoral institutions, we once again constructed matched pair experiments using the five criteria listed above. To these pairs, however, we included a sixth attribute, requiring that matched individuals be equally productive or prominent in the five years prior to their initial placement.

Focusing first on researcher productivity, we considered whether placing at a more prestigious institution causally drives the total number of publications produced in a researcher’s first five years post-hire to be greater than a peer with similar training and past productivity. Applying all matching criteria, we formed 194 matched pairs of researchers, and found that the individual who placed at the more prestigious institution is more productive by, on average, 5.11 papers in
their first five years post-hire \((p<0.001, \text{ }t\text{-test})\), with 57.4\% of trials exhibiting an advantage of any magnitude \((p<0.05, \text{ binomial test})\). Once again, this result is robust to the exclusion of various matching attributes. Under all matching conditions, the magnitude of the productivity advantage was significantly correlated with the difference in prestige rankings of the matched individuals, with the higher-placing individual producing nearly one additional paper (0.74 papers) for every 10-rank difference over their matched peer \((p<0.005, \text{ }t\text{-test})\).

![Figure 4.3: Early-career productivity advantages quickly emerge for prestigious faculty: Average yearly productivity differences for 194 matched pairs of faculty. Positive values indicate higher productivity for faculty at more prestigious institutions, relative to similarly-trained faculty. Shaded region denotes 95\% confidence interval for the mean. Productivity differences appear after only one year following initial placement.](image)

The differences in yearly productivity for the matched pairs in our experiment (Figure 4.3) reveal that an almost immediate advantage is conferred to the higher-placing individual. In particular, we noted significant differences in years \(y = \{1, 2, 4, 5\} \) \((p<0.05, \text{ }t\text{-test})\), with a trend towards widening advantages over time. Past studies of career productivity trajectories, however, suggest a common change-point in productivity shortly following the five-year window we analyze here. We leave for future work to investigate the role of institutional prestige on mid- and late-career productivity. Whether these advantages persist or change over time and in response to tenure decisions represents an interesting and important question with potential implications for the future of departmental tenure policies.
Next, we investigated similar effects on researchers’ prominence, matching individuals according to their total number of citations in the five years prior to initial placement and comparing their citations counts in the five years following the year of their initial appointment. Applying all matching criteria, we formed 97 matched pairs and found that, like productivity, researchers who place higher become significantly more prominent than their peers with similar training and pre-hire prominence. Specifically, the researcher who placed higher received an average of 258.6 additional citations in the five-year period, though the median difference was much more modest at 11 additional citations. Yearly differences (Figure 4.4), as noted with productivity, begin to appear almost immediately after placement and become significant by three years post-hire. For there to be a lag in this measure, however, is unsurprising given the nature of citations and the rate at which research is recognized, incorporated, and improved upon in the literature.

Figure 4.4: Early-career prominence advantages conferred to prestigious faculty: Average yearly prominence differences for 97 matched pairs of faculty. Positive values indicate greater prominence for faculty at more prestigious institutions, relative to similarly-trained faculty. Shaded region denotes 95% confidence interval for the mean. Significant differences emerge three years following initial placement and continue to rise in the years thereafter.

Together, the results of our matched pair experiments suggest that previous studies noting a correlation between doctoral prestige and post-hire productivity and prominence are likely to have measured the indirect effect of doctoral prestige driving the prestige of researchers’ initial appointments, which in turn drives each of these performance measures. In contrast, we find that
doctoral prestige confers no significant continued advantage to researchers’ productivity and prominence beyond facilitating their initial placement, and potentially a small boost in prominence due to pre-hire work completed at prestigious institutions. Instead, researchers’ current environments ultimately shape their individual performances. For this reason and with the goal of informing policy to facilitate broad success in research at all levels of the prestige hierarchy of academia, we now shift our focus to explore the specific ways that prestigious institutions drive higher productivity and prominence.

4.2.2 Mechanisms Behind Employer Prestige

In the previous section, we found that placing at higher-ranked institutions causes researchers to become more productive and more prominent early in their careers. There are, however, several possible mechanisms that could give rise to such advantages. Broadly speaking, these mechanisms can be characterized into two general categories, corresponding to departments either requiring or facilitating greater success. These two categories align with what Allison and Long called the “selection hypothesis” and the “departmental effects hypothesis,” which the authors note are not incompatible but likely differ in their impact [2]. Knowing at least the relative impact of these mechanisms, however, would inform departmental tenure policies and hiring strategies at institutions hoping to enhance the success of their faculty.

There are four primary ways that a department might cause its faculty to be more productive.

(1) The department can select already-productive faculty at the hiring stage.

(2) Faculty might adapt or conform to the productivity standards set by their departmental peers.

(3) The department can select productive faculty once again during tenure/retention evaluations.

(4) The department might facilitate the productivity of its faculty by providing a variety of resources and incentives.
The first three mechanisms represent the ways that a department might either implicitly or explicitly require its faculty to be more productive, and the fourth encapsulates the set of actions it can employ to help its researchers realize their full potential. In the sections that follow, we briefly investigate whether these first three mechanisms alone can explain the early-career performance advantages afforded to faculty at elite institutions. The extent to which advantages can not be explained by these requirement mechanisms, then, serves as an indication of how well success is being facilitated by departments.

**Selection on hiring.** Previous studies, including our own [111], have noted significant relationships between researchers’ pre-hire productivity and the prestige of their initial faculty appointments. These investigations, however, often mix both junior and senior faculty, which introduces selection bias and a focus on the “survivors” of academia, individuals who were hired, retained through tenure, and continued working as faculty long enough to be included such studies. These survivors make up only part of the larger faculty workforce, and their pre-hire productivities are not necessarily representative of junior faculty. To address this limitation, in this and following sections we focus on the 555 faculty in our dataset who held the title of “assistant professor” during the 2011 sample year, a title that generally signifies pre-tenure status. These faculty, too, represent the select group of individuals who secured tenure-track faculty positions. We therefore investigate the extent to which placement in the prestige hierarchy sorts individuals according to their pre-hire productivities, keeping in mind that our data exclude individuals who sought but failed to achieve employment in the professoriate.

Analyzing individual performances in the five years pre-hire, we found that productivity correlates significantly (Spearman $\rho = -0.195$, $p < 0.001$) with the prestige of researchers’ initial appointments (shown in Figure 4.5, $p < 0.05$, $t$-test). The effect size, however, was modest: for every 10-rank increase in employer prestige, faculty produced on average an additional 0.28 papers over the five-year period. Prominence (Figure 4.6) also correlates with prestige during this period (Spearman $\rho = -0.359$, $p < 0.001$), with every 10-rank increase in prestige corresponding to 18.59 additional citations. Normalizing each researcher’s citation counts by their publication counts, we
Figure 4.5: Pre-hire productivity correlates with prestige: Total publications in the five years pre-hire versus the prestige of the individual’s initial appointment. Black line denotes ordinary least squares regression, with slope indicating that new faculty are on average 0.28 papers more productive for every 10-rank increase in prestige.

measured an average advantage of 1.59 additional citations per publication in the five-year period per 10-rank increase.

Figure 4.6: Pre-hire prominence correlates with prestige: Total citations received in the five years pre-hire versus the prestige of the individual’s initial appointment. Black line denotes ordinary least squares regression, with slope indicating that new faculty receive on average 18.59 more citations for every 10-rank increase in prestige.

These observations were largely unaffected by the inclusion of dummy variables to each regression. Specifically, we added variables corresponding to researchers’ gender and whether they
received postdoctoral training. Gender’s effect was significant in regressions of both publication and citation counts ($p < 0.05$, $t$-test), however, normalizing citation counts by publications eliminated the effect, suggesting that women receive similar numbers of citations per publication as men in our dataset, yet produce fewer papers. Further, we found that postdoctoral experience significantly impacts publication counts, but not citations (normalized or unnormalized). Postdoctoral experience was significantly linked to employment at elite institutions ($\pi < 50; p < 0.05$, $\chi^2$), yet conditioning on this experience did change our prior conclusions.

Together, our results suggest that researchers are similarly productive and prominent in the years leading up to placement, with only modest productivity advantages afforded to individuals graduating from and placing into prestigious institutions. These results also support the results of our matched pair analyses, which revealed that researchers’ current locations affect their productivity, suggesting that similar effects benefit researchers’ students. Next, we considered a second, more implicit form of requirement by investigating the extent to which researchers conform to departmental publishing standards.

Conforming to departmental productivity. Previous studies have indicated that faculty productivity is affected by social pressures to conform or adapt to the performances of departmental peers [67, 4, 9]. Though not an explicit requirement, social pressure could drive the advantages in productivity and prominence for researchers at elite institutions. In this section, we measure the extent to which adaptation occurs among computer science faculty, moving researchers closer to their departmental publishing standards.

Because selection during tenure evaluations might artificially signal adaptation by individuals, it is important to only consider pre-tenure faculty in this analysis. For this reason, we focus again on junior faculty, requiring that they both held the title of “assistant professor” and were at least five years post-hire in 2011, allowing us to evaluate their early-career performances in the context of their departmental peers. Further, we restricted our analysis to only include departments with at least three other faculty, so as to provide robust estimates of departmental publishing norms.

Applying these restrictions, we compared the early-career publication counts of 92 pre-tenure
faculty to their departmental peers. Specifically, we ranked all faculty according to their productivities in the five years before and, separately, the five years after being hired. We then determined for each pre-tenure faculty, whether their pre-hire or post-hire productivity rank was closer to the average post-hire rank of the faculty at their initial employing institution. Intuitively, adaptation would move individuals towards their departmental average. We found, however, that only 50 of the 92 faculty (54.3%; \( p = 0.23 \), one-tailed binomial test) appear to have adapted to their environments, either rising or lowering their productivities to match departmental norms. This suggests that social pressures, at best, subtly affect researcher productivity, and past studies may have instead measured the effects of selection at tenure stages enforcing productivity requirements. We now investigate the extent to which productivity predicts tenure decisions and affects both where researchers continue to be employed and the productivity standards at those institutions.

**Selection on retention.** Faculty are evaluated and selected at two stages of their career, first upon their initial hiring, and again when they apply for tenure. Past studies have investigated both the effects of tenure on future productivity and productivity’s impact on tenure outcomes \[39, 36, 71\], and universally, more productive researchers are more likely to achieve tenure status than their less productive peers. Having measured modest effects for both productivity advantages at the onset of faculty’s careers and adaptation to departmental norms, we next investigated the impact of early-career productivity performances on the retention and relocation of faculty in our dataset by exploring how well these performances predict 2016 statuses.

As previously noted, 474 of the 555 junior faculty (85.4%) in our follow-up sample were still employed as normal faculty at one of the 205 in-sample institutions. Among these faculty, gender was not significantly linked to continued employment \(( p = 0.36, \chi^2)\), whether employed by an elite institution or not. Of the 474 individuals still employed, 399 (71.9% of the original 555) were still located at their 2011 institution, 47 (8.5%) had moved to a more prestigious institution, and 28 (5.0%) had moved to a less prestigious institution. Men and women were once again distributed similarly over these categories, as were individuals from elite versus non-elite institutions. Shown in Figure \[4.7\] the likelihood of faculty leaving their initial appointment was distributed evenly by
Figure 4.7: Faculty leave their appointments at similar rates across the prestige hierarchy: Binned by departmental prestige, faculty leave their initial appointments, either for other institutions or out of academia altogether, at similar rates. Error bars indicate 95% confidence intervals around the means. Dotted purple line indicates the total fraction of department faculty from all 205 institutions.

After taking inventory of the 2016 outcomes for junior faculty, we used supervised machine learning to determine the extent to which each outcome class could be predicted based on early-career productivity and other attributes. First, to predict faculty who will leave the academy altogether, we applied 6-fold cross-validation and trained logistic regression classifiers on individuals’ productivity z-scores, calculated relative to their departmental peers. The AUC score for this task was 0.62, indicating that productivity alone allows for modest predictions of whether individuals will depart academia around tenure evaluations. Perhaps unsurprisingly, researchers with low productivity tend to be filtered out at this career stage (Figure 4.8). The inclusion of other covariates, like gender, prominence, and the prestige of the employing institution had little effect on AUC for this prediction task.

Using a similar setup, we found that the other classes of outcomes were predictable with similar accuracies. We achieved the highest prediction accuracy predicting transitions up the prestige hierarchy (AUC=0.65), using productivity z-scores and the rank change from individuals’ doctoral and initial employing institutions as feature in our prediction model. Interestingly, researchers who
incurred large rank changes in their initial placements tended to partially or completely reverse that rank change through relocation (Figure 4.9), suggesting that individuals who do not conform to departmental norms may self-sort into more appropriate publishing climates.

Our prediction results here indicate that while productivity and prominence offer some clues as to which faculty will be retained or will relocate in the early years of their career, these predictions are once again modest. Further, the rates at which faculty leave academia, possibly due to being denied tenure, are relatively consistent across the prestige rankings, suggesting that top-ranked institutions do not rely on selection at retention to maintain their high standards of productivity.

4.3 Discussion

In this study, we constructed a series of matched pair experiments and found that, while doctoral prestige correlates with higher publishing and citation rates, it has no significant, continued effect on graduates after they become faculty. To the contrary, these measures are significantly and causally affected by factors correlating with the prestige of their employing institutions. We then investigated three possible mechanisms that could give rise to the observed advantages, asserting that a fourth, facilitation, must be responsible for the portion that cannot be explained by the other
Figure 4.9: Faculty relocation often reverses initial rank differences. Faculty who relocate tend to undo the rank change incurred by their initial placement, returning to an institution similar in rank to where they received their doctorate. This reversal of rank-changes is similar for individuals placing in the top-50 (purple) and otherwise (black).

Having found modest effects these competing mechanisms, we conclude that researchers at elite institutions are not simply required to be more productive. Rather, our results suggest that elite institutions do facilitate greater performances, and potentially in ways that can be adopted by lower-ranked universities. Our work here therefore justifies future investigation into the precise resources and mechanisms by which elite institutions advance the careers of their faculty.

Despite having found modest effects for the non-facilitation mechanisms, our current analyses cannot offer concrete insight into the relative strengths of the four proposed drivers of productivity/prominence advantages. To address this limitation, future work might build on our analyses here to simulate faculty’s early-career productivity, modeling adaptation behavior of individuals and facilitation by departments. Such a simulation would allow for inferences at the level of individual departments, and commonalities between the departments exhibiting strong facilitation would provide insight into which resources and environmental conditions are most effective at eliciting greater productivity.
One additional mechanism, which we do not discuss but may benefit researcher productivity is collaboration. Over time, scientific research is increasingly performed by teams, rather by individuals, and elite institutions may foster more efficient collaboration among peers, giving rise to advantages in researcher productivity and prominence. Our current publication data lack author affiliations for individuals not included in our collection of faculty profiles, yet incorporating this information would allow future studies to examine collaborations effects on publishing performances and the retention or relocation of faculty.

Citations are an especially crude measure of impact or quality of research. Nevertheless, they remain popular, due in large part to their availability, and ours and past results demonstrate that they do yield predictive power. A factor that limits the usefulness of citations as a measure of success is that a citation can signal many different forms of acknowledgement, with varying interpretations of importance. The increasing availability of full-text articles online, however, presents opportunities for future work to not just count citations, but to examine their context and potentially infer their meanings.

Finally, our analyses here rely on the assumption that researchers in computer science act to maximize their own prestige, with departments ultimately deciding who is hired as faculty. Placements and relocations down the prestige hierarchy, however, are much more easily realized than transitions up the rankings, suggesting that researchers possess some notion of choice in opting for less-prestigious positions. The extent to which researchers might prefer movements down the hierarchy is interesting and has important implications for departments hoping to increase their productivity and prominence. In the event that a researcher is drawn to an institution because of their existing culture, their hiring may serve only to reinforce it, further securing the department’s position in the prestige hierarchy.

This raises an interesting question: is the status hierarchy of universities currently in stable equilibrium? And if so, is there anything that departments can do to increase their prestige? A department could, for example, decide to increase its productivity requirements, but doing so may drive away applicants who might prefer a similarly-ranked institution with less demanding
requirements. The department’s current prestige also appears to dictate its hiring power, meaning that simply hiring faculty from higher-ranked, more productive institutions may not be possible.

Finally, as previously mentioned, faculty may seek employment in a department specifically because of its existing culture.

This problem is not an unfamiliar one to the business world, though. Few businesses are poised to dethrone the leaders of industry, and yet many successful companies exist. Often, these companies succeed not by directly outcompeting larger, more established institutions, but through innovation and often specialization in a particular area. Might universities find success in similar approaches? Currently, computer science departments in particular tend to hire faculty that span the breadth of the discipline. Future research, however, should investigate whether explicit specialization might allow lower-ranked institutions to increase their standing within the prestige hierarchy.
Chapter 5

Conclusions and Future Directions

5.1 Conclusions

The results of the preceding chapters have shed new light on several key areas of research in the science of science domain, particularly those giving rise to systematic imbalances in the composition and productivity of computer science faculty, as well as interactions between these two attributes. In Chapter 2, we explored the complex role that gender plays in faculty hiring and found that gender does not appear to directly impact placement outcomes, but instead correlates with disparities in productivity, postdoctoral experience, and other qualifications. This is both reassuring in the sense that overt discrimination does not appear to play a hand in computer science faculty hiring, and helpful in the sense that future studies can focus their attention on differences in the preparation of faculty in order to pinpoint the sources of our noted disparities.

One such disparity identified in Chapter 2 was a sustained productivity gap separating male and female researchers in computer science. As a first step in understanding the development of early-career research productivity, Chapter 3 revisited the long-held assumptions for individual trajectories based largely on aggregate performances. Applying a simple, descriptive model to individual trajectories we found that, despite the persistence of the aggregate trend, individuals are hugely diverse, with less than a third of careers following the expected shape or “conventional narrative” present through the literature. Our results call for existing models of productivity to be revised, and crucially, to the extent that existing policies rely on assumptions derived from these models, they too will need to be revisited. Further, we found no significant differences in the early-
career trajectories belonging to male and female researchers beyond a lower initial productivity, which, again, suggests future work should focus on earlier stages in the academic career timeline to understand the mechanisms driving early-career disparities.

Finally, Chapter 3 revealed stark differences in the production rates of faculty at elite versus non-elite institutions. Faculty in prestigious departments are more productive at the beginning of their careers, grow their productivity faster than peers at less-prestigious universities, and, as a result, achieve higher peak productivities. New faculty hired to elite institutions also tend to train at elite institutions, however, making it difficult to separate the effects of training at versus being employed by top-ranked and highly-productive departments. Chapter 4 addressed this issue directly by constructing matched pair experiments to isolate and infer causal effects of the prestige of doctoral and employing institutions on researcher productivity and prominence. We found that, in opposition to previous correlative studies, researchers’ doctoral institutions confer no significant productivity advantage to their graduates’ early-career productivity and prominence. Instead, we find that the prestige of researchers’ current, employing institutions drives productivity and prominence, which we argue is only partially due to selection, that these elite institutions facilitate more successful careers. Our results here validate and lay the foundations future work to identify the exact underlying mechanisms through which universities enable greater productivity, which will inform departmental policies for institutions throughout the prestige hierarchy of academia.

In terms of specific policy recommendations, our investigations collectively suggest that differences between men and women’s publishing experiences emerge in graduate school and persist throughout their careers. We have found that such differences do affect placement outcomes and may also determine who applies tenure-track faculty positions and where they apply. With these considerations in mind, we suggest that departments make every effort to ensure that all graduate students—regardless of race, gender, ethnicity, or otherwise—be afforded the opportunity to publish and to attend academic conferences themselves and to share in the experiences of their peers. To facilitate earlier experiences, departments might find success in establishing required courses for developing students’ communication skills, both verbally and in writing, and for research design.
Such courses are largely absent from graduate curricula, and students are left to gradually develop these skills by working with their advisors.

In regards to productivity, our results show rich diversity in the timing of researchers’ success in publishing. Productivity can occur at any career stage, and policy-makers should be careful not to assume that an individual will never become (or re-become) productive based solely on their age or the number of publications listed on their CV. This recommendation is underscored by recent work showing that an author’s most influential work can occur at any point in their career [100]. In short, success in science is only somewhat predictable, and our results recommend against policies that attempt to make too strong of predictions or requirements for productivity.

Finally, our analyses of the environmental effects on productivity and prominence demonstrate that early-career success is largely determined by a researcher’s current location, not their doctoral institution. For departments looking to hire successful scientists, this implies that simply hiring the applicant with the most prestigious credentials is not necessarily an optimal strategy. Instead, priority should be given to candidates poised to maximally benefit from the institution’s resources. For smaller, non-elite departments especially, such a strategy may stand in conflict to the general approach of assembling departments that span the breadth of computer science research. Future studies should investigate whether departments would benefit from more explicit specialization, particularly in light of the ever-broadening scope of computer science research.

5.2 Future directions

The research of this dissertation provides some much needed clarity to the sources and repercussions of systematic imbalances arising from the complex forces that shape the composition and productivity of computer science faculty. As an integral part of this work, we have identified a number of interesting and important directions for future work to explore. First, our studies have focused exclusively on the field of computer science, and while we expect that our major findings to apply broadly throughout academia, subtle differences certainly exist and are worthy of future investigation. Beyond simply applying these same analyses to other fields, however, our work featured
can be extended in many ways, a few of which we will now discuss in the following sections.

5.2.1 The role of social networks in faculty hiring

Our analysis of faculty hiring in Chapter 2 made use of the fact that graduates in computer science report explicitly that they seek to maximize the prestige of their initial placements as tenure track faculty. Further, they report prioritizing appointments based on attributes that correlate with prestige, including access to quality graduate students, higher salaries, and larger startup packages. There are, of course, other factors that don’t necessarily correlate with prestige but do have a significant impact on where candidates apply and the offers that they accept. Of particular interest, new hires indicate that advisor recommendations and other social ties affected both where they applied and which offers they accepted. Hiring departments, too, benefit in these scenarios, due to increased exposure to applicants, which allows them to make more informed decisions that appease the desires of current faculty hoping to bring friends and collaborators into the department.

The data that we used in Chapter 2 are limited in ways that make it difficult to investigate how social ties impact hiring and retention of faculty. Namely, our data represent a snapshot of all faculty in 2011, which fails to provide a detailed characterization of the composition of departments in the years in which particularly older faculty were hired. To overcome this limitation, future research should make use of, at minimum, multiple snapshots to provide more accurate depictions of the departments to which candidates where applying. Ideally, studies would analyze time-series of yearly department rosters, though the construction of this dataset will require the automation of identity resolution and tracking of faculty in academia. One might imagine, however, that such an automation platform could easily extend to assemble records for all fields of science, allowing for comprehensive and comparative studies of the evolution of the scientific workforce and the connectedness of the disciplines they represent.

Integration of other data types, too, will allow future studies to overlay faculty hiring networks with social connections representing many kinds of interactions. For example, two faculty may know
one another from their doctoral (or earlier) studies, having been part of the same lab, co-authored papers together, or cited each other’s work. They may also interact on social media platforms, publish in and attend similar conferences. Each of these interactions corresponds to a different level of familiarity and may indicate the strength or sentiment of faculty relationships. Future work will reveal how these various relationships impact the hiring, retention, and communication of scientific ideas, driving the composition and productivity of faculty.

5.2.2 A generative model for career productivity

Chapters 3 and 4 both investigate the progression of researchers’ productivity and prominence as a function of their career age. As demonstrated, individual trajectories exhibit an incredibly diverse set of patterns, and yet the average trajectory of all researchers together is remarkably stable. This observation calls for the revisiting models of scholarly productivity and indicates that past efforts may have overemphasized and therefore overfit to aggregate trends while ignoring the heterogeneity present among individuals.

To motivate future efforts, consider the following model, which builds on three mechanisms and intuitions derived from both past work and the studies included in this dissertation. First, an individual’s productivity changes slowly over time, with a researcher’s productivity in one year being roughly equivalent to their productivity in the previous year, as in Markov chains or a simple random walk model. Second, as a researcher advances in her career, her service and teaching obligations may increase coinciding with increases in external, non-work-related factors that compete for individuals’ time and attention, often including childcare, assistance with aging parents, and, eventually, the decline of their own health. These factors suggest a dampening factor that increasingly attenuates productivity with age, which might be modeled as a bias in our random walk. And third, we noted in Chapter 4 that tenure decisions tend to filter out the least productive faculty in academia, as in first passage models.

Together these three intuitions suggest a simple model for the development of faculty productivity trajectories. Specifically, they prescribe a biased random walk model with a minimum
productivity barrier corresponding to tenure, typically five or six years into a researcher’s career. Such a model would require specifying as few as four parameters: (i) an initial starting productivity, (ii) a distribution for random walk step sizes, (iii) an attenuation or dampening factor, and (iv) a cut-off for productivity at the tenure evaluation stage.

![Figure 5.1: Random walk models capture aggregate productivity trends: Using normally-distributed step sizes, linear attenuation, and a strict minimum threshold for productivity, random walk models recapitulate aggregate trends for career productivity.](image)

For demonstration, we offer perhaps the simplest realization of this model, seeding productivity from empirical values and using normally-distributed step sizes, linear attenuation with age, and a strict lower bound on productivity for modeling retention. As shown in Figure 5.1, these simple mechanisms alone are sufficient to reproduce aggregate trajectories for computer scientists' productivity, while also producing a rich diversity of individual trajectories. The model tends to overestimate, however, the role of selection at tenure evaluations, filtering out an excessive fraction of faculty. This suggests at least one additional mechanism is missing from our simple model, one that captures the effects of faculty learning and working to grow their productivity. Future work might find success by incorporating a simple growth mechanism to our proposed model of career productivity, and including prestige as a parameter in that mechanism would serve to approximate departmental effects on the development of new faculty.
5.2.3 Better measures of impact

Finally, in Chapter [4] we explored researcher prominence, quantified as the number of citations received as a measure of their impact on the scientific discourse. Previous studies have identified a number of limitations to this approach [], and yet citations remain a common way of studying impact in scientometric studies and are prominently featured on author profiles on sites like Google Scholar. Part of the staying power of citations stems from the fact that they are readily available and easily counted. It seems unlikely that citations will fall out of favor in the scientific community, but there are several opportunities to make citation counts more useful as a measure of impact.

First, publicly-accessible, full-text articles are increasingly available online, enabling future studies to analyze not just the number of citations a paper receives, but also the nature of those citations. In practice, citations signal one of many types of relationships between articles. An author may cite an article that proposed a standard method for solving a particular problem; another that created a public data set; another that represents the earliest related work; and another that made a recent contribution, which inspired the author’s current work. Each of these citations represents a different intention on the part of the author, and with it, a different form of importance to the greater scientific literature.

Along these same lines, the nature of a paper (e.g. review article, methods paper, dataset, etc.) surely affects the nature and number of citations it receives. Being able to partition citations even crudely into these categories would provide additional and important context for citation counts by providing a mechanism for comparing similar types of papers and their relative amount of impact. Finally, the nature of citations that an article receives may shift over time as the work becomes increasingly distant from the current state of the art and serves a primarily historical role in the literature. This suggests a sort of lifecycle for the role or purpose that articles serve over time. Future work should investigate how these lifecycles depend on the discipline of science and how lifecycles themselves have evolved over time.
Finally, historically and throughout this dissertation, productivity and prominence are quantified separately for each individual faculty. As noted in previous chapters, however, scientific research is increasingly performed by teams, and our data suggest that the sizes of such teams have increased steadily over time, meaning that the already difficult task of disentangling and measuring the contributions of individuals will grow to be even more difficult in the future. In order to encourage researchers to take part in large, ground-breaking studies, evaluations of faculty will need to evolve to reward researchers for taking part in studies where their contributions are not easily separated from others’. Already, probabilistic topic models allow for inferences of author contributions to full-text articles \[93\], yet future efforts should seek to identify complementary forms of faculty contributions. Team sports offer, potentially, an interesting case study for disentangling individual contributions, where players adopt differing roles to assist their team. In basketball, for instance, player contributions are measured not just in points, but also in assists, rebounds, and steals, among other statistics. In academia, faculty teach classes, mentor students, perform research, and fulfill a wide variety of service requirements, though these contributions are rarely considered in concert. As detailed records of each becomes available, however, these total contributions should be explored together.

5.2.4 Hedgehogs and Foxes

Recent work suggests that interdisciplinary research has consequences for an individual’s productivity and prominence. Specifically, navigating large, multi-faceted projects that span several disciplines tends to be demanding and difficult, which limits the number of papers produced by the authors. However, these projects are often more prominent, more highly-cited, rewarding the authors for their investments. What’s less clear, however, are the consequences of researchers being “multi-disciplinary” or working on multiple research directions either at the same time or in sequence. An analogy offered by the Greek poet Archilochus would label researchers as hedgehogs or foxes. The fox, as the saying goes, “knows many things, but the hedgehog knows one big thing.” In terms of a career in science, is it more beneficial to be a hedgehog, who spends their entire career
focused on a single area? Or to be a fox, who explores many?

Intuition would suggest that productivity might suffer as a result of frequently changing research directions, which requires individuals to familiarize themselves with new corners of the literature and develop sufficient domain expertise to make novel contributions. On the other hand, studies have shown that researchers leverage diverse perspectives to devise clever solutions to difficult problems. Future studies should investigate the long-term outcomes and benefits of either strategy. In the case that diverse repertoires reign supreme, sabbatical policies should seek to promote the development of diverse and complementary skill sets.
Bibliography


Appendix A

Supplementary Material for Chapter 3

A.1 Collection of CV data

Performing the DBLP coverage analysis and adjustment discussed in Supplemental Text A.2 required a benchmark dataset with complete coverage of the publications histories of a representative subset of researchers. This Supplemental Text describes the relevant details of the collection of that benchmark dataset. We manually extracted lists of publication dates from publicly available curricula vitae (CVs) belonging to a random 10% of the $N=1091$ researchers with career lengths between 10 and 25 years and having publications in at least three distinct years. This last condition ensures that the piecewise linear model can be fit to the individual’s trajectory and excludes just 32 researchers from our analysis. Because of the high diversity of productivity trajectories, we chose 10% of individuals, uniformly at random, from each of the quadrants designated by the signs of the two slope parameters, $m_1$ and $m_2$, as shown in Figure 3.4. Specifically, names of researchers from each quadrant were randomly shuffled and then collected, in order, until reaching 10% of the total. However, individuals for whom a CV could not be found, or whose publicly available CV was last updated before 2011 were skipped, and other faculty were randomly selected in their place. Success rates for this exercise ranged between 66.3% and 86.6%, measured as the number of successfully extracted publication lists versus the total number of attempts. The majority of these failures were due to researchers having out-of-date CVs. Future studies should consider whether such partial records of researcher productivity are sufficient for analysis, as their inclusion would greatly improve success rates during collection.
A.2 General trends in productivity data

Meaningful trends in publication rates over a career can be confidently identified from raw publication counts only if two conditions are met. First, raw publication counts must be exhaustive, containing all peer-reviewed publications. Second, field-wide publication rates must be stationary over time. Due to the facts that DBLP data do not satisfy the former, and that computer science as a field does not satisfy the latter, raw publication counts recorded in the DBLP dataset must be adjusted to compensate before they can be analyzed. This Supplemental Text explains the details of two compensatory adjustments to DBLP data. We justify the first adjustment by providing a detailed analysis of time-varying fraction of publications covered by the DBLP dataset, anchored by a hand-collected benchmark CV dataset (see Supplemental Text A.1). We then justify the second adjustment by identifying a clear and significant overall growth in publication rates over forty years of computer science publication data. We conclude by discussing several possible explanations for why researcher productivity increases over time.

DBLP has indexed the overwhelming majority of current computer science publications, including peer-reviewed conferences and journals. However, while excellent today, this coverage has increased systematically over time, meaning that DBLP coverage is less complete for older publications and faculty. In order to quantify trends in the time-varying coverage of our DBLP data, we hand-collected the CVs of 109 faculty, representing 10% of individuals whose trajectories were shown in Figure 3.4, providing a set of benchmark publication lists (see Supplemental Text A.1).

For each year of our dataset, we compared the number of publications listed on individuals’ CVs to the number of publications listed on their corresponding DBLP profiles, selecting only peer-reviewed conference and journal publications from CVs. Comparing DBLP data with CV benchmarks two sets of counts reveals that DBLP coverage has increased linearly from around 55% in the 1980s to over 85% in 2011 (Figure A.1). DBLP coverage has grown at a rate of approximately 1.06% of additional coverage per year, with a 95% confidence interval indicating that this rate falls between 0.8% and 1.4%. Because the ratio of DBLP publications $y_{DBLP}$ to CV publications $y_{CV}$ is
Figure A.1: DBLP coverage improves for more recent publications. Fraction of all publications found in DBLP data compared to publication lists extracted from CVs of corresponding researchers, separated by year. Regression of these fractions reveals that DBLP coverage improves by approximately 1.06% each year. Shaded region denotes the 95% confidence interval for the regression.

well described by the line

\[
\frac{y_{DBLP}(t)}{y_{CV}(t)} = m_\alpha t + b_\alpha ,
\]

we use Eq. (A.1) to convert all non-benchmarked DBLP publication counts to CV-equivalent publication counts, with estimated parameters of \(m_\alpha = 0.010588\) and \(b_\alpha = -20.434804\).

After linearly adjusting all raw publication counts to correct for the expected DBLP coverage in a given year [Eq. (A.1)], we analyzed how individual researcher productivity has changed over the years spanned by our dataset. Due to the fact that we extracted and adjusted DBLP records for only authors in the faculty hiring dataset [19][111], a straightforward analysis of the number of per-person publications in each calendar year would feature a different mixture of career ages in each year. For example, adjusted publication counts from the 1970s would include only early-career researchers; late-career researchers in the 1970s retired long before our dataset was collected. Because the main text of this paper reveals systematic trends in productivity by career age, this straightforward counting technique would introduce bias.

To quantify the expansion of publication rates over time, without introducing career-age bias, we selected “indicator” career ages at which to measure productivity, and compared how productivity at specific points in the academic career has changed over time. The trend in publication growth is consistently positive for all indicator career ages, and as in the main text, publication rates after
year 5 tend to be higher than publication rates in the first four years. When these publication rates were normalized by their 2011 values, all indicator sets collapsed onto a common growth line (Figure A.2). Thus, while the indicator sets of \{0, 1, 2, 3, 4\}, \{5\}, and \{5, 10, 15\}, revealed that productivity grows at a rate between 0.84 and 1.48 additional papers per person per decade, their rates of growth are directly proportional to 2011 productivity, meaning that the shape of the canonical trajectory has not changed over time, but has simply expanded proportionally.

![Figure A.2: Individual productivity has increased over time. After adjusting for DBLP coverage (see Supplemental Text A.2), evaluation of average individual performances in select indicator years reveals that researchers have become more productive over time, growing at a rate of approximately one additional paper per decade. Shaded regions denote the 95% confidence interval for each regression.](image)

The relative slopes of publication rate expansion are consistent with the relative publication rates in the canonical “average” productivity trajectory (Figure 3.2). Indeed, growth in researchers’ first five years of productivity is relatively modest, which is expected since researchers, both historically and more recently, spend these years building their research programs by applying for funding and recruiting graduate students and postdoctoral researchers. On the other hand, productivity in year five—the year of or immediately preceding tenure evaluations at most institutions and, perhaps not coincidentally, the modal year of peak researcher productivity—grows at a faster rate of 1.48 additional papers per person per decade. These rates are both similar comparing years 5, 10, and 15, which describes changes across a larger window of career productivity.
Figure A.3: Annual publication rates have grown steadily. For faculty in this study, per-person annual publications have increase over time at a rate of approximately one additional paper every 10 years. This rate of growth affects researchers at all levels of prestige rank $\pi$ [19]. Slopes represent least-squares linear regressions, with shaded regions denoting corresponding 95% confidence intervals.

The relationship between 2011-equivalent publications and past publications is consistently linear over the time spanned by our dataset (Figure A.2) and is modeled well by

$$\frac{y_{CV}(t)}{y_{2011}(t)} = m_\beta t + b_\beta.$$  \hspace{1cm} (A.2)

We therefore used Eq. (A.2) to convert all CV-equivalent publication counts to 2011-equivalent publication counts, with estimated parameters $\hat{m}_\beta = 0.131873$ and $\hat{b}_\beta = -258.286620$. Thus, we applied the two transformations of Eq. (A.1) (Figure A.1) and Eq. (A.2) (Figure A.2) in series to publication data to produce the adjusted publication counts used in the main text, unless otherwise specified; benchmark CV data was used for the 109 individuals for whom we collected it, and for those individuals, only Eq. (A.2) was applied.

In the main text, we noted that, after applying these two linear adjustments, the median number of early-career publications per person per institution increases over time and correlates with prestige. Figure A.3 illustrates this trend, stratifying individuals into three levels of prestige, and revealing that production growth rates between higher and lower prestige departments have widened slightly but significantly over time ($p < 0.05$, two-tailed $t$-test).

This trend, observed for early-career publications (i.e., publications within the first ten years
of a career, for individuals with careers of ten years or longer) is no different from the trend for all post-hire publications and all researchers (Figure A.4). Median lifetime career productivity correlates significantly with prestige, and, as in early-career productivity, public and private institutions are similarly affected by this relationship. Using an ordinary least squares regression of productivity versus prestige including dummy and interaction terms for public/private status we found that the relationship between productivity and prestige is not significantly affected by public/private status ($p > 0.05$, $t$-test, for both public/private dummy and public/private-prestige interaction).

Figure A.4: Publications correlate with prestige of employing institutions. Dots indicate median number of publications per person per institution for all years post-hire, adjusted for growth in publication rates over time and ordered by institutional prestige. Effects of prestige are statistically indistinguishable for private (black) and public (orange) institutions. Shaded region denotes the 95% confidence interval for least squares regression.

While production rates have increased steadily over time and for all levels of prestige, we find that the imbalance in research production has decreased in recent years. Figure A.5 illustrates this inequality across researchers and time by showing the fraction $Y$ of all publications in our sample that were produced by the most productive fraction $X$ of all faculty (a Lorenz curve), for faculty first hired in each of the four decades that our data span and restricting analysis to only publications produced in the first five years of an individual’s career. As referenced in the main text, the Gini coefficients for productivity imbalance have declined, from 0.62 in the 1970s to 0.40 in the 2000s.
Many factors could potentially drive such a shift towards more balanced research production in science (including, for example, technological advancements and corresponding declines in research costs that may have leveled the playing field for researchers at less-prestigious universities), and we welcome future studies that explore this shift in more detail.

Finally, several possibilities exist that might explain why researchers are becoming more productive. First, the average number of co-authors per publication has steadily increased over time, allowing researchers to work on a larger number of projects. Second, the number of publication venues has also grown, providing more outlets for researchers’ work and potentially facilitating more specialized communities with faster peer review. Third, technological advances including improvements in computer architecture have benefitted researchers universally, increasing the speed at which results are both generated and published. Lastly, perhaps the perceptions of what constitutes the minimum publishable unit of research has changed over time, resulting in a larger number of shorter, more narrowly focused publications in recent years. Each of these possibilities represents
an interesting line of research that we leave for future investigations.

### A.3 Modeling framework

Equation (3.1) is the simple model used in the main text to parameterize the adjusted publication counts over the course of a career. Reproduced below, it consists of two lines with slopes $m_1$ and $m_2$ that intersect at time $t^*$. 

\[
\begin{align*}
    f(t) &= \begin{cases} 
    b + m_1 t & 0 \leq t \leq t^* \\
    b + m_1 t^* + m_2 (t - t^*) & t > t^* 
    \end{cases} 
\end{align*}
\] (A.3)

In this manuscript, we fit Eq. (A.3) to adjusted count data using least squares. However, there exist other regression frameworks that correspond to generative models for time series data. In this Supplemental Text, we discuss some of these alternatives which may be useful for future work that probes the stochastic mechanisms behind individual productivity trajectories.

#### A.3.1 Adjusting models instead of counts

Although publication time series are naturally count data, a regression framework that is naturally suited for counts, such as Poisson or Negative-Binomial regression, is not advised. Directly fitting raw counts using a Poisson or Negative-Binomial model would neglect the adjustments for both the coverage of DBLP and the time-varying changes in publication rates (Supplemental Text A.2). On the other hand, adjusted publications are non-integers, rendering them inappropriate for count regressions. However, there are alternatives that would allow for both count regressions and the adjustments of Supplemental Text A.2. These adjustments come with a price, however, due to the assumptions and free parameters that they introduce.

One alternative solution to fitting a model to adjusted publication data is to fit an adjusted linear model to raw publication data. In other words, adjust the model instead of the data. Due to the fact that this approach would preserve the data as counts, this model would be amenable to Poisson and Negative-Binomial regression frameworks. Adjusted publications $y_{2011}$ are related to
raw publications by the adjustment

\[ y_{2011}(t) = y_{\text{DBLP}}(t) \frac{1}{\hat{m}_\alpha t + \hat{b}_\alpha} \frac{1}{\hat{m}_\beta t + \hat{b}_\beta} \]  

(A.4)

where \( m_\alpha \) and \( m_\beta \) are slopes and \( b_\alpha \) and \( b_\beta \) are intercepts of the linear adjustments for DBLP coverage and publication expansion, respectively, and hats indicate that variables have been estimated from data (Supplemental Text A.2). Applying this adjustment to the model \( f \), which is to hold for adjusted publications, we get

\[ f_{\text{DBLP}}(t) = (\hat{m}_\alpha t + \hat{b}_\alpha)(\hat{m}_\beta t + \hat{b}_\beta)f(t - t_0) \]  

(A.5)

where \( t_0 \) is the initial year of a particular faculty member’s career, and \( t \) is the calendar year.

We now turn to details of Poisson and Negative-Binomial frameworks for fitting the adjusted model Eq. (A.5), and discuss the assumptions and parameters that they introduce. Future work that focuses on creating generative models to explain productivity trajectories must do so with full consideration of these assumptions.

A.3.2 Poisson Model

Consider a Poisson fit of Eq. (A.5) to a set of data given by \( \{t_i, y_i\} \). To simplify, let us be explicit about the dependence of \( f_{\text{DBLP}} \) on the four parameters, \( m_1, m_2, b, \) and \( t^* \), which we collectively refer to as \( \theta \).

\[ f_{\text{DBLP}}(t; \theta) = q(t)f(t - t_0; \theta) \]

where we have made clear that \( q(t) = (\hat{m}_\alpha t + \hat{b}_\alpha)(\hat{m}_\beta t + \hat{b}_\beta) \) does not depend on the model parameters \( \theta \). The likelihood is then

\[ P(\{t_i, y_i\}|\theta) = \prod_i \frac{e^{-q(t_i)f(t_i - t_0; \theta)} [q(t_i)f(t_i - t_0; \theta)]^{y_i}}{y_i!} \]  

(A.6)

Rather than maximizing \( P \), we will maximize log \( P \). Taking the natural log of both sides, we get

\[ \log P(\{t_i, y_i\}|\theta) = \sum_i \left\{ -q(t_i)f(t_i - t_0; \theta) 
+ y_i [\log q(t_i) + \log f(t_i - t_0; \theta)] - \log y_i! \right\} \]  

(A.7)
Note that the terms $-\log y_i!$ and $y_i \log q(t_i)$ do not depend on the parameters $\theta$, so they affect the value of the maximum but not its location in parameter space. Dropping them yields a log-likelihood score $\mathcal{L}$ of

$$\mathcal{L}({\{t_i, y_i}\}|\theta) = \sum_i -q(t_i)f(t_i - t_0; \theta) + y_i \log f(t_i - t_0; \theta). \quad (A.8)$$

Note that for any trajectory, $q(t_i)$ can be precomputed and does not depend on the parameters $\theta$. Thus, fitting the 2011-equivalent Poisson model requires that we maximize Eq. (A.8) with respect to $m_1$, $m_2$, $b$, and $t^*$. This equation must be maximized numerically.

While this adjusted Poisson model is attractive because it naturally fits count data, it imposes assumptions on the data-generating process that are not justified empirically. Namely, the variance and mean of a Poisson distribution are equal, meaning that the Poisson regression expects the same of the data it explains. Future work that revisits the mechanistic models of faculty publication trajectories using generative processes will need to consider this rather strong assumption.

### A.3.3 Negative Binomial Model

The Poisson model above enforces the constraint that the mean is equal to the variance. However, there is no indication that the data support this assumption so we introduce the standard alternative, the Negative Binomial model. This model requires both a mean $\mu$ and a heterogeneity parameter $\zeta$, such that the probability of a single observation $y$ is

$$P(y) = \frac{\Gamma(y + \frac{1}{\zeta})}{\Gamma(y + 1)\Gamma\left(\frac{1}{\zeta}\right)} \left(\frac{1}{1 + \zeta \mu}\right)^{\frac{1}{\zeta}} \left(\frac{\zeta \mu}{1 + \zeta \mu}\right)^y \quad (A.9)$$

As in the Poisson regression, we will once more parameterize the mean using the piecewise linear model as $\mu_i = \mu(t_i) = q(t_i)f(t_i - t_0)$. However, we must also introduce a model for $\zeta(t_i)$.

The easiest way forward, mathematically, is to set $\zeta(t_i) = \zeta$. This assumes equal heterogeneity around the expected value $\mu(t_i)$ for all time points in a career $t_i$. Note that this assumption decouples the heterogeneity from the mean, while under the Poisson model they are directly coupled. One might think of the Poisson model therefore as fitting the parameters $\theta$ to both the trend and the
fluctuations together. The fixed-$\zeta$ Negative Binomial model, on the other hand, fits the parameters $\theta$ to the trend and uses a fixed $\zeta$ to accommodate all fluctuations. In this sense this Negative Binomial approach is more flexible, and uses an additional parameter to gain that flexibility.

The $\zeta(t_i) = \zeta$ assumption results in a log probability of

$$
\log P(\{t_i, y_i\}|\theta, \zeta) = \sum_{i=1}^{T} \left\{ \log \Gamma \left( y_i + \frac{1}{\zeta} \right) - \log \Gamma(y_i + 1) \\
- \log \Gamma \left( \frac{1}{\zeta} \right) - \frac{1}{\zeta} \log \left[ 1 + \zeta q(t_i) f(t_i - t_0) \right] \\
+ y_i \left( \log \left[ \zeta q(t_i) f(t_i - t_0; \theta) \right] - \log \left[ 1 + \zeta q(t_i) f(t_i - t_0; \theta) \right] \right) \right\} \tag{A.10}
$$

and we note that $\sum_{i=1}^{T} \log \Gamma \left( \frac{1}{\zeta} \right) = T \log \Gamma \left( \frac{1}{\zeta} \right)$, and that both $\sum_{i=1}^{T} \log \Gamma(y_i + 1)$ and $\sum_{i=1}^{T} y_i \log q(t_i)$ are constants that do not depend on either $\theta$ or $\zeta$, allowing us to write a log-likelihood score of

$$
\mathcal{L}(\{t_i, y_i\}|\theta, \zeta) = -T \log \Gamma \left( \frac{1}{\zeta} \right) + \sum_{i=1}^{T} \left\{ \log \Gamma \left( y_i + \frac{1}{\zeta} \right) \\
- \frac{1}{\zeta} \log \left[ 1 + \zeta q(t_i) f(t_i - t_0) \right] \\
+ y_i \left( \log \left[ \zeta f(t_i - t_0; \theta) \right] - \log \left[ 1 + \zeta q(t_i) f(t_i - t_0; \theta) \right] \right) \right\} \tag{A.11}
$$

Progress here, however, is obstructed by the difficulties of taking derivatives of Gamma functions. Thus, the above equation must be optimized numerically over the parameters $\theta$ and $\zeta$.

While these calculations may be helpful in seeding a way forward in future work, it is important to note that the fixed-$\zeta$ Negative Binomial model also makes strong assumptions about the generative process that created the data. Indeed, the assumption that fluctuations are uniform over an entire career is strong, and is not justified by data. One could also avoid this assumption, but this introduces additional problems, which we now discuss.

The temptation to let each point $t_i$ have a parameterized value of $\mu(t_i)$ and a free parameter of $\zeta(t_i)$ results in overfitting. Note that this would allow each point in the time series $(t_i, y_i)$ to be fit by a negative binomial distribution with a mean given by Eq. (A.5) and an arbitrarily large or small $\zeta(t_i)$, resulting in dramatic overfitting. This approach therefore makes few assumptions, but provides little value to the modeler.
A middle ground between fixed $\zeta$ and unrestricted $\zeta(t_i)$ would be to parameterize $\zeta(t_i)$ using a lower dimensional model. Using the same model for $\zeta(t_i)$ as we used for $\mu(t_i)$ would be similar, in principle, to the Poisson regression. Using a different model for $\zeta(t_i)$ is an option, but would require, again, a deep focus on the underlying mechanisms hypothesized to explain fluctuations in productivity.

### A.3.4 Modeling outlook

Generative models for productivity trajectories would be enormously valuable. In this Supplemental Text, we emphasized the assumptions made by the generative models underlying various regression frameworks. In particular, we derived models that are able to be fit directly to raw count data by including the inverse of the adjustment derived in Supplemental Text A.2, a quadratic term referred to as $q(t)$.

In terms of impacts, fitting the Poisson model and the fixed-$\zeta$ Negative Binomial model to the trajectories investigated in this paper do affect the parameters of individuals’ trajectories. However, they do not diminish the diversity of trajectories that we observe. Indeed, the example trajectories shown in Fig. 3.4 are only subtly affected by the use of one type of generative model or another. An in-depth investigation of parameterized stochastic generative models for faculty productivity trajectories is left open for future work.

### A.4 Sensitivity to timing of publications

Publication generally signifies the conclusion of a research project but the exact date when an article is published can depend on many factors, including the availability of reviewers, graduation deadlines for graduate students, delays between acceptance and publishing, synchronization with conference submission deadlines, as well as non-academic constraints, such as the impending birth of a child. Each of these factors might advance or delay a publication’s appearance in the literature, and furthermore, the effort associated with each publication may span weeks, months, or years. As a result, publication years serve as a noisy indicator of when productivity occurs.
To ensure that our findings are not due to coincidence in the timings of researchers’ publications, we examined the sensitivity of our results to the addition of small amounts of noise. For each researcher and for each of their publications, we added noise drawn from a normal distribution ($\mu = 0$, $\sigma = 0.7413011$) to the publication year and then rounded to the nearest whole-year. In expectation, this process leaves one half of publication years unaffected and shifts 22.8% by one year in either direction, 2.1% by two years, and 0.04% by three years. We repeated this process 200
times for each researcher (examples shown in Figure A.6), and found that the median number of trials in which the trajectory changed shape compared to the noise-free fit (i.e., $m_1$ or $m_2$ changed sign) was 9 (4.5%; see Figure A.7).

While the typical individual’s parameters are robust to noise, those individuals whose trajectories featured few publications or whose noise-free model parameters were near zero were far more likely to change shape. In fact, for 10.5% of individuals, noise led to shape change more often than not, i.e. in greater than 50% of noisy repetitions. Therefore, as an additional check, we asked whether the model parameters inferred for researchers’ noise-free trajectories differ significantly from those inferred for their 200 noise-added trajectories. We used Fisher’s method to combine $p$-values for each researcher and found no significant differences in any of the four model parameters ($m_1$, $m_2$, $b$, and $t^*$). Evaluated separately, fewer than one percent of researchers’ inferred model parameters differ significantly under the noise-free and noise-added fits. We conclude that the general shape of productivity trajectories is robust to small differences in publication year.

Figure A.7: Trajectory shapes are robust to perturbations in publication years. Applying the piecewise linear model to 200 noise-added publication trajectories for each researcher, the median fraction of trials resulting in sign changes of model parameters $m_1$ or $m_2$ compared to noise-free fits is 0.045.

A.5 Model selection

When the complexity of a model exceeds the complexity of the underlying data, some parameters of the model may no longer be interpreted as meaningful. Although the piecewise linear model
of Eq. (3.1) has only four parameters—two slopes, $m_1$ and $m_2$, an intercept $b$, and a change point $t^*$—it may nevertheless overfit productivity trajectories that are actually linear. This Supplemental Text provides additional details for our model selection procedures that avoid the overinterpretation of the piecewise change point parameters $t^*$.

If a publication trajectory is generated by a straight line with added noise, then fitting a piecewise model will result in $m_1$ approximately equal to $m_2$, and the location of the change point $t^*$ will be arbitrary. We apply model selection to identify individuals with careers lengths between 10 and 25 years ($N = 1091$) whose productivity trajectories are both stable (see Supplemental Text [A.4]) and consistently better modeled by the piecewise model Eq. (3.1) than a straight line (i.e., ordinary least squares or OLS). This filter includes only those trajectories whose change points $t^*$ can be interpreted with confidence.

To perform model selection, we consider three information theoretic model selection techniques: Akaike information criterion (AIC), Akaike information criterion with finite-size correction (AIC$_c$), and Bayesian information criterion (BIC), defined as

- **AIC**: $n \log(SSE/n) + 2k$
- **AIC$_c$**: $n \log(SSE/n) + 2k + \frac{2k(k+1)}{n-k-1}$
- **BIC**: $n \log(SSE/n) + k \log(n)$,

where $n$ is length of the individual’s career, $k$ is the number of model parameters [$k = 2$ for OLS, and $k = 4$ for Eq. (3.1)], and $SSE$ is the sum of squared errors (differences) between actual and modeled publication counts. AIC is the least conservative of the three methods, finding that 52.6% ($N = 574$) of individuals are better modeled by Eq. (3.1) than OLS regression. By contrast, AIC$_c$ and BIC select only 30.2% ($N = 329$) and 39.3% ($N = 429$) of individuals, respectively. In the main text, we adopt the most conservative approach, AIC$_c$, but the other two methods nevertheless produce qualitatively similar distributions for $t^*$, with the modal year for $t^*$ remaining at year 5.
A.6 Detection of alphabetized publication venues

Conventions of author order vary widely in computer science. In the first/last convention, first authorship is reserved for the lead author or primary contributor to the study, while last authorship indicates the senior author who oversaw or advised the work. In the alphabetical convention, borrowed from mathematics, a paper’s authors are arranged alphabetically by last name. If the trend from first-authorship toward last-authorship over a career is to be reliably interpreted (Figure 3.8), publications with alphabetical author orders must be discarded. This Supplemental Text explains the methods used to statistically identify and remove publication venues that are highly enriched with alphabetical conventions.

Our approach is to count the number of multi-author papers in each publication venue with alphabetically ordered authors and compare this count to the number expected by chance. (We note that all single-author papers are ignored in this analysis.) A paper with $M$ authors will list its authors alphabetically by chance with probability $1/M!$. Noting the number of authors on each multi-author paper published by a particular venue and assuming independence of ordering decisions, we derive an empirical distribution representing the number of coincidentally-alphabetized author lists and ask whether the venue adopts the convention significantly more often than would be expected by chance. Additionally, we require that the number of observed alphabetized lists be at least twice the expected value. These two conditions ensure that the alphabetical convention is both significant and widespread in its adoption in a particular venue. We find that 631 of the 5622 (11.9%) distinct venues in our dataset alphabetize their author lists. These venues account for 27,232 of the 177,437 (15.3%) multi-author conference or journal publications for which the publication venue is known.

Manually inspecting the list of alphabetized venues reveals that popular theoretical venues like STOC (Symposium on Theory of Computing), FOCS (Foundations of Computer Science), STACS (Symposium on Theoretical Aspects of Computer Science), and SODA (Symposium on Discrete Algorithms) adhere to the alphabetical convention, while WWW (World Wide Web Con-
ference), CSCW (Conference on Computer-Supported Cooperative Work and Social Computing), KDD (Conference on Knowledge Discovery and Data Mining), CHI (Conference on Human Factors in Computing Systems), and AAAI (AAAI Conference on Artificial Intelligence) do not, matching our expectations.

### A.7 Least squares fit of $f(t)$

A least-squares fit of the continuous piecewise-linear equation $f(t)$ given in Eq. (3.1) involves minimizing the sum of squared errors over four parameters, $m_1, m_2, b$, and $t^*$. In this Supplemental Text, we provide details that make this fitting process rapid and maximally accurate.

The model fit consists of two steps. First, we assume that $t^*$ is fixed, and find the optimal values of $m_1, m_2, b$. In the second step, we search for the $t^*$ whose corresponding optimal parameters provide the best fit. The sum of squared error $\varepsilon$ is given by

$$\varepsilon = \frac{1}{2} \sum (m_1 t_i + b - y_i)^2 + \frac{1}{2} \sum (m_1 t^* + m_2 (t_i - t^*) + b - y_i)^2$$  \hspace{1cm} (A.12)

where $t^*$ is the change point, $\sum$ denotes the sum for all $t_i < t^*$, and $\sum'$ denotes the sum for all $t_i \geq t^*$.

In the first step, we imagine $t^*$ to be fixed and simply take partial derivatives with respect to the three parameters, set each equal to zero, and solve. Setting $\nabla \varepsilon = 0$ yields three equations,

$$m_1 \left[ \sum t_i^2 + \sum t^* \right] + m_2 \sum t^* (t_i - t^*) + b \left[ \sum t_i + \sum t^* \right] = \sum y_i t_i + \sum y_i t^*$$

$$m_1 \sum t^* (t_i - t^*) + m_2 \sum (t_i - t^*)^2 + b \sum (t_i - t^*) = \sum y_i (t_i - t^*)$$

$$m_1 \left[ \sum t_i + \sum t^* \right] + m_2 \sum (t_i - t^*) + b \left[ \sum 1 + \sum 1 \right] = \sum y_i + \sum y_i \cdot$$  \hspace{1cm} (A.13)

Thus, for a fixed value of $t^*$, the above equations provide a linear system of three equations for the three unknowns $m_1, m_2, b$. The system can be solved numerically using any linear solver.

In the second step, we embed the optimization above within a search for the optimal $t^*$ value. For each proposed value of $t^*$, we use Eq. (A.13) to find the optimal $m_1, m_2, b$ and
then compute the associated error using Eq. (A.12), choosing the $t^*$ that minimizes $\varepsilon$. Initially, we propose a coarse grid at the level of $\Delta t^* = 0.1$, and then refine the grid by an order of magnitude locally around the best result, repeatedly, until the optimal $t^*$ is known to single precision. This procedure is fast, and due to the result of Eq. (A.13), limits numerical search to a one-dimensional line.
Appendix B

Supplementary Material for Chapter 4

B.1 Measuring subfield similarity

Computer science is a large and diverse field of study in which publication and therefore citation rates vary according to subfield. For this reason and to ensure that the faculty included in our matched pair analyses represented fair matches, we required that the individuals publish in similar areas. To this end, we used topic vectors inferred for individual faculty as part of our previous study on faculty hiring [111]. In summary, for each faculty in our dataset, we combined all of her paper titles into a single document, forming a bag of words characterization of the individual’s research. We then used latent dirichlet allocation [10] with $k=10$ topics to jointly infer (i) topics, as distributions over words representing subfields or themes in the faculty’s research, and (ii) individuals’ distributions over those topics, denoting their area(s) of expertise.

Next, to identify faculty with similar topic vectors, we computed the Jensen-Shannon divergence between all possible pairs of individuals. Figure B.1 shows the resulting distribution, from which we defined “similar” as having a divergence score in the 90th percentile of similarity with respect to the total population. Finally, having calculated Jensen-Shannon divergence as the symmetrized variant of the Kullback-Leibler divergence using a logarithm with base two, our threshold for the 90th percentile of subfield similarity was 0.294 bits.
Figure B.1: Subfield similarities for all faculty: Distribution of Jensen-Shannon divergence scores for all possible pairs of faculty in our dataset. The shaded (purple) region denotes scores within the 90th percentile of similarity (see text), and only faculty pairs within this range were included in our matched pair analyses.