Collaboration, Stars, and the Changing Organization of Science: Evidence from Evolutionary Biology *

Ajay Agrawal
University of Toronto and NBER
ajay.agrawal@rotman.utoronto.ca

John McHale
National University of Ireland, Galway
john.mchale@nuigalway.ie

Alexander Oettl
Georgia Institute of Technology
alex.oettl@scheller.gatech.edu

July 2013

Abstract

We report a puzzling pair of facts concerning the organization of science. The concentration of research output is declining at the department level, but increasing at the individual level. For example, in one field of science, over the period 1980 to 2000 the fraction of citation-weighted publications produced by the top 20% of departments fell from approximately 75% to 60% but over the same period rose for the top 20% of individual scientists from 70% to 80%. We speculate that this may be due to the rising burden of knowledge and falling communication costs, which together could increase the returns to collaboration, particularly across institutions, and amplify the role of stars by enabling a more finely disaggregated division of labor. We report descriptive evidence that is consistent with our conjecture on the rising role of stars, their increasing propensity to collaborate and to do so over increasing distances and differences in department rank, and the effect of their individual locations decisions on the overall distribution of human capital. We speculate on the efficiency of the emerging distribution of scientific activity given the localized externalities generated by stars.

JEL Classifications: O31, O33, I23, J24, L23.

*This research was funded by the Centre for Innovation and Entrepreneurship at the Rotman School of Management and the Social Sciences and Humanities Research Council of Canada. We thank Adam Jaffe, Ben Jones, and seminar participants at the NBER pre-conference workshop on "The Changing Frontier: Rethinking Science and Innovation Policy" for valuable feedback. Errors remain our own.

1 Introduction

The spatial organization of science is undergoing a fundamental transformation. New patterns of institutional participation, division of labor, and star scientist centrality are emerging. Given the essentially combinatory nature of invention and innovation, changes in organization that affect access to knowledge and ease of collaboration to produce new knowledge are potentially of great importance to aggregate technological progress and economic growth.[1]

In this paper, we document and discuss significant changes in the spatial organization of science over recent decades in the field of evolutionary biology. We identify two trends that appear contradictory at first glance. The first is the falling concentration of scientific output across institutions. As the base of science broadens, more institutions are participating in scientific research, and relatively more activity is migrating to lower-ranked institutions. The second is the rising concentration of output across individual scientists. Publications and citations have always been highly skewed towards star performers. However, we find that the relative importance of stars has increased in recent decades.

What could explain these trends? We see two large, seemingly opposing forces acting on the spatial organization of knowledge production. However, although they act in opposite directions on the relative returns to co-location, they both increase the relative returns to collaboration. The first is the rising "burden of knowledge" (Jones, 2009). The increasing depth of knowledge required to work at the scientific frontier is leading to rising returns to specialization. This in turn raises the returns to collaboration, given the need to combine ideas and skills to produce new ideas. Indeed, utilizing the sudden unexpected release of previously hidden knowledge caused by the collapse of the Soviet Union as a natural experiment, (Agrawal, Goldfarb, and Teodoridis, 2013) report evidence that an outward shift in the knowledge frontier does in fact cause an increase in collaboration, consistent

[1]For influential work that emphasizes the role of combining ideas in the generation of new knowledge, see Romer (1990), Jones (1995), Weitzman (1998), and Mokyr (2002).

with the knowledge burden hypothesis. Furthermore, all else equal, to the extent that co-location lowers the cost of collaboration, the rising burden of knowledge increases the returns to co-location.

The second is the improvement in collaboration-supporting technologies that reduce the barriers created by distance, e.g., email, low-cost conferencing, file-sharing technologies (Agrawal and Goldfarb, 2008; Kim, Morse, and Zingales, 2009). All else equal, these advances allow a greater physical dispersal of collaborating scientists. Thus, these forces potentially conflict in terms of their impact on the relative returns to co-location but both should lead to rising collaboration.

We report evidence on the extent to which collaboration has grown. Furthermore, we find that the average distance between collaborators has grown in terms of both physical distance and the rank-separation of collaborating institutions. We also show how the base of institutional participation has grown, including the entry of institutions from emerging economies. Notwithstanding such broadening in the reach of collaborative activity, we find that individual-level output is becoming more concentrated. In particular, we observe an increasing importance of star scientists, albeit stars operating increasingly as part of cross-institutional teams.

Finally, drawing on parallel work on the causal impact of stars on departmental performance (Agrawal, McHale, and Oettl, 2013), we speculate on the efficiency of the emerging spatial distribution of scientific activity. Recognizing the existence of knowledge, reputational, and consumption externalities associated with the location decisions of star scientists, we make no presumption that the resulting spatial distribution of stars is efficient. We find that stars attract other stars and also that the recruitment of a star can have positive effects on the productivity of co-located incumbents working in areas related to the star. These effects appear to be particularly strong when the recruitment takes place at non-top-ranked institutions.

However, while strong forces may lead to star agglomeration due to localized externalities, lower-ranked institutions may have strong incentives to compete for stars as a core part of strategies aimed at climbing the institutional rankings. We document significant movement up the rankings for a select set of institutions that begin outside the top-ranked institutions. There may also be congestion effects from star co-location due to clashing egos and increasing returns to "vertical collaboration" across skill sets located at different institutions. Overall, we find a tendency towards reduced concentration of the fields best scientists at its top-ranked institutions. Thus, fears of excessive concentration of stars due to positive sorting might be overblown, although research on the normative implications of the observed changes in the organization of science is still at an early stage.

We structure the rest of the chapter as follows. In Section 2, we describe the construction of our evolutionary biology data at the institutional and individual levels. In Section 3, we report evidence of the broadening institutional and international base of scientific activity and describe the increasing concentration of individual productivity. We document the rising importance of stars in Section 4. In Section 5, we report the striking patterns of increased collaboration in terms of average numbers of collaborating scientists, the average physical distance between collaborating scientists, and the average distance between the ranks of collaborating scientists. Finally, in Section 6, we provide a more speculative discussion of possible normative implications of these participation, concentration, and collaboration patterns, with an emphasis on the role of the location of stars.

2 Data

Our study focuses on the field of evolutionary biology, a sub-field of biology concerned with the processes that generate diversity of life on earth (e.g., the origin of species). As in most fields of science, research in evolutionary biology consists of both theoretical and experimental contributions. In addition to specializing in particular topic areas, empiricists also often specialize in working with particular organisms such as *Biston betularia* (English moth), *Drosophila melanogaster* (fruit fly), and *Gasterosteus aculeatus* (three-spined stickleback fish). The returns to species specialization result from, for example, the upfront costs of learning how to work with a particular species (including, in many cases, learning where to find them and how to catch and care for them to facilitate reproduction in order to observe, for instance, the variation in genotypes and phenotypes of offspring for multiple generations) as well as setting up the infrastructure in a lab or in nature to study them.

2.1 Defining Evolutionary Biology

We examine 255 evolutionary biology departments over a 29-year period: 1980–2008, inclusive. We delineate the field, identify participants, impute department membership, and construct individual and department-level output measures using bibliometric data from the ISI Web of Science. Specifically, employ the following approach.

We collect data on all articles published during the period 1980 through 2008 in the journals associated with the four main societies associated with the study of evolutionary biology: Society for the Study of Evolution, Society for Systematic Biology, Society for Molecular Biology and Evolution, European Society of Evolutionary Biology. Their journals are *Evolution*, *Systematic Biology*, *Molecular Biology* and *Evolution*, and *Journal of Evolutionary Biology*, respectively. We focus on these four society journals to ensure that every article published within them is an evolutionary biology article and of relevance to evolutionary biologists. In other words, unlike general interest journals such as Science and Nature, which include papers from evolutionary biology but also research from many other fields, these four journals focus specifically on our field of interest. This process yields 15,256 articles.

We next collect all articles that are referenced at least once by these 15,526 society journal articles. There are 149,947 unique articles that are referenced at least once by the set of 15,256 evolutionary biology society articles. We call this set the corpus of influence in that all of these referenced articles have had some impact on an evolutionary biology article.

We then weight this corpus of influence by how many times each article has been cited by an article published in the set of 15,256 evolutionary biology society journal articles within five years of publication. There are 501,952 references from the 15,256 society journal articles to the 149,946 corpus of influence articles. We use the 149,946 articles to construct our publication count measures and the 501,952 references to these articles to construct our citation-weighted publication count measures.

The key benefit of this approach as opposed to simply using the ISI Journal Citation reports' field definitions is that it allows us to include general journals that evolutionary biologists are likely to publish in, such as *Science*, *Nature*, and *Cell* (among others).

2.2 Identifying Authors

We next attribute the 149,946 articles in the corpus of influence to individual authors. One problem with the ISI Web of Science data is that until recently it listed only the first initial, a middle initial (if present), and the last name for each author. Since our empirical objective is to trace the movement of evolutionary biologists across departments, it is first necessary to disambiguate authors (that is, to distinguish J Smith from JA Smith). We rely on heuristics developed by Tang and Walsh (2010) to disambiguate between authors who share the same name. The heuristic considers backward citations of two focal papers. If two papers reference similar papers (weighted by how many times the paper has been cited, i.e., how obscure or popular it is), then the likelihood of the papers belonging to the same author increases, and we link the two papers to the same author. We repeat this process for all papers with authors who have the same first initial and last name. We exclude scientists who do not have more than two publications linked to their name.

2.3 Identifying Scientist Locations

Using the generated unique author identifiers for each evolutionary biology paper, we next attribute each scientist to a particular institution for every year they are active. A scientist is active from the year they publish their first paper to the year they publish their last paper. Here again, we must overcome a data deficiency inherent within the ISI Web of Science data; until recently, the Web of Science did not link institutions listed on an article to the authors. Instead, we impute author location using reprint information that provides a one-to-one mapping between the reprint author and the scientists affiliation. In addition, we take advantage of the fact that almost 57% of evolutionary biology papers are produced with only a single institution listing. We thus are able to directly attribute the location of all authors on these papers to the focal institution.

We should note that this method of location attribution is more effective for evolutionary biology than many other science disciplines since articles in this field are generally not produced by very large teams relative to other disciplines in the natural sciences (3.32 average authors per paper).

3 Participation: A Broadening Base

The first trend in the organization of evolutionary biology we document is a decline in the skew of the distribution of output. This may reflect increasing emphasis in knowledge production across previously lesser-producing institutions concerned about rankings who thus increasingly emphasize research output as a factor in promotion and tenure, changing preferences of faculty who have spent increasing time developing specialized research expertise, and/or mounting political pressure to distribute government funding more evenly across institutions and political jurisdictions. In addition, we find a dramatic increase in research activity in emerging economies, possibly reflecting a broader movement towards higher value-added activities as part of the economic development process.

In Figures 1 to 3, we report evidence of the broadening base of science in terms of the declining department-level concentration of scientists, publications, and citations, respectively. Specifically, we plot Gini coefficients to illustrate the distribution of scientists (publications, citations) across departments by year. The pattern of falling concentration is pronounced for the period between 1980 and 2000, although there is some indication of a turnaround in this pattern after 2000.

In Figure 4, we plot department-level Lorenz curves for publications and citations. These curves illustrate the overall shift in the distribution over time. For example, the top 20% of departments produced 60% of all publications in 1980 but only 50% in 2000. Similarly, the top 20% produced 75% of all citation-weighted publications in 1980 but only 60% in 2000. It is important to note that we use a balanced panel for these analyses, including only the 255 institutions that published in evolutionary biology throughout the period under study. In other words, we do not allow for entry of new institutions part way through the study period. Since most institutions that ever were meaningful contributors to this field were active throughout our study period, this is not a serious restriction.

However, we relax the no-entry restriction for the data we use in the next graph where we plot output by country because several institutions in previously low-income countries were not active in the early years but have since become increasingly important in the overall production of knowledge. We plot the increasing importance of institutions based in emerging markets in Figure 5. The growth rate of publications from institutions based in BRIC countries (Brazil, Russia, India, China) began to increase dramatically from the early 1990s onwards and increased fortyfold by 2000.

These decentralization findings from university-based research in evolutionary biology are broadly consistent with prior findings on the decentralization of innovative activity more broadly (Rosenbloom and Spencer, 1996; Bresnahan and Greenstein, 1999). For example, in a recent study of innovation in ICT over almost the identical period as our study (1976-2010), Ozcan and Greenstein (2013) examine US patent data and find that although the top 25 firms account for 72% of the entire patent stock and 59% of new patents in 1976 they account for only 55% and 50%, respectively, by 2010. The decline is even more dramatic when they restrict the sample to the ownership of high-quality patents (82% down to 62%). They interpret their results as supporting the view that decentralization is resulting from "more widespread access to the fundamental knowledge and building blocks for innovative activity" (p.5).

Overall, we interpret our data as reflecting a decline in the concentration of output at the department level. In other words, the top institutions are producing a declining fraction of the overall output, and previously lesser-producing institutions are now contributing an increasing portion of overall output. However, this is not the case at the individual level. We turn to this unit of analysis next.

4 Concentration: The Increasing Importance of Stars

With greater democratization in knowledge production across departments, is science becoming a less elite activity, with a falling centrality of stars as they compete with scientists from an ever-widening base? One might expect the broadening base of science at the department level to be accompanied by a reduction in the concentration of output at the individual level. However, we find evidence of the opposite.

We again plot Gini coefficients by year using citation-weighted publications, but this time at the individual level. These data, illustrated in Figure 6, indicate a significant increase in concentration during the 1980s and then relative stability during the following decade. Then, in Figure 7, we plot individual-level Lorenz curves for 1980, 1990, and 2000 with the same data to illustrate how the full distribution shifts over time. Again, we see individual-level output increasing over time. For example, the top 20% of scientists produced 70% of output in 1980 but 80% by 2000, with most of the shift occurring in the first decade. Furthermore, in Figure 8, we illustrate the increasing spread between the top-performing scientists and the rest by comparing the number of citation-weighted publications required to be in the Top 50, which increases fivefold, to the average number of citation-weighted publications, where the increase over the same time period is negligible.

How might we reconcile decreasing concentration at the department level but increasing concentration at the individual level? The answer may lie in the changing patterns of collaboration. Recall that although the rising burden of knowledge and declining communication costs exert opposing forces on the returns to co-location, both increase the returns to collaboration. We turn to the topic of collaboration next.

5 Collaboration: Increasing Across Distance and Rank

The trend towards increasing collaboration is a well-documented feature of the changing organization of science (for example, see Wuchty et al., 2007). Moreover, this collaboration has been increasingly taking place across university boundaries (Jones et al., 2008). We document this phenomenon in our setting in Figure 9. Specifically, this figure illustrates the steady increase in the average number of authors on evolutionary biology papers, rising from 2.3 in 1980 to 3.8 in 2005.

We also observe a dramatic trend in the average rank difference between authors on co-authored papers (Figure 10). This increase in rank separation is most pronounced for papers that include a star. The average rank separation roughly doubles from 30 to 60 for star papers between 1980 and 2005 (Figure 11). There has also been a large increase in the rank separation – from under 30 to above 50 – for non-star papers. Furthermore, we find evidence of increasing distance between collaborators over time. We illustrate this in Figure

12 where the average distance between coauthors increases from 325 to 500 miles over the period 1980 to 2005.

6 Discussion: Normative Implications of Star Location

Our review of the basic trends in participation, concentration, and collaboration reveals the dramatically changing organization of scientific activity in the field of evolutionary biology. The emerging picture also points to the increasingly central role played by stars in collaboration and overall output. Moreover, stars, like the overall research community, appear to be increasingly collaborating across distance and institution rank. Overall, we see evidence of a developing cross-institutional division of scientific labor, with stars playing a leadership role in institution- and distance-spanning multi-author research teams.

The rising centrality of stars raises questions about the efficient distribution of stars across institutions. We thus finally reflect on the efficiency of the emerging pattern of the division of labor, drawing on both the factual picture just documented and parallel work on the causal impact of star scientists at the departmental level (Waldinger, 2012, 2013; Agrawal, McHale, and Oettl, 2013). A key question is whether the emerging spatial distribution of stars is efficient from the perspective of maximizing the value of scientific output.

We do not presume that the distribution will be efficient, given the free location choices of individual scientists and the productivity, reputational, and consumption externalities associated with those choices. We note in particular that the reputational spillover from locating at top-ranked institutions could lead to an excessive positive sorting of stars at these institutions. Such inefficiency, if it exists, could be ameliorated by easier cross-institution collaboration, effectively making the location of stars less important to knowledge production. Even so, given the ongoing costs of distance-related collaboration, a concern still remains that there may be excessive concentration from a social welfare perspective.

In Agrawal, McHale, and Oettl (2013), we show that the arrival of a star, whom we define as a scientist whose output in terms of citation-weighted publications is above the 90th percentile of citation-weighted stock of papers published up until year t_{-1}, leads to a significant increase in the productivity of co-located scientists. More specifically, we show this effect operates through two channels: knowledge and recruiting externalities. We show that the arrival of the star leads to an increase in the productivity of incumbents, those scientists already working in the department prior to the arrival of the star, but only for those incumbents working on topics related to those of the star. We do not find any evidence of productivity gains by incumbents working on topics unrelated to those of the star. These effects are robust to including controls for broader departmental and university expansion. Furthermore, they are robust to placebo tests for the timing of the effect; there is no evidence of a pre-trend in terms of increasing productivity prior to the arrival of the star. Moreover, the results are also robust to using a plausibly exogenous instrument for star arrival.

The stars arrival also leads to a significant increase in subsequent joiner quality (recruits hired after the arrival of the star), which is most pronounced for related joiners but also occurs for unrelated joiners. These results also hold when subjected to the robustness tests described above. These recruiting results raises a concern about the possibility of reputation-driven positive sorting at top institutions, with stars attracting stars irrespective of productivity-increasing knowledge spillovers. This in turn raises a concern about lost opportunities for stars to seed focused, dynamic research clusters at lower-ranked institutions.

But are these opportunities lost? Given the apparent role of star recruitment in department building – which our evidence suggests would be particularly effective where the institution already has a cadre of incumbents working in related areas to the star and has a sufficient flow of new openings to take good advantage of star-related recruitment externalities – an offsetting force to excessive concentration could come from the incentive of lower-ranked institutions to use star-focused strategies to ascend departmental rankings.

Figure 13 shows how departmental rankings changed between 1980 and 2000. While these data imply a reasonably high degree of rank persistence, they also show that some institutions made significant movements up the rankings. Anecdotal evidence suggests that the recruitment of stars may have played an important role here.

Furthermore, stars may increasingly benefit institutions they do not join, but where they have collaborative relationships. Azoulay, Graff Zivin, and Wang (2010) and Oettl (2012), who both use the unexpected death of star scientists to estimate their effect on the productivity of their peers, report evidence that stars significantly influence the productivity of their collaborators. Moreover, Agrawal and Goldfarb (2008) show that the greatest effect of universities connecting to Bitnet, an early version of the internet, in terms of influencing cross-institution collaboration patterns, was not between researchers at tier 1 institutions, but rather tier 1 – tier 2 collaborations. One interpretation of this result is that lowering communication costs particularly benefits vertical collaboration, suggesting an increasingly vertically disaggregated division of labor as communication costs fall. Perhaps, for example, declining communication costs increase the returns for individuals at top institutions specializing in leading major research initiatives, identifying key research questions, and writing grant applications, where their collaborators at lower ranked-institutions run experiments, collect and analyze data, and work together with all collaborators to interpret and write their results. The results reported by Kim, Morse, and Zingales (2009) are consistent with this when they document the rise of lesser-ranked universities.

To obtain more direct evidence of changes in star concentrations, we plot in Figure 14 the share of the top 100 evolutionary biology scientists at the top 50 evolutionary biology departments. The basic pattern shows, if anything, a fall in the concentration of stars at top institutions, somewhat allaying fears of excessive concentration due to reputation-driven positive sorting.

Our examination of the efficiency of the emerging organization of activity in the field of evolutionary biology is unavoidably preliminary and speculative given current levels of knowledge. The broad pattern of increased spatial and cross-institution collaboration – often centered on a star – is pronounced in the data. However, despite institution-level evidence of reputation-based sorting, we do not observe the feared rise in concentration at top institutions. Given the importance of the spatial and institutional distribution of stars to the workings of collaborative science, we expect the normative implications of the changing spatial distribution of scientific activity – and its stars – to be an active area of future research on the organization of science.

References

Agrawal, A. and A. Goldfarb (2008, September). Restructuring research: Communication costs and the democratization of university innovation. *American Economic Review 98*(4), 1578–90.

Agrawal, A. K., A. Goldfarb, and F. Teodoridis (2013). Does knowledge accumulation increase the returns to collaboration? evidence from the collapse of the soviet union. University of Toronto mimeo.

Agrawal, A. K., J. McHale, and A. Oettl (2013, July). Why stars matter. University of Toronto mimeo.

Azoulay, P., J. Graff Zivin, and J. Wang (2010). Superstar extinction. *Quarterly Journal of Economics 125*(2), 549–589.

Bresnahan, T. F. and S. Greenstein (1999). Technological competition and the structure of the computer industry. *The Journal of Industrial Economics 47*(1), 1–40.

Jones, B. F. (2009). The burden of knowledge and the "death of the renaissance man": Is innovation getting harder? *The Review of Economic Studies 76*(1), 283–317.

Jones, B. F., S. Wuchty, and B. Uzzi (2008). Multi-university research teams: shifting impact, geography, and stratification in science. *science 322*(5905), 1259–1262.

Jones, C. I. (1995, August). R & d-based models of economic growth. *Journal of political Economy 103*(4), 759–784.

Kim, E. H., A. Morse, and L. Zingales (2009). Are elite universities losing their competitive edge? *Journal of Financial Economics 93*(3), 353–381.

Mokyr, J. (2002). *The gifts of Athena: Historical origins of the knowledge economy.* Princeton and Oxford: Princeton University Press.

Oettl, A. (2012). Reconceptualizing stars: Scientist helpfulness and peer performance. *Management Science 58*(6), 1122–1140.

Ozcan, Y. and S. Greenstein (2013). The (de)concentration of sources of inventive ideas: Evidence from ict equipment. Northwestern University mimeo.

Romer, P. M. (1990, October). Endogenous technological change. *Journal of Political Economy 98*(5), S71–S102.

Rosenbloom, R. S. and W. J. Spencer (1996). *Engines of innovation: US industrial research at the end of an era.* Harvard Business Press.

Tang, L. and J. P. Walsh (2010). Bibliometric fingerprints: Name disambiguation based on approximate structure equivalence of cognitive maps. *Scientometrics 84*(3), 763–784.

Waldinger, F. (2012). Peer effects in science: Evidence from the dismissal of scientists in nazi germany. *The Review of Economic Studies 79*(2), 838–861.

Waldinger, F. (2013). Bombs, brains, and science: The role of human and physical capital for the creation of scientific knowledge. University of Warwick mimeo.

Weitzman, M. L. (1998). Recombinant growth. *The Quarterly Journal of Economics 113*(2), 331–360.

Wuchty, S., B. F. Jones, and B. Uzzi (2007). The increasing dominance of teams in production of knowledge. *Science 316*(5827), 1036–1039.

Figure 1: Gini Coefficients by Year for the Distribution of Scientists across Departments

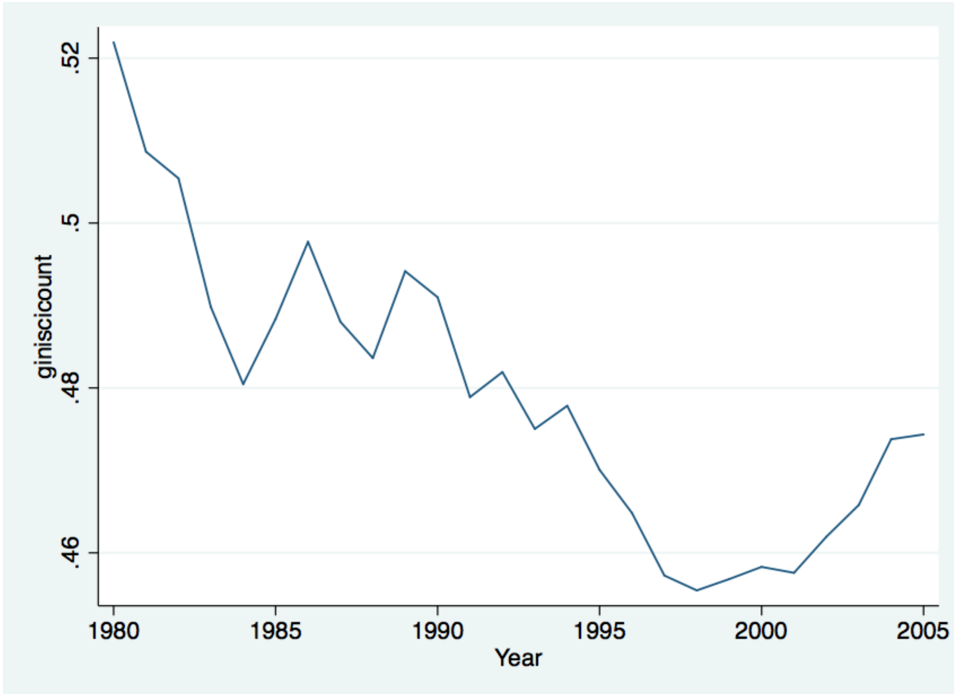

Figure 2: Gini Coefficients by Year for the Distribution of Publications across Departments

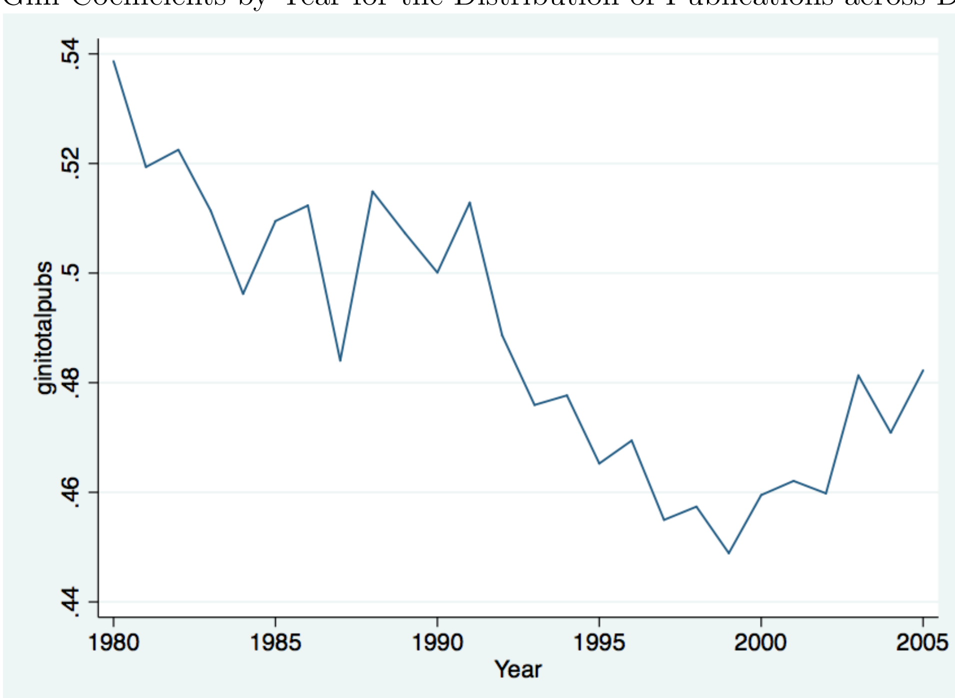

Figure 3: Gini Coefficients by Year for the Distribution of Citations Received across Departments

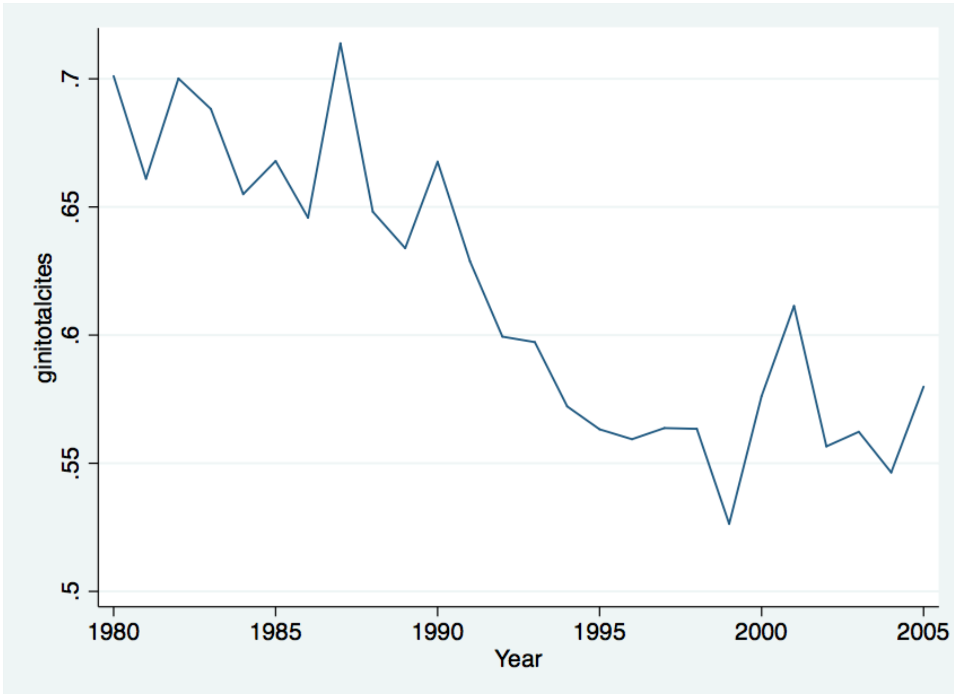

Figure 4: Lorenz Curves by Department for Publications and Citation-Weighted Publications

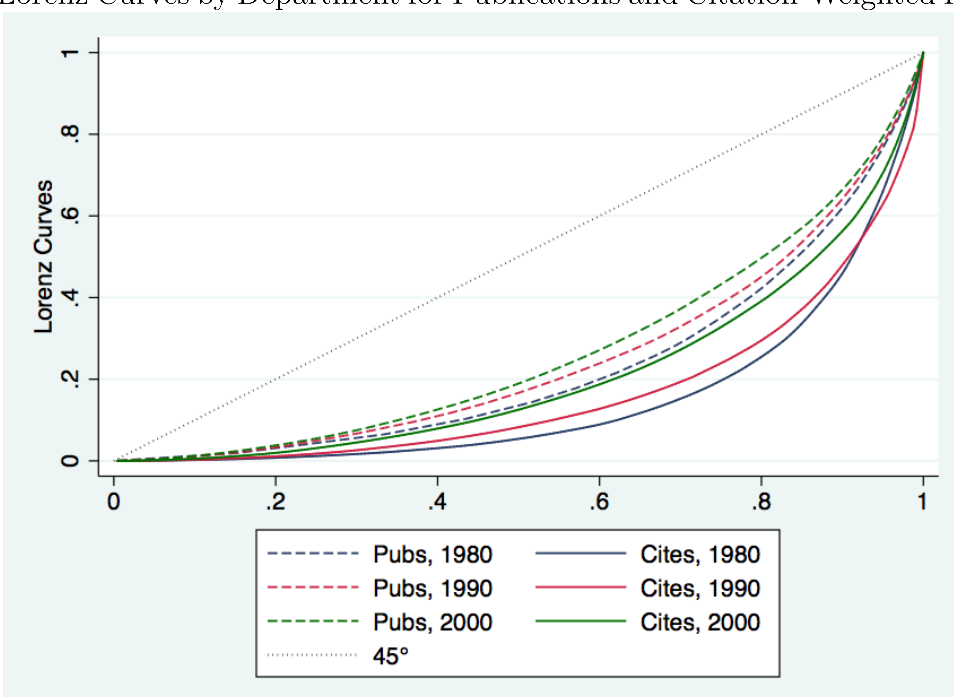

Figure 5: Publication Count by Country by Year Normalized Relative to Output in 1980

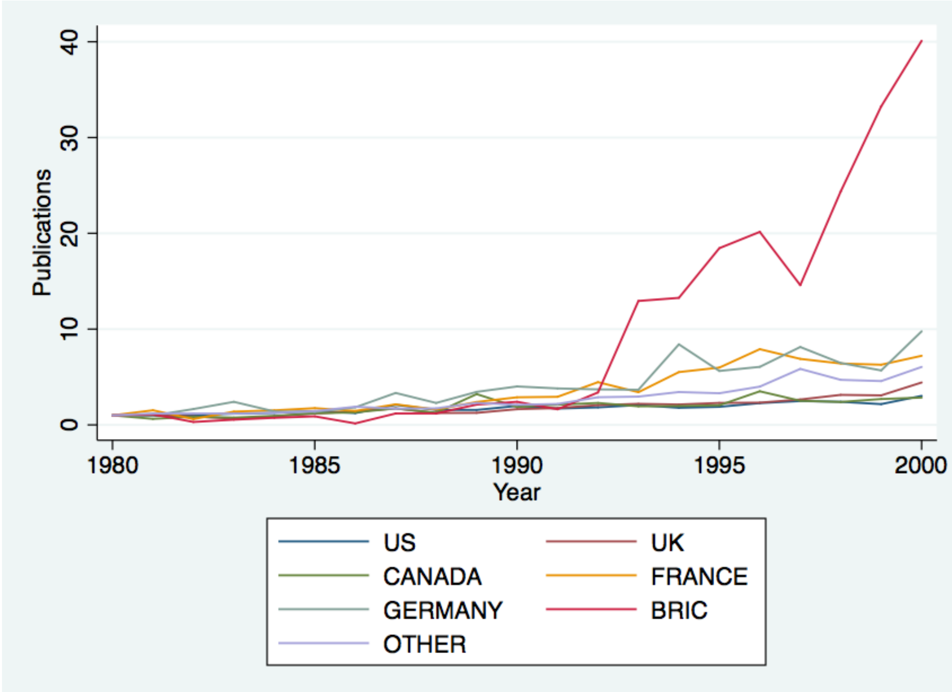

Figure 6: Gini Coefficients for the Distribution of Citation-Weighted Publications across Individuals

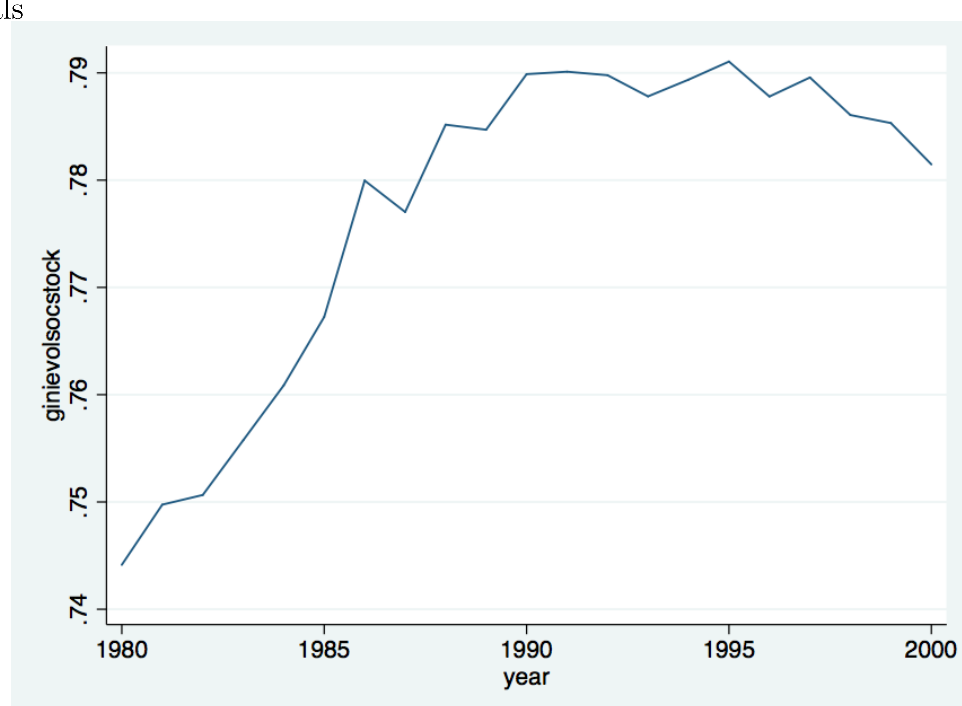

Figure 7: Lorenz Curves by Individual for Citation-Weighted Publications

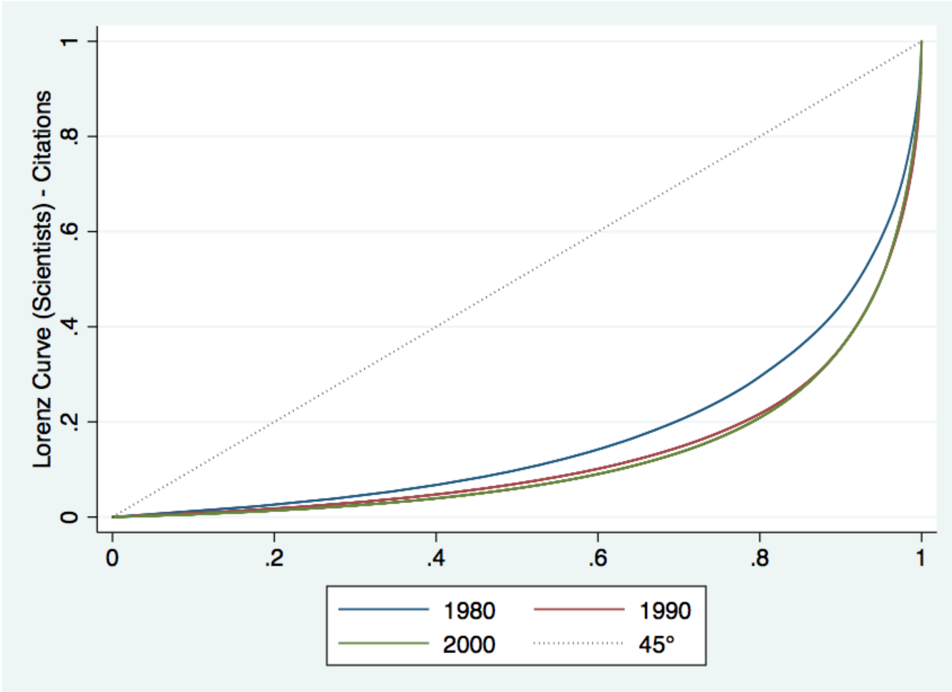

Figure 8: Publication Stock of 50th Ranked Scientist

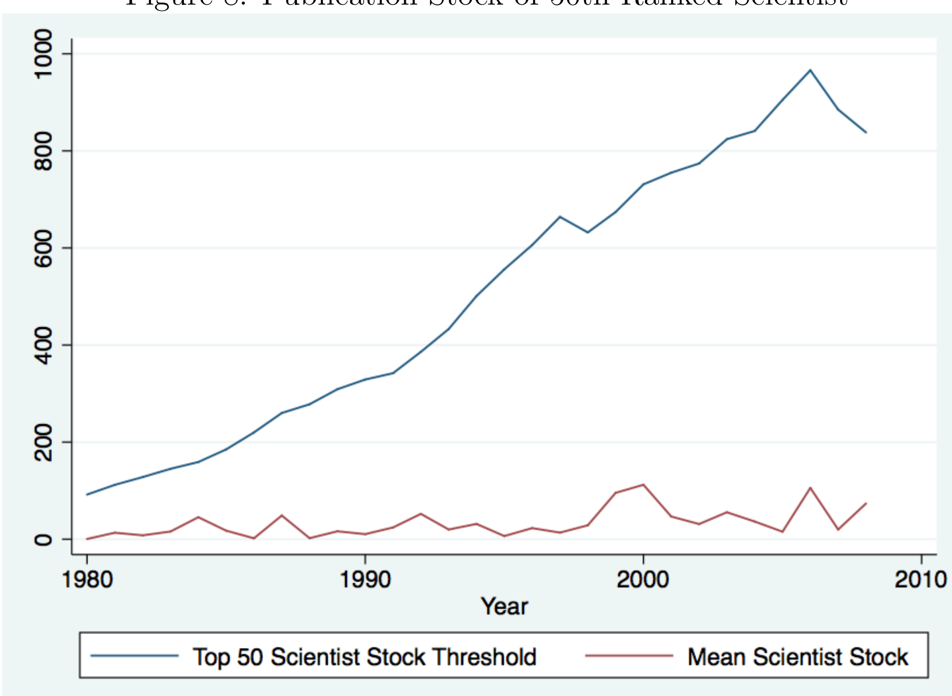

Figure 9: Mean Number of Authors Per Paper

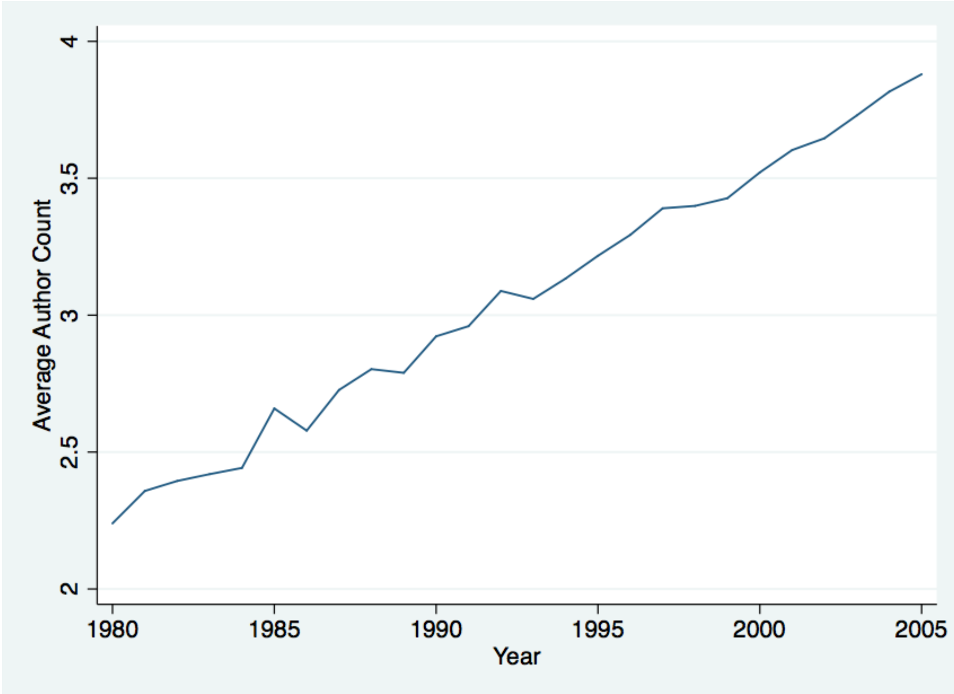

Figure 10: Mean Difference in Institution Rank Between Coauthors

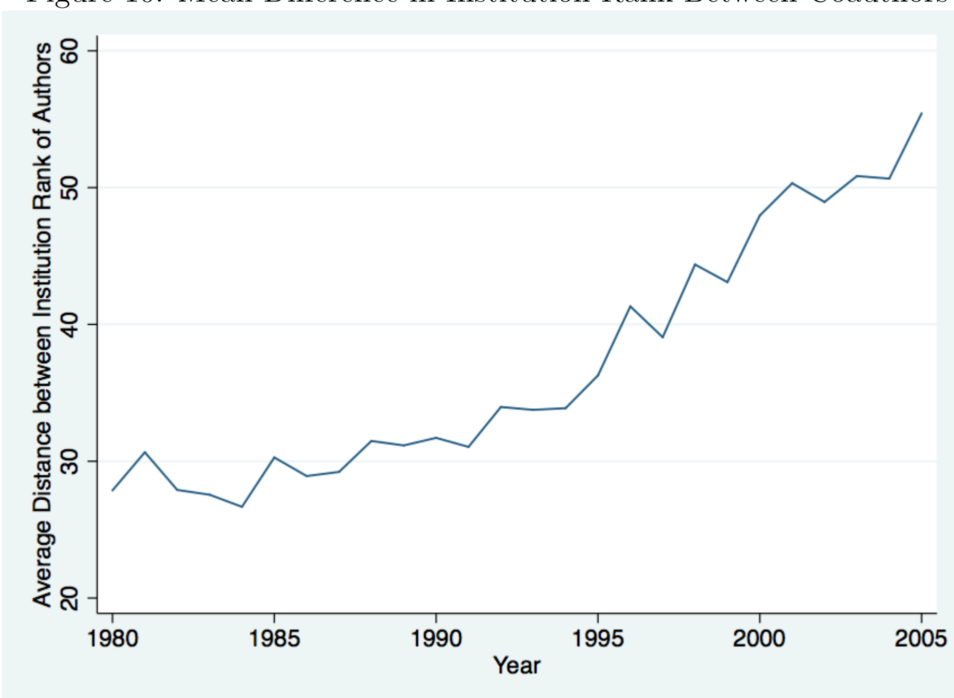

Figure 11: Mean Difference in Institution Rank Between Coauthors (Star versus Non-Star)

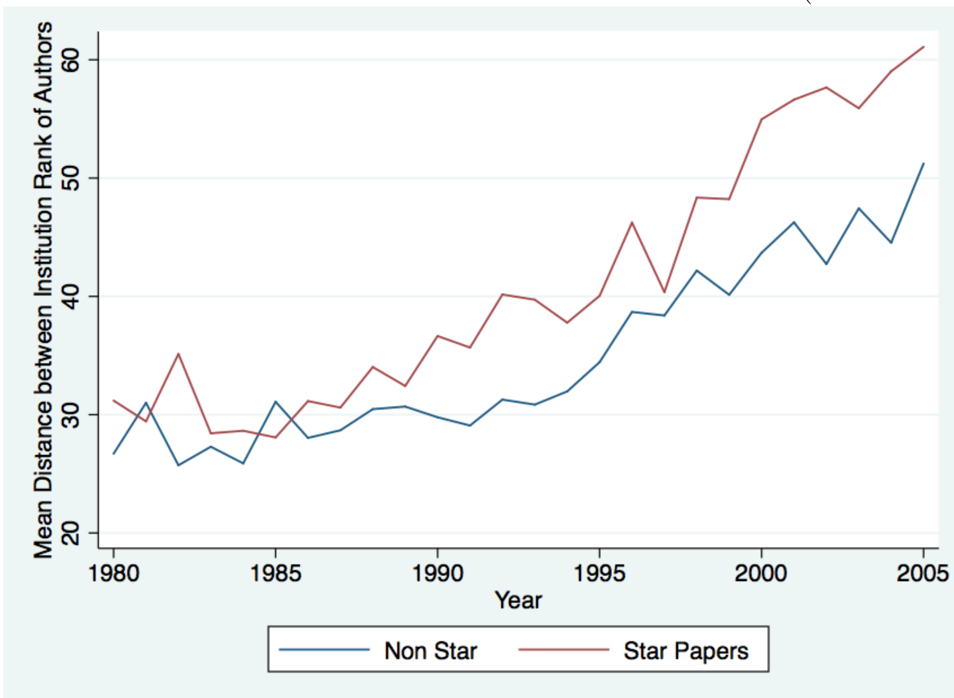

Figure 12: Mean Distance Between Coauthors (miles)

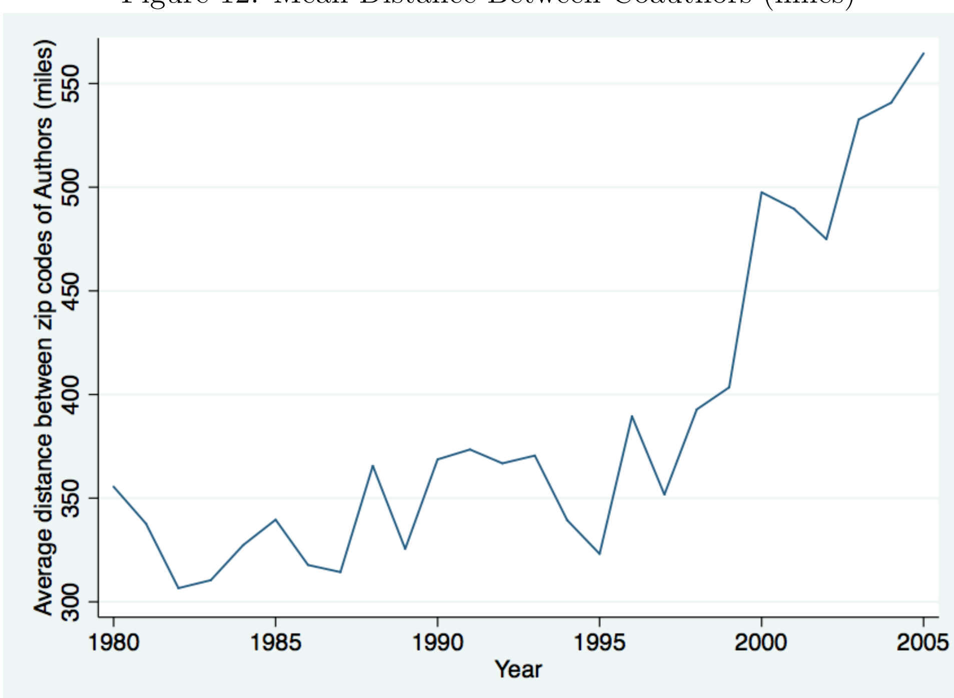

Figure 13: Department Level Rank in Evolutionary Biology: 1980 vs 2000

Figure 14: Fraction of Top 100 Ranked Researchers at a Top 50 Ranked Department