Submitted to *Operations Research*
manuscript OPRE-2014-06-377-R2

Tenure Analytics: Models for Predicting Research Impact

Dimitris Bertsimas
Sloan School and Operations Research Center, Massachusetts Institute of Technology, dbertsim@mit.edu

Erik Brynjolfsson
Sloan School, Initiative on the Digital Economy, and Operations Research Center, Massachusetts Institute of Technology, erikb@mit.edu

Shachar Reichman
Recanati Business School, Tel Aviv University and Sloan School, Massachusetts Institute of Technology, shachar@mit.edu

John Silberholz
Operations Research Center, Massachusetts Institute of Technology, josilber@mit.edu

Tenure decisions, key decisions in academic institutions, are primarily based on subjective assessments of candidates. Using a large-scale bibliometric database containing 198,310 papers published 1975–2012 in the field of operations research (OR), we propose prediction models on whether a scholar would perform well on a number of future success metrics using statistical models trained with data from the scholar's first five years of publication, a subset of the information available to tenure committees. These models, which use network centrality of the citation network, co-authorship network, and a dual network combining the two, significantly outperform simple predictive models based on citation counts alone. Using a dataset of the 54 scholars who obtained a PhD after 1995 and held an assistant professorship at a top-10 OR program in 2003 or earlier, these statistical models, using data up to five years after the scholar became an assistant professor and constrained to tenure the same number of candidates as tenure committees did, made a different decision than the tenure committees for 16 (30%) of the candidates. This resulted in a set of scholars with significantly better future A-journal paper counts, citation counts, and h-indexes than the scholars actually selected by tenure committees. These results show that analytics can complement the tenure decision-making process in academia and improve the prediction of academic impact.

Key words: citation analysis; academic impact; analytics; networks

1. Introduction

In academia, some of the most important decisions facing personnel and funding committees concern young researchers. Personnel committee members typically must decide whether to grant tenure based on evidence from less than a decade of research output following graduation with a doctorate, while funding committees must decide whether to provide crucial early-career grants to scientists based on a few years of research. Typically, the decision process is based on subjective assessments of the committee regarding the quality of a candidate's research and support letters, and the use of quantitative methods in this process is typically limited.

The impact of these decisions is not solely limited to scholars' careers, but also influences the ranking of departments, the prestige of universities, and the functioning of the scientific enterprise. The financial and organizational implications of these early-career academic decisions are large. A tenured faculty member will receive millions of dollars in career compensation and will occupy a faculty spot for decades. Meanwhile, the National Science Foundation provided \$5.8 billion in research funding in 2014 (National Science Foundation 2014b), including \$220 million specifically for young researchers (National Science Foundation 2014a). Given the stakes, we feel it is time for a "Moneyball moment" in academia, in which models predicting future academic outcomes are used to support decisions regarding early-career faculty.

Considering the importance of academic decisions, it is not surprising that the measurement of scholars' impact has received extensive attention in the literature. Most notably, Hirsch (2005) presented the h-index, where a scientist has index h if h of her N papers have at least h citations each, and the other $N - h$ papers have no more than h citations each. Several papers have offered extensions, modification and alternatives to the h-index (see Bornmann et al. (2008) for a comparison of nine different variants of the h-index). Podsakoff et al. (2008) produced a ranking of scholars in the field of management based on the total number of citations per author taking into account the attributes of the researcher's academic career (years in the field, graduate school attended, editorial board memberships, etc.).

Although early-career prediction of a researcher's future academic success is of particular interest to personnel and funding committees, it has received limited attention in the literature. Most studies that predict a researcher's future citations have relied on data from later in a researcher's career, often requiring a decade or more of observation (Dorsey et al. 2006, Hirsch 2007, Hönekopp and Khan 2012, Mazloumian 2012). Similarly, research to predict recipients of prestigious research awards has relied on data from late in a researcher's career. Garfield and Welljams-Dorof (1992) found that high ranking of an author by number of citations in a specific field is positively correlated with receiving Nobel Prizes. In Acuna et al. (2012), the authors present a model that attains high-quality predictions of future academic results for young life scientists. Follow-up research in a population of physicists showed that the model's performance deteriorates considerably on scholars very early in their careers (Penner et al. 2013); authors of these studies discuss the strengths and weaknesses of models to predict scientists' future impact in Acuna et al. (2013). In Yang et al.

(2011), the authors predict research productivity of urology researchers 2-4 years after residency given their publication history during residency, a time frame not representative of their medium- or long-term academic success. Importantly, no studies to date have addressed whether early-career predictions of scholars' future academic success can be used to improve the decisions made by tenure committees today.

In this work, we study how the network centrality of papers in the citation network, authors in the co-authorship network, and both papers and authors in a dual network combining the two, can be integrated into a future impact prediction algorithm. The idea to include network indexes into prediction methods stems from the fact that a citation represents a flow of information, and a research idea presented in one paper is built upon in another paper. Recent literature relates network structure properties to information dissemination in networks (Valente 1996, Katona et al. 2011). In the co-authorship network, centrality of an author may indicate better access to new information, better opportunities for new collaborations, and multidisciplinary research interests (Newman 2004). Structural importance of an author (e.g., higher centrality) may indicate that the author connects structural holes — sub-networks that are not otherwise connected (Burt 1995) — and is a broker of information. Additionally, people with high centrality have been found to have a competitive advantage over their peers and are more likely to be recognized as top performers (Burt 1995, 2005). Moreover, Goldenberg et al. (2012) found that in a content dual network structure, nodes with higher centrality bridge these structural holes and facilitate content exploration.

Our contributions include:

1. Using a bibliometric database of papers in the Operations Research (OR) literature, we predict if a scholar will perform well on a number of success metrics using statistical models trained only on publications within the scholar's first five years of publication.

2. We evaluate whether these statistical models could be used to improve the future publication metrics of scholars tenured at highly ranked OR programs. Using a dataset of the 54 scholars who obtained a PhD after 1995 and held an assistant professorship at a top-10 OR program in 2003 or earlier, these statistical models, using data up to five years after the scholar became an assistant professor and constrained to tenure the same number of candidates as tenure committees did, made a different decision than the tenure committees for 16 (30%) of the candidates. This resulted in a set of scholars with significantly better future A-journal paper counts, citation counts, and h-indexes than the scholars actually selected by tenure committees.

Our paper is structured as follows. In Section 2, we describe our data sources and the measures we use to perform network analysis. In Section 3, we present models to predict a scholar's future success using only early-career data, and in Section 4 we analyze if these models could be used to improve the future publication metrics of scholars tenured at top OR programs. Finally, we discuss the implications and limitations of this work in Section 5.

2. Data and Measures

Before describing and evaluating prediction models for future academic success, we describe our data sources for this study and the measures we use to perform network analysis and to evaluate the success of scholars.

2.1. Bibliometric Database

We collected data from the Thompson Reuters Web of Science (WOS) on all papers in journals or conference proceedings labeled as part of OR. In total, we obtained records for 198,310 papers published between 1975 and 2012. We additionally collected 398,871 papers in the WOS that are not in the OR field but that cite one of the OR papers and 400,850 papers in the WOS that are not in the OR field but that are cited by one of the OR papers. Due to overlaps between the non-OR papers citing and cited by OR papers, we obtained a final dataset of 752,562 papers.

In addition to information about papers, we obtained records of 1,489,509 citations to an OR paper and 1,293,378 references from the OR papers. Due to OR papers referencing other OR papers, we obtained a total of 2,206,116 citations.

2.2. Name Disambiguation

Bibliometric entries from the WOS provide either the first initial and last name or the first and last name for the authors of each paper, leaving ambiguity as to whether two authors sharing the same first initial and last name are in fact the same person. To address this issue, we performed name disambiguation, which associates each author on each paper with an *author cluster* that represents a single person. Details of our name disambiguation approach are provided in Appendix A.

Across the 198,310 WOS papers and 427,127 author/paper pairs in the OR field, we identified 136,313 author clusters using this approach. In Section 3, in which we predict future success of authors, we limit analysis to the 43,047 authors whose first publication was dated 1995 or earlier.

2.3. Network Analysis

After performing name disambiguation, we built three networks:

- We built the *citation network*, in which nodes represent papers and directed edges (p_1, p_2) represent paper p_1 citing paper p_2. The full citation network contains 752,562 nodes representing all of the papers in WOS in the field of OR, citing a paper in OR, or cited by a paper in OR, as well as 2,206,116 directed edges representing all citations to or from any of the 198,310 OR papers.
- We built the *co-authorship network*, in which nodes represent authors and undirected edges (a_1, a_2) represent authors a_1 and a_2 coauthoring at least one paper. The full co-authorship network contains 136,313 nodes representing authors of any of the 198,310 OR papers, as well as 290,301 undirected edges.

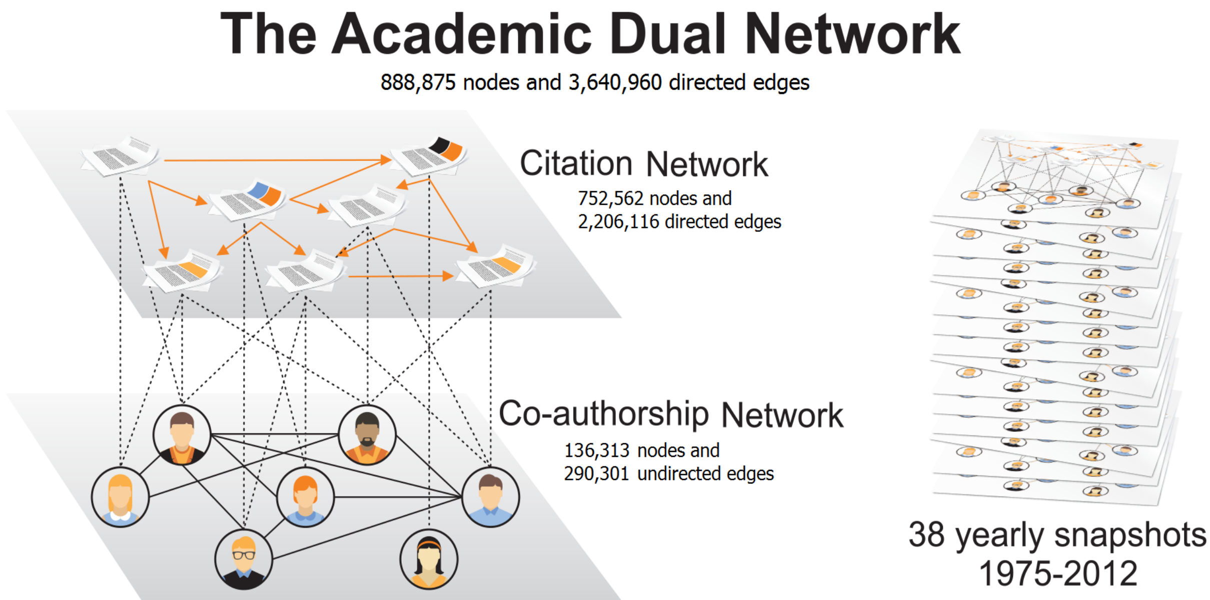

Figure 1 The full set of networks used in the network analysis. From 1975–2012, we compute snapshots of the *citation network*, in which nodes represent papers and directed edges represent citations, the *co-authorship network*, in which nodes represent authors and undirected edges represent co-authorship of at least one paper, and the *academic dual network*, which combines the citation and co-authorship networks and includes undirected edges representing authorship of a paper.

- We built the *academic dual network*, which consists of the union of the citation and co-authorship networks, in addition to undirected edges (a,p), which indicate that author a wrote paper p. To combine networks with directed and undirected edges, we replaced each undirected edge (a,b) with a pair of directed edges (a,b) and (b,a). With the addition of the 427,121 undirected edges for which an author wrote one of the WOS OR papers, the full academic dual network contains 888,875 nodes and 3,640,960 directed edges.

Because the citation network, co-authorship network, and academic dual network evolve over time, it is important to examine how the role of papers and scholars in the flow of knowledge may have also changed over time. We therefore created a set of yearly snapshots of the citation network from 1975 to 2012, where the papers that were published each year are added to the network of the former year, along with edges representing the references of these new papers. Networks snapshots of the co-authorship and dual networks are generated in a similar way, including any collaboration or authorship instance up to the snapshot's year. Figure 1 depicts the full set of networks used in the analysis.

For each network snapshot, we computed four centrality indices that are commonly used in the literature to characterize network structures and effectiveness (Barabási 2012, Newman 2003, Wasserman and Faust 1994):

- The *betweenness centrality* of a node is a measure of the number of the shortest paths between any two nodes in the network in which this node is included (Freeman 1977). Formally, $BC(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$, where σ_{st} is the number of shortest paths from node s to node t, $\sigma_{st}(v)$ is the number of shortest paths from s to t that pass through node v, and V is the set of all nodes. We defined the *normalized betweenness centrality* of node v as the betweenness centrality of v divided by the maximum betweenness centrality of any node in the network: $nBC(v) = \frac{BC(v)}{\max_{s \in V} BC(s)}$.
- The *closeness centrality* of a node is the inverse of the average shortest path between this node and any other node in the network (Freeman 1979). Formally, $CL(v) = \frac{|V|-1}{\sum_{s \neq v \in V} d_{vs}}$, where d_{vs} is the length of the shortest path from node v to node s.
- The *clustering coefficient* of a node quantifies how close a node's neighbors are to forming a clique, meaning neighbors of the node are also neighbors of each other (Watts and Strogatz 1998). Formally, $CC(v) = \frac{2L_v}{k_v(k_v-1)}$, where L_v is the number of edges between the k_v neighbors of node v.
- The *PageRank* of a node measures the node's relative importance in the network (Brin and Page 1998). Formally, $PR(v) = \frac{1-d}{|V|} + d \sum_{s \in B_v} \frac{PR(s)}{L(s)}$, where B_v is the set of all nodes linking to node v, $L(s)$ is the number of outgoing links of node s, and d is a damping factor that we set to 0.85, the suggested value from Brin and Page (1998).

2.4. Dataset for Tenure Impact Analysis

To evaluate whether prediction models for future academic success could be useful to tenure committees, we manually built a dataset of OR scholars who obtained a PhD since 1996. To obtain a more homogenous set of scholars, we limited our analysis to scholars who held an assistant professorship at a top-10 university for OR, as determined by the number of INFORMS fellows at the university — the set of universities used in this analysis were Carnegie Mellon, Columbia, Cornell, Georgia Tech, MIT, Michigan, Princeton, Stanford, UC Berkeley, and UCLA.

We first identified scholars who at some point had a position in one of the target universities. We assume that any OR professor at a top university will have attended the INFORMS annual conference some year between 1996 and 2011, and we scraped a set of 41,103 presentation records from this time period from the informs.org website. We obtained a more limited set of 15,178 records by filtering presentation records to ones potentially containing the name of one of our target universities. We manually reviewed this set of presentation records, obtaining a set of 685 scholars who had at some point held a tenure-track position at one of our target universities.

For each of the 685 scholars, we searched publicly available information on the Internet to obtain the year range of each tenure-track position that scholar has held, as well as the year they obtained their PhD. While we were often able to use the education and employment history sections of a scholar's CV or website, we also used affiliation information from INFORMS conference presentations, employment histories on LinkedIn profiles, and previous versions of departmental websites available through archive.org. We removed 370 scholars from our dataset because they had obtained their PhD before 1996, 99 scholars because they are currently assistant professors, eight scholars because they switched to a target university after receiving tenure elsewhere, and five scholars because we were unable to adequately determine their employment history. In this way, we obtained a more limited set of 203 scholars.

OR is an interdisciplinary field, so the INFORMS annual conference is attended by members of other academic communities. The 203 scholars in the limited dataset included members of communities as varied as computer science, public health, and chemical engineering. Because the citation database described in Section 2.1 is limited to OR publications, some of these scholars' publication records are not well represented by our dataset, so we removed any scholars with fewer than half of their journal publications present in our dataset. To obtain the list of all publications for each scholar, we primarily used publication lists available in CVs and research websites, but in cases where these were not available we used datasets like Google Scholar and, where appropriate, dblp. To obtain the publications of each scholar in our publication dataset, we manually linked scholars' publications to papers in our dataset; we did not use the automated name disambiguation described in Section 2.2. The requirement that half of a scholar's journal articles be present in our dataset limited our analysis to 75 scholars. In Section 4, when we evaluate if prediction models of future scholar success can we used to improve tenure decisions, we limit analysis to the 61 scholars who became an assistant professor in 2003 or earlier. This set of 61 scholars was constructed before performing any experiments to compare model-driven tenure decisions against the decisions of tenure committees.

Finally, we labeled the 61 scholars in our limited dataset based on whether they had received tenure at their first top-10 institution. All 35 scholars who obtained tenure at their first top-10 institution were labeled as having received tenure. For the remaining 26 scholars who left their first top-10 institution without attaining a tenured position, we determined if they either left because tenure was not granted/was not going to be granted or for personal/other reasons. For all scholars about whom team members did not have knowledge of this determination, we performed an interview with a senior faculty member at the scholar's institution using e-mail to determine whether the scholar left due to tenure-related reasons. IRB approval was obtained from both MIT and Tel Aviv University to perform these interviews, and informed consent was obtained from all interviewees. Seven scholars were excluded from the analysis due to having left their first position for personal reasons, and the remaining 19 scholars were labeled as having been denied tenure. One of these scholars did not take another academic position after being denied tenure and the remaining 18 took positions at universities outside our list of top-10 institutions.

Category	**Prediction of Author Outcomes (Section 3)**	**Analysis of Tenure Decisions (Section 4)**
Short-term publication metrics	Nine years after first publishing: • Paper count • A-journal paper count • h-index • Citation count	Nine years after becoming an asst. prof.: • Paper count • A-journal paper count • h-index • Citation count
Medium-term publication metrics	Sixteen years after first publishing: • Paper count • A-journal paper count • h-index • Citation count	Projected value for 16 years after becoming an assistant professor: • Paper count • A-journal paper count • h-index • Citation count
Research	INFORMS fellow (career award)	INFORMS research award within 10 years of becoming an assistant professor
Teaching	—	Teaching award within 10 years of becoming an assistant professor
Service	—	A-journal editor in March 2014
Demographics	—	Gender
Publication details	—	Mean number of coauthors on papers Author subfield(s)

Table 1 Metrics used to evaluate models

2.5. Scholar Metrics

In Section 3, we predict the future success of the 43,215 scholars in our dataset whose first publication was in 1995 or earlier; due to the large number of scholars in this analysis we were limited mainly to metrics that could be computed from our bibliometric database. In Section 4 we evaluate the future success of 54 scholars who had assistant professorships in top OR programs; given the small size of this dataset we computed a wider range of success metrics. However, scholars in this analysis became assistant professors as late as 2003, so we are limited to metrics that can be measured within the first decade of a scholar's career. The metrics selected for these two analyses are summarized in Table 1.

We measure short- and medium-term success in publication using a scholar's paper count, number of A-journal publications (defined as publications in *Management Science*, *Mathematical Programming*, *Mathematics of Operations Research*, or *Operations Research*), h-index, and citation count nine and 16 years after either first publishing (Section 3) or becoming an assistant professor (Section 4). In the tenure analysis, most scholars obtained an assistant professorship too recently to observe 16-year publication metrics. As a result, we instead use hierarchical linear models to obtain projected 16-year metrics for each scholar. Details of these projections are provided in Appendix B. Table 2 reports the breakdown of these publication metrics among the 54 scholars used in the tenure analysis.

	5-year			**9-year**			**Proj. 16-year**		
	Q1	Med.	Q3	Q1	Med.	Q3	Q1	Med.	Q3
Tenured ($n = 35$)									
Paper count	4	5	8	10	12	16	15	18	28
A-journal paper count	1	3	4	3	5	8	4	8	12
Citation count	10	14	33	49	96	145	157	228	389
h-index	1	2	3	4	5	6	5	7	9
Not tenured ($n = 19$)									
Paper count	1	2	5	5	7	12	8	11	18
A-journal paper count	1	1	2	3	4	5	4	6	7
Citation count	0	9	14	23	63	93	44	145	240
h-index	0	1	2	3	4	5	4	6	7

Table 2 First quartile (Q1), median (Med.), and third quartile (Q3) of publication metrics among the 54 scholars in the tenure dataset, broken down by whether they were tenured at their first top-10 university.

In Section 3, we also predict whether a scholar has become an INFORMS fellow, an award that is given for outstanding lifetime achievement in OR and management science. We collected data on scholars who were announced as recipients of this award from 2002, the first year it was offered, to 2013. We omitted scholars for whom we had no publication data before 1996 or when the author's name was ambiguous, leaving 193 recipients of the award. This metric was not appropriate for use in Section 4, because none of the scholars considered in that analysis are senior enough to have won this career award.

In Section 4, we evaluate tenure decisions using a number of other characteristics of scholars that are of interest to tenure committees but that need to be collected manually for each scholar and are therefore not practical to use in Section 3. Using the INFORMS Award Recipients website (INFORMS 2014), we evaluated whether a scholar had won a research-related award either from INFORMS or from one of the INFORMS sections or societies within 10 years of becoming an assistant professor. Of the 54 scholars, 17 had received a research award in this time frame. To assess superlative teaching, we reviewed scholars' CVs and websites to determine whether they had won a teaching award either from their university or from INFORMS within 10 years of becoming an assistant professor. Of the 40 scholars who listed awards and honors, 24 had received a teaching award in this time frame. To assess service to the community, we identified whether each scholar was an editor in chief, area editor, or associate editor at an A journal in March 2014. Of the 54 scholars, 29 were identified as editors. Though this metric does not control for the year when a scholar became an assistant professor, being an editor was not significantly associated with first year of assistant professorship in bivariate logistic regression analysis ($p = 0.44$). To identify the gender of scholars, we used data from CVs or author bios where available and otherwise used a scholar's photograph on their research website or first name. Of the 54 scholars, 43 are male (80%).

As part of the evaluation in Section 4, we also collected additional publication details, labeling all 54 of the scholars with the mean number of coauthors on their papers and a subfield classification.

Methods	**Contextual and Crosscutting Areas**	
• Decision Analysis	• Accounting	• Judgment and Decision Making
• Optimization	• Behavioral Economics	• Marketing
• Simuation	• Business Strategy	• Military and Homeland Security
• Stochastic Models	• Computational Economics	• Operations and Supply Chains
	• Entrepreneurship/Innovation	• Organizations
	• Env., Energy, and Sustainability	• OR Practice
	• Finance	• Policy Mod./Public Sector OR
	• Games, Info., and Networks	• Transportation
	• Information Systems	

Table 3 Set of subfield labels assigned to scholars

A scholar was labeled with a subfield s if at least half of the scholar's papers from the database in Section 2.1 fell in s. To make this determination, we manually reviewed the title and abstract of all 855 papers for these 54 scholars. Possible subfield labels were any editorial area from the journals *Operations Research* or *Management Science*, with 21 subfields in total (see Table 3). Because it is often difficult to determine the specific methodologies used from the abstract of an applied OR paper, we only labeled a paper with a methodological subfield (Decision Analysis, Optimization, Simulation, or Stochastic Models) if no other subfield label could apply. For instance, a paper about using optimization for revenue management would be labeled Operations and Supply Chains, while a paper about column generation strategies would be labeled Optimization. Papers could be labeled with multiple subfields. Forty-eight scholars were labeled with a single subfield, with the most common being Operations and Supply Chains (26 scholars) and Optimization (seven scholars). Six scholars were labeled with no subfield and two were labeled with two subfields.

As part of a sensitivity analysis performed in Section 4, we collected additional publication information about the 54 scholars. As detailed in Section 2.4, we reviewed the publication lists of all scholars and identified any publications not in the bibliometric database from Section 2.1. We labeled each additional publication with its publication year and whether it is in a highly ranked journal according to field-specific journal rankings. To make this determination, we included the A+ journals from finance and economics (Currie and Pandher 2011, Ritzberger 2008), the A* journals from computer science (Excellence for Research in Australia 2010), the top four journals from marketing, information systems, and statistics (Hult et al. 2009, R. Kelly Rainer and Miller 2005, Theoharakis and Skordia 2003), and the interdisciplinary journals *Science* and *Nature*. In total, we identified 24 additional A-journal publications through 2012; 17 of these papers had been published within the first five years of assistant professorship.

3. Prediction of a Scholar's Future Success

For a model to be useful to a hiring or tenure committee, it must be able to accurately predict the future success of a scholar based on early-career data. In this section, we define statistical models to predict the set of nine metrics identified in Section 2.5 using only centrality measures available within five years of an author's first publication. To be able to observe long-term career outcomes, we limited our dataset to the 43,215 scholars identified in Section 2.2 whose first paper was published in or before 1995. We randomly split this set of scholars so that 70% were in the training set and 30% were in a testing set.

The eight publications metrics (paper count, A-journal paper count, h-index, and citation count nine and 16 years after first publication) are all continuous outcomes. For each publication metric m, we defined a baseline prediction model that predicted the metric using only citation count five years after first publication, C_5. To capture non-linear relationships between C_5 and each publication metric m, we trained random forest models (Breiman 2001) using the R `randomForest` package (Liaw and Wiener 2002), selecting parameters using 10-fold cross-validation with the R `caret` package (Kuhn 2015).

For each of the eight outcomes, we compared this baseline model against two other random forest models, each of which also had parameters selected via 10-fold cross-validation. The first model (four-metric model) used the four publication metrics measured five years after first publication: the number of citations of the author's papers (C_5), the h-index of the author (h_5), the number of publications by the author (P_5), and the number of A-journal publications by the author (A_5). The second model (centrality model) used these four metrics in addition to centrality measures for both the author in the co-authorship and dual networks and the author's papers in the citation and dual networks. We defined the following additional independent variables, each of which was measured five years after the first publication by the authors:

- The betweenness centrality of the author in the co-authorship network (BC_5^{cA}) and the dual network (BC_5^{dA}), and the (arithmetic) mean betweenness centrality of the author's papers in the citation network (BC_5^{cP}) and in the dual network (BC_5^{dP})
- The normalized betweenness centrality of the author in the co-authorship network (nBC_5^{cA}) and the dual network (nBC_5^{dA}), and the mean normalized betweenness centrality of the author's papers in the citation network (nBC_5^{cP}) and in the dual network (nBC_5^{dP})
- The closeness centrality of the author in the co-authorship network (CL_5^{cA}) and the dual network (CL_5^{dA}), and the mean closeness centrality of the author's papers in the citation network (CL_5^{cP}) and in the dual network (CL_5^{dP})
- The clustering coefficient of the author in the co-authorship network (CC_5^{cA}) and the dual network (CC_5^{dA}), and the mean clustering coefficient of the author's papers in the citation network (CC_5^{cP}) and in the dual network (CC_5^{dP})
- The PageRank of the author in the co-authorship network (PR_5^{cA}) and the dual network (PR_5^{dA}), and the mean PageRank of the author's papers in the citation network (PR_5^{cP}) and in the dual network (PR_5^{dP})

	Baseline		Four-Metric		Centrality	
Metric	**RMSE**	**MAPE**	**RMSE**	**MAPE**	**RMSE**	**MAPE**
9-year paper count	2.8	79	1.3	20	1.3	21
	(2.8, 2.8)	(78, 81)	(1.3, 1.3)	(20, 21)	(1.3, 1.3)	(20, 21)
16-year paper count	5.7	127	4.3	56	4.1	54
	(5.7, 5.8)	(124, 131)	(4.3, 4.5)	(54, 58)	(4.1, 4.2)	(53, 56)
9-year A-journal paper count	1.0	32	0.4	5	0.4	5
	(1.0, 1.0)	(31, 33)	(0.4, 0.4)	(5, 6)	(0.4, 0.4)	(5, 6)
16-year A-journal paper count	1.4	38	0.8	10	0.8	11
	(1.4, 1.5)	(37, 39)	(0.8, 0.8)	(10, 11)	(0.8, 0.8)	(10, 11)
9-year citation count	7.3	58	6.8	54	6.3	45
	(7.2, 7.6)	(58, 60)	(6.8, 7.2)	(53, 55)	(6.3, 6.7)	(45, 47)
16-year citation count	39.0	217	36.5	178	33.5	148
	(38.9, 40.3)	(207, 226)	(36.3, 38.5)	(169, 184)	(33.3, 34.8)	(146, 159)
9-year h-index	0.6	31	0.4	22	0.4	18
	(0.6, 0.6)	(30, 31)	(0.4, 0.4)	(22, 22)	(0.4, 0.4)	(18, 19)
16-year h-index	1.3	53	1.0	38	0.9	34
	(1.3, 1.3)	(53, 54)	(1.0, 1.0)	(38, 39)	(0.9, 0.9)	(34, 35)

Table 4 Testing-set root-mean-square error (RMSE) and mean absolute percentage error (MAPE), with bootstrap 95% confidence intervals. Models compared are the baseline prediction model, which uses the total number of author's papers' citations, the four-metric model which uses citation counts, h-index, publication counts, A-journal publication counts, and the network centrality model, which uses citation counts, h-index, publication counts, A-journal publication counts, and centrality measures in the co-authorship, citation, and dual networks from the same time frame. All three models use data from the five years following an author's first publication.

Table 4 displays the testing-set root-mean-square errors (RMSE) and mean average percentage errors (MAPE) of the three models for each of the eight publication metrics, with bootstrap percentile 95% confidence intervals (Davison and Hinkley 1997) computed using the R `boot` package (Canty and Ripley 2014). To accommodate outcomes with value 0 in the computation of MAPE, we used modified formula $MAPE(\hat{y}, y) = \sum_{i=1}^{n} 100 \frac{|\hat{y}_i - y_i|}{\max(y_i, 1)}$, where $\hat{y}$ is the vector of predicted outcomes, y is the vector of true outcomes, and n is the number of observations. The network centrality model obtained the largest improvements in MAPE over the baseline model on 9- and 16-year A-journal paper count and on 9- and 16-year paper count, decreasing the MAPE by more than half in all cases. The improvement over the baseline was more modest on the 9- and 16-year citation count and h-index metrics, but the network centrality models still improved the MAPE by more than 10% compared to the baseline model in all cases. The network centrality model had nearly identical performance to the four-metric model when predicting the 9- and 16-year paper count, A-journal paper count, and h-index outcomes, yielding RMSE improvements of 0.4 (95% CI 0.3–0.6) when predicting 9-year citation count and 3.0 (95% CI 2.4–4.5) when predicting 16-year citation count. For the 16-year citation outcome, adding the centrality measures to the four-metric model yielded a 29% improvement in MAPE (95% CI 19%–30%). 16-year citation count proved to be the hardest metric to predict, with the MAPE of all models exceeding 100%. As expected, 16-year outcomes were more difficult to predict than 9-year outcomes in all cases.

In addition to the publication metrics, we predicted the probability that each scholar would become an INFORMS fellow (p^f) using the same independent variables. Due to the small number of positive observations, we used logistic regression models instead of random forests for these models.

Our baseline model was:

$$\ln \frac{p^f}{1-p^f} = \beta_0 + \beta_{C_5} C_5 + \epsilon$$

Our four-metric model was:

$$\ln \frac{p^f}{1-p^f} = \beta_0 + \beta_{C_5} C_5 + \beta_{h_5} h_5 + \beta_{P_5} P_5 + \beta_{A_5} A_5 + \epsilon$$

Our network centrality model was:

$$\ln \frac{p^f}{1-p^f} = \beta_0 + \beta_{C_5} C_5 + \beta_{h_5} h_5 + \beta_{P_5} P_5 + \beta_{A_5} A_5 + \sum_{t \in \{cA, dA, cP, dP\}} \sum_{M \in \{BC, nBC, CL, CC, PR\}} \beta_{M_5^t} M_5^t + \epsilon$$

Figure 2 displays the receiver operator characteristic (ROC) curve for testing-set predictions of a scholar becoming an INFORMS fellow. Especially at high sensitivities, the network centrality model outperforms both the baseline and the four-metric models — the network centrality model could identify 90% of INFORMS fellows in the testing set with a false positive rate of 7%, improving over the false positive rate of 56% for the baseline model (bootstrap 95% CI for improvement 43%–50%) and 17% for the four-metric model (95% CI for improvement 4%–20%). As determined by the area under the ROC curve (AUC), the network centrality model can differentiate between a randomly selected future INFORMS fellow and non-INFORMS fellow 97% of the time using publication data from their first five years of publication, improving over the 82% performance of the baseline model (95% CI for improvement 13%–15%) and the 94% performance of the four-metric model (95% CI for improvement 1%–4%).

4. Evaluating Data-Driven Tenure Decisions

In Section 3, we compared models for predicting future research impact that use data from the first five years of a scholar's academic career, and we found that models trained with a variety of publication measures and network centrality measures outperform models trained only using citation information. However, it remains to be seen if these models can be useful to tenure committees, as committees have access to information not available to the models from Section 3, including

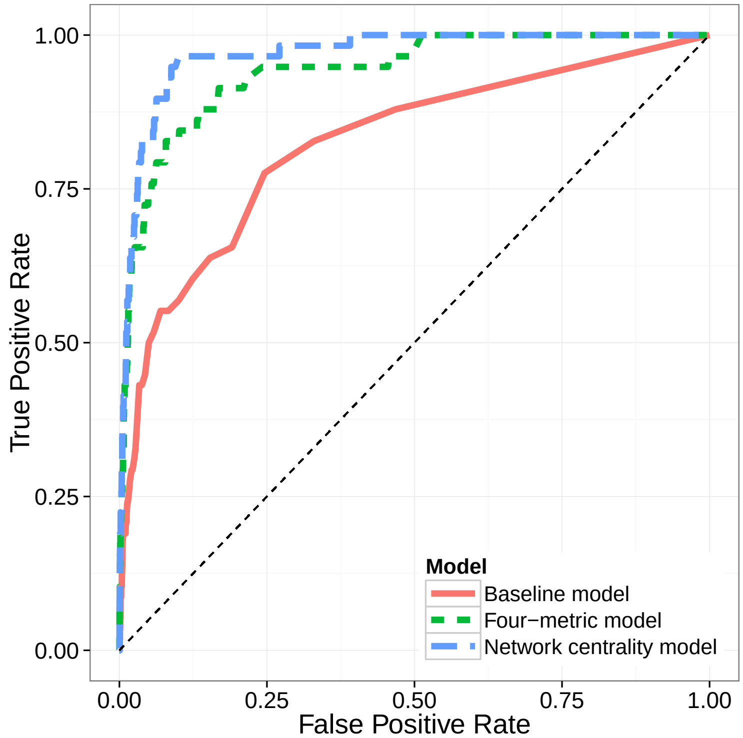

Figure 2 Testing-set receiver operator characteristic curve for predicting if a scholar will become an INFORMS fellow. Models evaluated are the baseline prediction model (AUC= 0.82), which uses citation totals in the five years following an author's first publication, the four-metric model (AUC= 0.94) which uses citation counts, h-index, publication counts, A-journal publication counts, and the network centrality model (AUC= 0.97), which uses citation counts, h-index, publication counts, A-journal publication counts, and centrality measures in the co-authorship, citation, and dual networks from the same time frame.

forthcoming papers, the text of published papers, teaching evaluations, and letters of support. To address this question, we built a dataset of the 54 scholars who obtained a PhD in 1996 or later and held an assistant professorship at a top-10 OR program in 2003 or earlier, as detailed in Section 2.4.

To compare the tenure decision making process currently being used by universities to the proposals made by the network centrality models from Section 3 for a set of scholars S, we first rank the scholars in S by their predicted value for each of the eight publication metrics used in Section 3, using publication information from five years after assistant professorship as the independent variables for each scholar. If t of the scholars in S were tenured at a top university, then we select the t scholars with the best average rank across the eight publication metrics as the "tenure selections" of the network centrality models. Among the 54 scholars in the dataset, 35 (65%) were tenured at a top-10 university. The network centrality models agreed with tenure

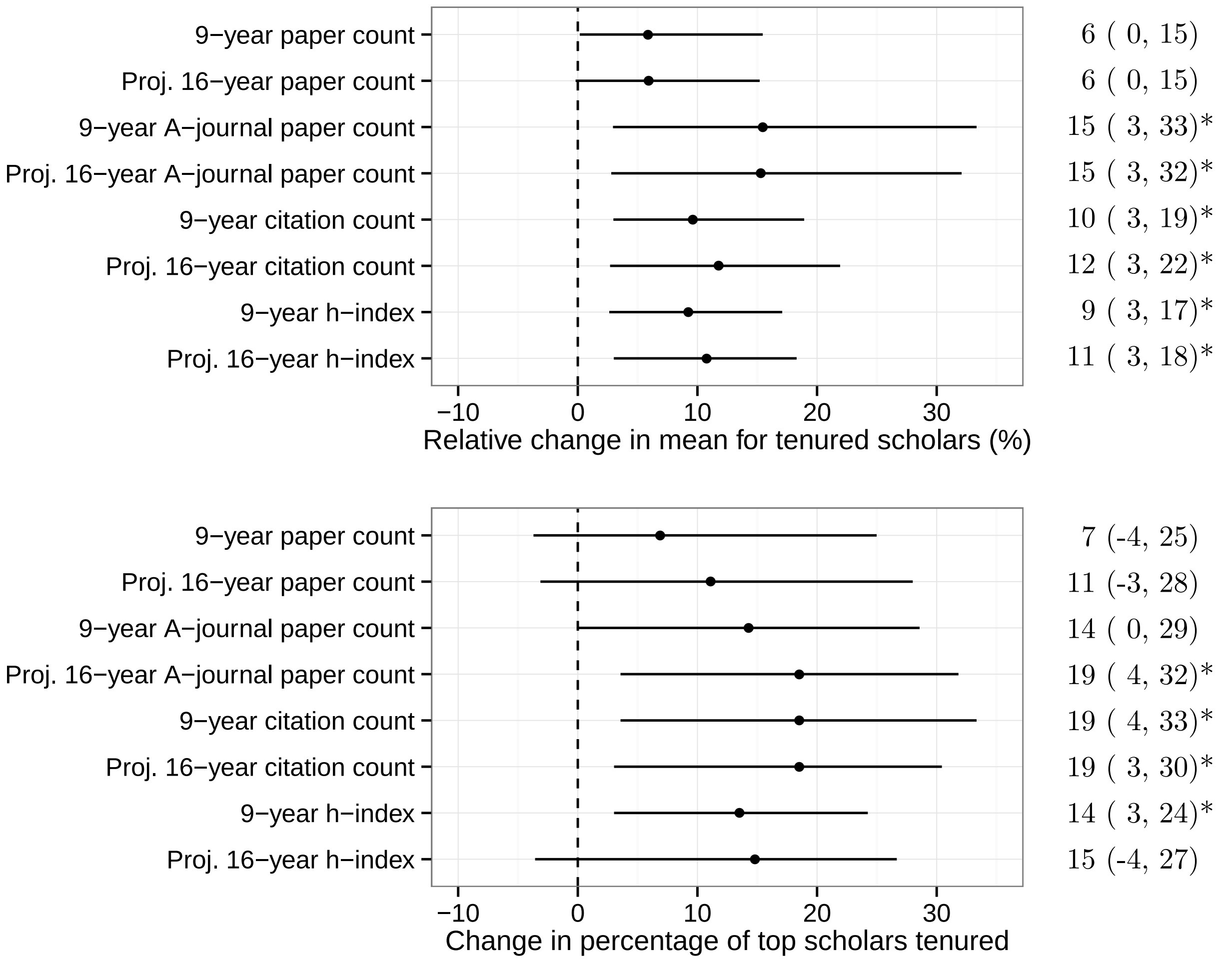

Figure 3 Difference (and bootstrap 95% CI) in future publication outcomes between the scholars selected for tenure by the network centrality models from Section 3 and the scholars selected by tenure committees. (Top) Relative change in mean value among tenured scholars for each metric. For instance, committees tenured scholars with average 16-year citation count 323 and our models would have tenured scholars with average 16-year citation count 361, a relative change of 12%. (Bottom) Change in the percentage of top-performing (above-median) scholars tenured for each metric. For instance, committees tenured 74% of top scholars for the 9-year citation count and our models would have tenured 93%, a change of 19%.

committees on 38 (70%) of the scholars, tenuring eight scholars not selected by the committees and not tenuring eight scholars selected by the committees.

Figure 3 compares the tenure decisions of the network centrality models against the decisions of tenure committees across the eight future publication metrics, displaying both the relative change in the mean value of the metric among tenured scholars (top) and the change in the proportion of above-median scholars given tenure (bottom). Bootstrap percentile confidence intervals (Davison and Hinkley 1997) were computed with the R boot package (Canty and Ripley 2014). The network centrality model obtained statistically significant improvements in mean values for the 9- and projected 16-year A-journal paper count, citation count, and h-index metrics. The models showed the largest relative change in mean A-journal paper count and the smallest relative change in paper count, and the models obtained an improvement of 1.4 papers (95% CI 0.0, 3.5), 1.3 A-journal papers (95% CI 0.2, 2.4), 38 citations (95% CI 9, 63), and 0.8 h-index (95% CI 0.2, 1.3) in projected 16-year outcomes over the scholars who were actually given tenure. The network centrality models increased the percentage of above-median scholars tenured by more than 10% in seven of the eight publication metrics, with four statistically significant improvements and one statistically non-inferior change.

For each type of publication metric, Figure 4 plots the 5-year and projected 16-year values for the 54 scholars, using color to indicate the tenure decision by committees and the network centrality models for that scholar. Nearly all scholars given tenure by a tenure committee but not the model (blue points in Figure 4) had below-median long-term metrics. Conversely, the majority of the scholars chosen by the model as replacements (purple points in Figure 4) had above-median long-term outcomes. The network centrality model did a better job of identifying scholars who are "diamonds in the rough," with below-median 5-year metric values and above-median projected 16-year values. Only three of the 29 blue points in Figure 4 with below-median 5-year outcomes had above-median 16-year outcomes, compared to six of the 10 purple points with below-median 5-year outcomes.

We also assess the performance of the other models from Section 3 in selecting scholars to tenure. If the four-metric models had been used instead of the network centrality models, exactly the same set of scholars would have been selected for tenure. On the other hand, consider a simpler baseline model, which selects scholars for tenure based on their 5-year citation count alone. To tenure 35 of the 54 scholars, the same number as tenure committees, the model tenures scholars with nine or more citations by year five. This simple model agrees with tenure committees on 38 (70%) of the tenure decisions. The simple model had similar performance to the network centrality models on the 9- and projected 16-year citation count and h-index metrics, but it performed worse on the metrics related to paper count. On the 9- and projected 16-year paper count and A-journal paper count metrics, the simple model had no statistically significant changes in mean value or proportion of top scholars tenured compared to tenure committees. Meanwhile, the network centrality model had three statistically significant improvements and three statistically non-inferior changes across these eight comparisons. While the citation-based approach is much simpler than the network centrality models and performs equally well on the citation count and h-index metrics, it has worse performance at identifying scholars who perform well on publication count metrics.

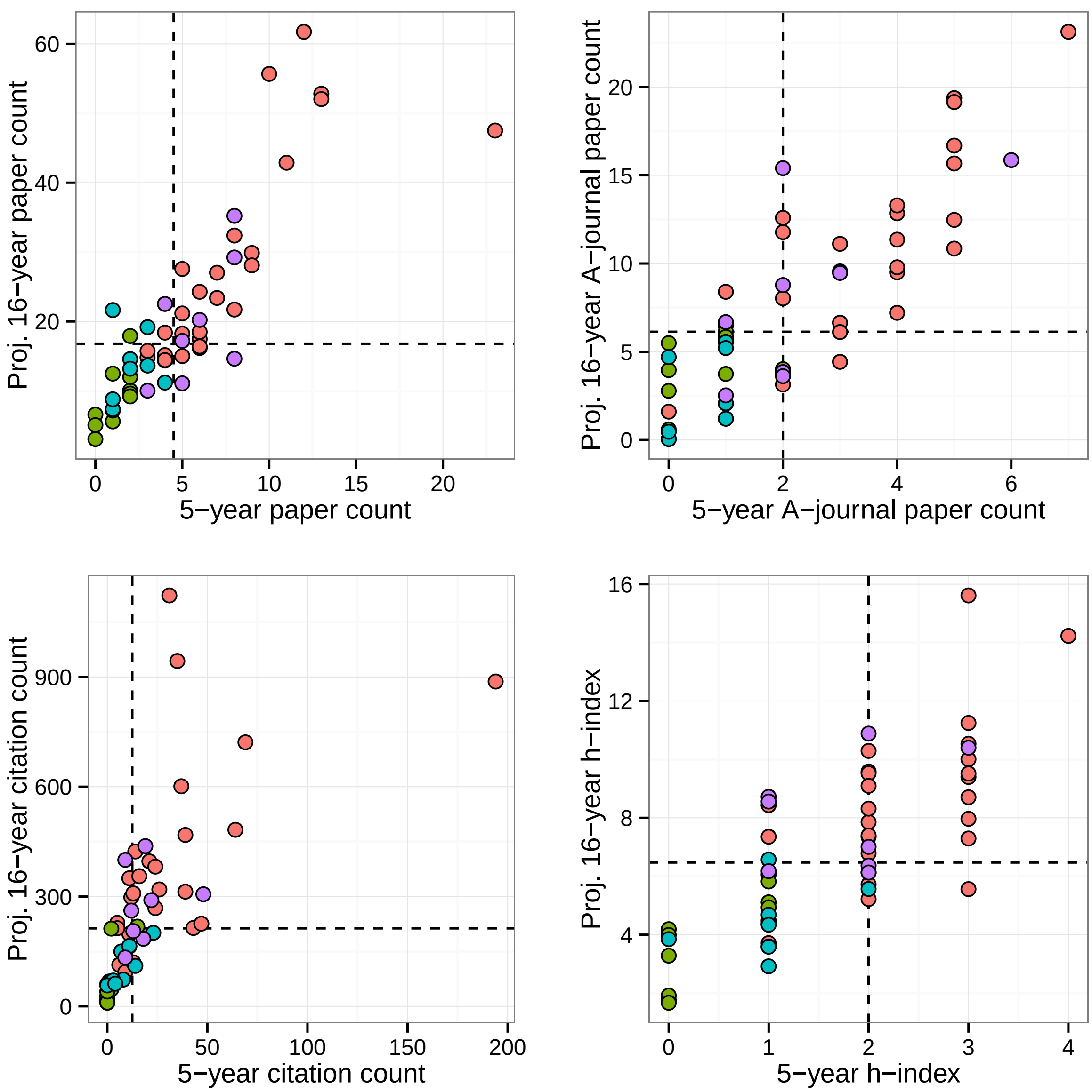

Figure 4 Comparison of 5-year and projected 16-year publication metrics for each scholar. The colors indicates the tenure decision by tenure committees and by the network centrality models — red indicates both gave tenure, green means neither gave tenure, blue means only the committee gave tenure, and purple means only the models gave tenure. The dashed lines indicate the median 5-year and projected 16-year outcomes.

The comparison of the network centrality model to tenure committees is robust to changes in the data source for future publication outcomes. As a sensitivity analysis, we augmented the 9- and 16-year A-journal publication counts with the 24 additional publications identified in Section 2.5. The performance of the proposed model compared to tenure committees was similar — the mean 9- and projected 16-year A-journal publication counts exhibited relative increases of 13% (95% CI

2, 29) and 13% (95% CI 2, 28), respectively, and the rate of above-median scholars given tenure for the 9- and projected 16-year A-journal publication counts increased by 13% (95% CI 0, 26) and 15% (95% CI 0, 28), respectively.

We found no statistically significant evidence that our approach would cause changes in the composition of scholars given tenure across the additional outcomes of interest identified in Section 2.5:

– Proportion of scholars receiving an INFORMS research award within 10 years of becoming an assistant professor (95% CI -0.21, 0.16).

– Proportion of scholars receiving a teaching award within 10 years of becoming an assistant professor (95% CI -0.25, 0.09). The unbalanced nature of this confidence interval likely stems from tenure committees' access to teaching evaluations.

– Proportion of scholars who were A-journal editors in March 2014 (95% CI -0.10, 0.19).

– Proportion of scholars who are male (95% CI -0.06, 0.14).

– Mean number of coauthors on publications (95% CI -0.17, 0.05).

– Proportion of scholars from each subfield (95% CI for each subfield in Table 3 contains 0).

5. Discussion and Conclusions

Using a bibliometric database of OR papers, we established that a scholar's publications early in their career can be used to predict later-career success and that these predictions could yield statistically significant improvements in the future publication metrics of scholars tenured by top OR programs. The latter result is especially noteworthy because the models developed in this paper did not have access to many of the sources of information available to tenure committees. This suggests that prediction models of future academic success could be useful to tenure committees.

It is important to note that tenure committees consider many criteria when making tenure decisions. While the models proposed in this work rank scholars based on predictions of various measures of future research productivity, they does not account for other important considerations for tenure, such as a scholar's service to their university, teaching ability, or personality. Some of these other criteria can be quantified, and in Section 4 we demonstrated that the scholars tenured by the proposed model do not statistically significantly differ from those selected by tenure committees in the rate of research awards, teaching awards, or A-journal editorships, nor do they significantly differ in the distribution of subfield, gender, or typical number of coauthors. However, other criteria, such as personality or creativity (Azoulay et al. 2011), are difficult to quantify, and tenure committees must rely on imprecise measures when evaluating candidates based on these factors. Criteria not related to research productivity can be important in the tenure decision — among the five pairs of scholars in our OR tenure dataset with identical 5-year research productivity values (paper count, A-journal paper count, citation count and h-index), one pair of scholars had different tenure outcomes (one was tenured and the other was not). Because the models presented in this work are limited to predictions of future research productivity and cannot evaluate candidates on all criteria of interest to tenure committees, they would be most useful as decision aids to complement the existing evaluation procedures used by tenure committees.

The analysis in Section 4 has several limitations. First, the total number of scholars in the analysis set is relatively small, making it difficult to obtain sharp estimates of the differences in long-term outcomes between the scholars tenured by their universities and the scholars selected for tenure by the models presented in this work. Further, the analysis evaluates the proposed model based on observed long-term outcomes for scholars, even in cases where the proposed model disagrees with the choice made by tenure committees. The initial tenure decision might in fact affect a scholar's long-term outcomes; for instance, failing to get tenure at a top-10 institution might decrease a scholar's research output as they work to adjust to a new university, or it might alternately provide motivation, yielding a boost in productivity. Finally, the analysis treats the number of tenure slots across the programs studied as a fixed resource, an assumption made to simplify the comparison of the proposed model's choices against those of tenure committees. In reality, no such limit exists.

The models described in this work could be expanded in a number of ways. First, the data sources in this work were limited in scope — we only considered publications and scholars from the field of OR, and we limited our study of the effectiveness of data-driven tenure decisions to top-ranked OR programs. While we also believe the proposed models could be useful in other fields and at lower-ranked programs, the only way to confirm the broader effectiveness of the proposed methodology is to test it in other settings. Further, we only considered models for the tenure decision. Similar models could be used in other contexts, such as hiring new assistant professors, evaluating candidates for grants and awards, and hiring scholars who previously held tenure-track positions at other institutions. Additional experimentation is needed to evaluate the usefulness of predictions of future research impact in making these decisions.

For the prediction models described in this work to be useful to tenure committees, they need to be implemented and separately calibrated for a broad range of academic disciplines using a large-scale bibliometric database. First, this implementation requires large-scale name disambiguation to be performed across the bibliometric database. Further, our network centrality models rely on network centrality measures computed over large citation and co-authorship networks. The networks considered in this paper consisted of fewer than 1 million nodes and 10 million edges, and we were able to compute all centrality measures required using a single personal computer. However, networks for other fields might be significantly larger, resulting in a larger computational burden to compute the centrality measures. Researchers have reported success in using parallel processing to speed up centrality computations in large networks (Bader and Madduri 2006), which could reduce this hurdle to implementing a decision support system. Given the significant effort and data required to implement the models presented in this work, the models would need to be developed and distributed as a complementary service to an existing bibliometric database like Google Scholar or the Thomson Reuters Web of Science. Alternately, tenure committees might favor the four-metric models, which do not rely on centrality and yielded similar predictive performance in Section 3 and identical performance at selecting scholars to tenure in Section 4. Models would need to be updated periodically, as patterns of publication change over time. If models relying on network centrality gain widespread use in the tenure decision-making process then candidates might change their publication behavior to boost their centrality in citation and co-authorship networks, prompting further recalibration of the proposed model.

Though broader evaluation is needed and hurdles remain to deliver the prediction models developed in this work to tenure committees across a range of academic disciplines, the demonstrated effectiveness of these models in the field of OR suggests potential for data-driven models as decision aids to academic personnel committees.

Appendix A: Name Disambiguation Algorithm

Bibliometric entries from the WOS provide either the first initial and last name or the first and last name for the authors of each paper, leaving ambiguity as to whether two authors sharing the same first initial and last name are in fact the same person. To address this issue, we performed name disambiguation, which associates each author on each paper with an *author cluster* that represents a single person. This process consists of two steps — first we predict the probability that two authors who share the same first initial and last name are in fact the same person, and then we use agglomerative clustering to assign each paper to an author cluster.

To predict the probability that two authors who share a first initial and last name are the same person, we collected records for 431,395 papers from the WOS database for the field of economics as well as 299,707 papers from the Social Sciences Research Network (SSRN), which is a popular database of preprints and working papers for a number of fields including economics. We considered papers between these two databases to be the same if they have the same title and same author last names, yielding 37,848 matches with a total of 72,657 author records. Among these 37,848 papers, there were 261,203 pairs of papers for which one author on each had the same first initial and last name. We labeled each pair as being the same person if each shared the same SSRN login id and a different person if they had a different login id; 214,577 pairs (82.1%) were labeled as matches.

For each pair of papers, we defined the following variables:

- x_1: Measure of how well middle names match — 3 if both are reported and matching, 2 if neither is reported, 1 if one is reported and one is not reported, and 0 if both are reported and non-matching

- x_2: Measure of how well first names match — 3 if both first names are fully reported and matching, 2 if both first names are abbreviated, 1 if one first name is fully reported and the other is abbreviated, and 0 if both first names are fully reported but don't match
- x_3: Measure of how well the emails match — 3 if both emails are fully reported and matching, 2 if neither email is reported, 1 if one email is reported and the other is not, and 0 if both emails are fully reported and not matching
- x_4: Whether the document type (proceedings paper, journal article, letter, or other) of the two articles is the same
- x_5: Cosine similarity between the titles of the two articles. The cosine similarity is the dot product of the word frequency vectors divided by the magnitude of each vector. The maximum value is 1, indicating an identical distribution of word frequencies, and the minimum value is 0, indicating no words in common
- x_6: Cosine similarity between the source names of the two articles
- x_7: Cosine similarity between the abstracts of the two articles
- x_8: Cosine similarity between names of the institutions of the two authors
- x_9: Cosine similarity between the author-provided keywords of the two articles
- x_{10}: Cosine similarity between the keywords of the two articles generated by WOS
- x_{11}: Cosine similarity between the sets of coauthor names for the two articles (represented by first initial/last name pairs)
- x_{12}: The minimum number of coauthors divided by the maximum number of coauthors between the two articles
- x_{13}: The difference in number of coauthors between the two articles
- x_{14}: The difference in number of citations through 2012 between the two articles
- x_{15}: The difference in publication year between the two articles
- x_{16}: The difference in number of pages between the two articles

Using a 70% random sample of the paper pairs, we trained a logistic regression model using variables x_1 through x_{16}. Variables x_1 through x_4 were modeled as factor variables and variables x_i, x_i^2, and x_i^3 were included in the model specification for $i \in \{5, \ldots, 16\}$. This model obtained a test-set AUC of 0.921, meaning it could differentiate between a randomly selected true positive and true negative pair 92.1% of the time.

Among the 198,310 papers classified in the OR field in the WOS dataset, there were 106,130 unique first initial/last name values. For each of these names n, there is a set of papers S_n containing authors with that first initial and last name. For each pair of papers i and j in a set S_n, the logistic regression model provides a predicted probability p_{ij} that the authors in this pair of papers with name n are the same person. The distribution of names is similar between OR and economics —

there is a cosine similarity of 0.75 between the vectors of last name frequencies in these two WOS datasets — so it is reasonable to expect the predictive performance of the logistic regression model to generalize to the new dataset. For a given clustering of S_n, we define y_{ij} to be 1 if papers i and j are in the same cluster and 0 otherwise. Then the likelihood of a given clustering, assuming independence between link probabilities, is $\Pi_{i<j}(y_{ij}p_{ij} + (1-y_{ij})(1-p_{ij}))$, and the log-likelihood is therefore $\sum_{i<j}\log(1-p_{ij}) + \sum_{i<j} y_{ij}\log(\frac{p_{ij}}{1-p_{ij}})$. Thus, we seek the clustering that maximizes the sum of $\log(\frac{p_{ij}}{1-p_{ij}})$ over all pairs of papers i and j that are assigned to the same cluster. We solve this problem with agglomerative clustering, beginning with no nodes assigned to any cluster and iteratively adding the node to the cluster that most improves $\sum_{i<j} y_{ij}\log(\frac{p_{ij}}{1-p_{ij}})$ over the nodes i and j assigned to clusters. If no improving addition can be made to any current cluster, a new cluster is created with one of the unassigned nodes. The agglomerative clustering yielded 136,313 author clusters.

We evaluated the quality of the clustering using a set of 166 scholars who obtained their assistant professorship in 1996 or later and who published at least one paper in the OR literature; we manually identified all WOS for these scholars using CVs and other available publication information while generating the dataset for the tenure analysis. For these scholars, we witnessed *lumping* (assigning two individuals to the same cluster) in 33 of the clusters (19.9%) and splitting (assigning papers from the same individual to different clusters) in 17 clusters (10.2%). Eight of the clusters (4.8%) demonstrated both lumping and splitting, and 13 of the 25 clusters that exhibited lumping had only one or two extra papers. While the majority of scholars were perfectly disambiguated using the method, there are still a number of scholars for whom the assigned clusters were incorrect.

Appendix B: Projected 16-year publication metrics

Using linear models, we seek to project 16-year publication metrics based on all available publication data for a scholar. To perform these projections, we use the 75 scholars identified in Section 2.4 who held assistant professor positions in the OR field starting in 1996 or after.

In Figure 5, we plot the number of years into each of these scholars' career against each of the four publication metrics of interest. Data are only plotted for the years available for a given scholar. For instance, no 13-, 14-, 15-, or 16-year paper counts were available for the scholar with 45 published papers after year 12. Each plot displays some degree of heteroscedasticity, in which the outcome metric's variance increases over time; this effect is particularly dramatic for the citation count outcome. To deal with this heteroscedasticity, we log-transform each outcome variable, predicting $\log(m+1)$ instead of m for each metric m. If we used linear model $\log(m+1) = \beta_0 + \beta_1 y + \epsilon$, where y is the number of years since a scholar became an assistant professor, we would be modeling an exponential growth of the outcome variables through time. However, metrics like a scholar's total

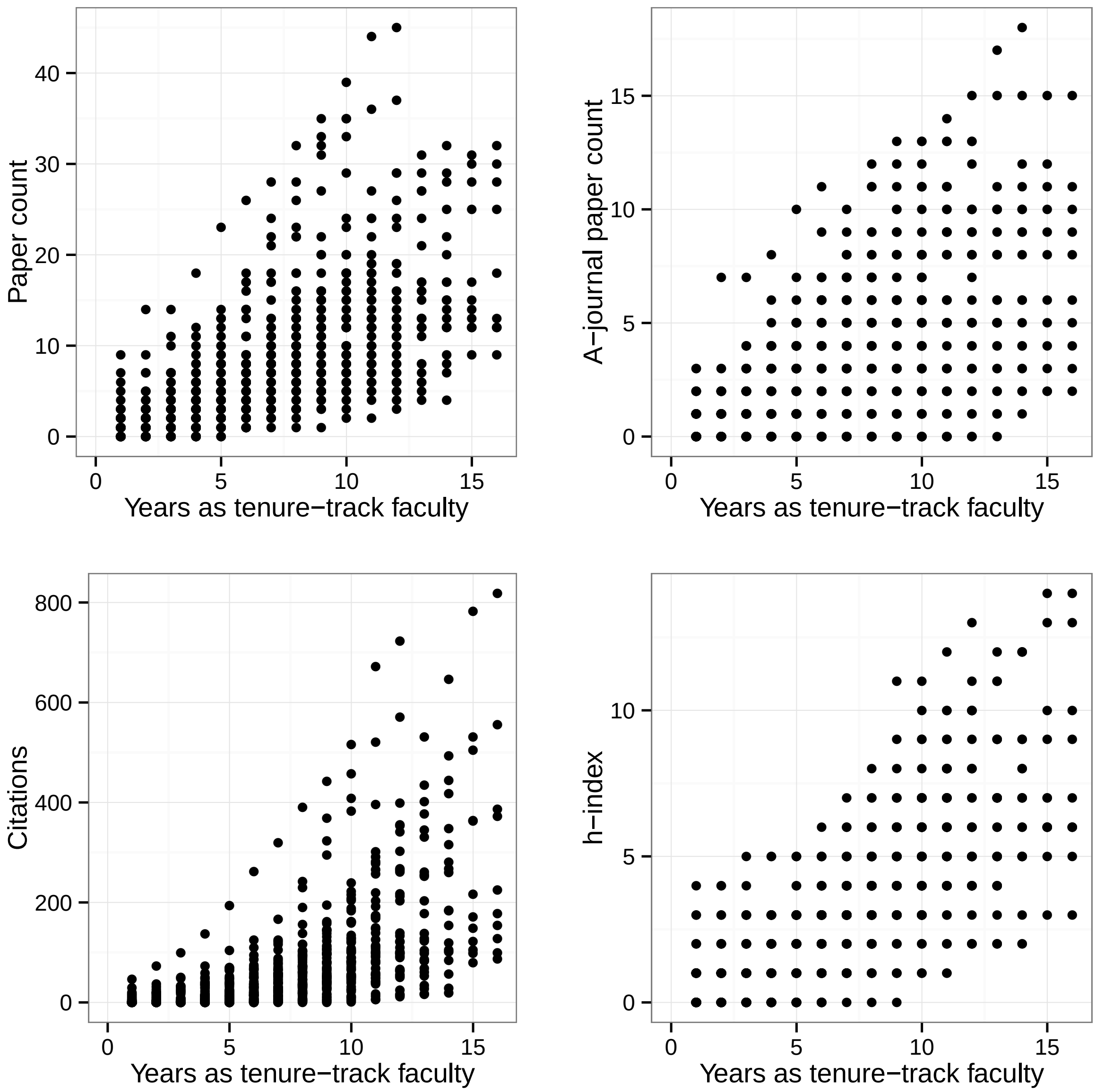

Figure 5 The publication metrics over time among the 75 scholars used to obtain projected 16-year outcomes.

citation count are known to grow polynomially through time (Hirsch 2007). As a result, we used a log-log regression model of form $\log(m+1) = \beta_0 + \beta_1 \log(y) + \epsilon$ for each metric m.

To obtain scholar-specific projections for 16-year publication metrics, we used hierarchical linear models. In these models, each observation is of publication metric value m_{sy} for a scholar s at year y after becoming an assistant professor, and observations are grouped by scholars in the multilevel model. To assess the need for random intercepts and slopes, for each scholar s we fitted a linear regression model $\log(m_{sy}+1) = \hat{\beta}_0 + \hat{\beta}_1 \log(y) + \epsilon$. The resulting intercept and slope estimates for the 75 regression models trained for each outcome metric are displayed in Figure 6. It is clear that both intercept and slope vary across scholars for all metrics, and further that these two coefficients

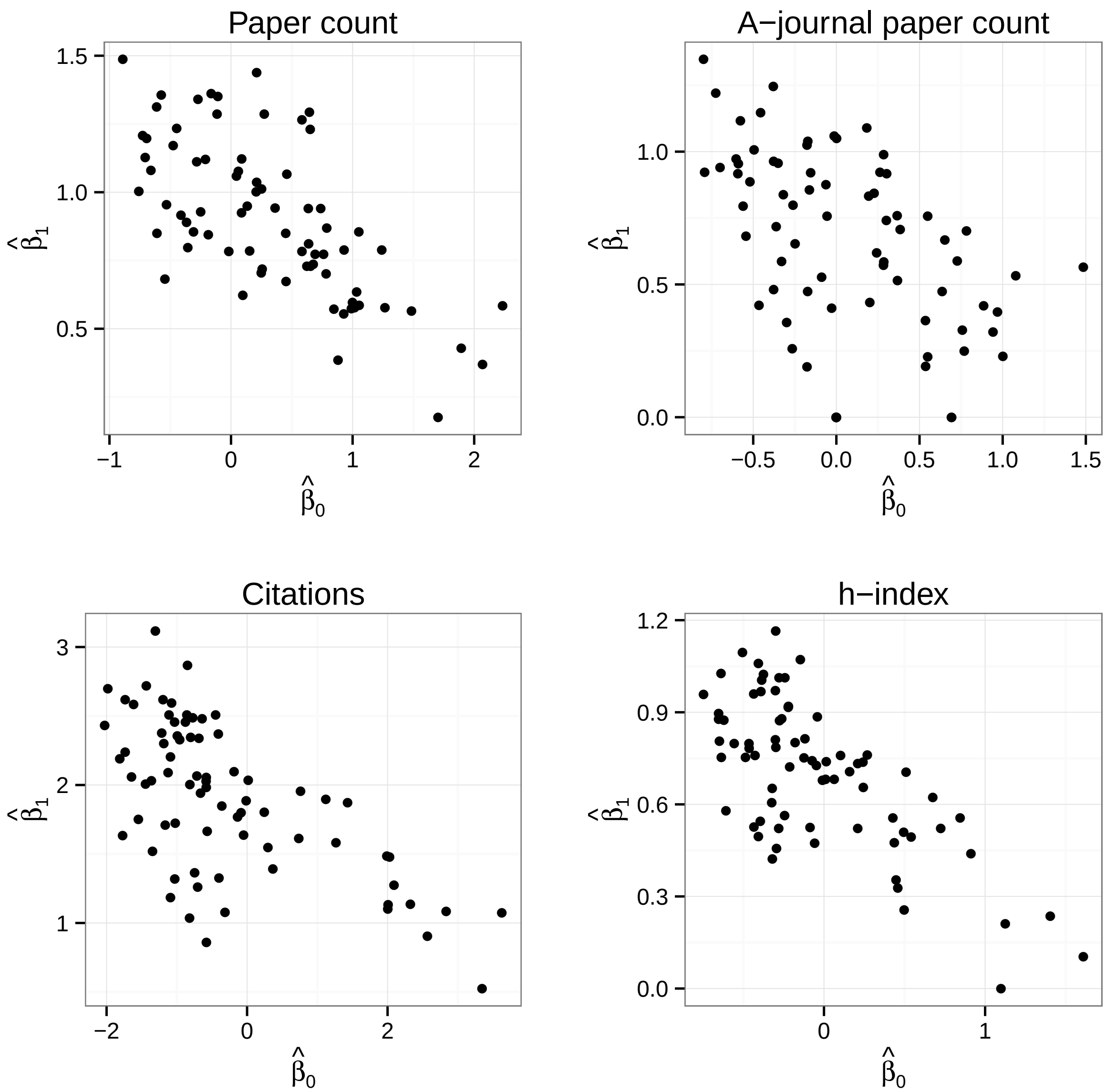

Figure 6 Estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ when fitting scholar-specific linear regression models $\log(m_{sy}+1) = \hat{\beta}_0 + \hat{\beta}_1 \log(y) + \epsilon$ to predict publication metric m_{sy} for scholar s, y years after becoming an assistant professor.

are negatively correlated. As a result, we used random slopes and intercepts in our hierarchical models and allowed correlation between the random intercepts and slopes.

We fit the hierarchical models with the R nlme package (Pinheiro et al. 2013). The models for paper count, A-journal paper count, citation count, and h-index had R^2 values of 0.93, 0.91, 0.91, and 0.88, respectively, and the model fits are presented in Figure 7. As a final step, we used the models to project the 16-year publication metrics for each scholar, using that scholar's random intercept and slope for the projection.

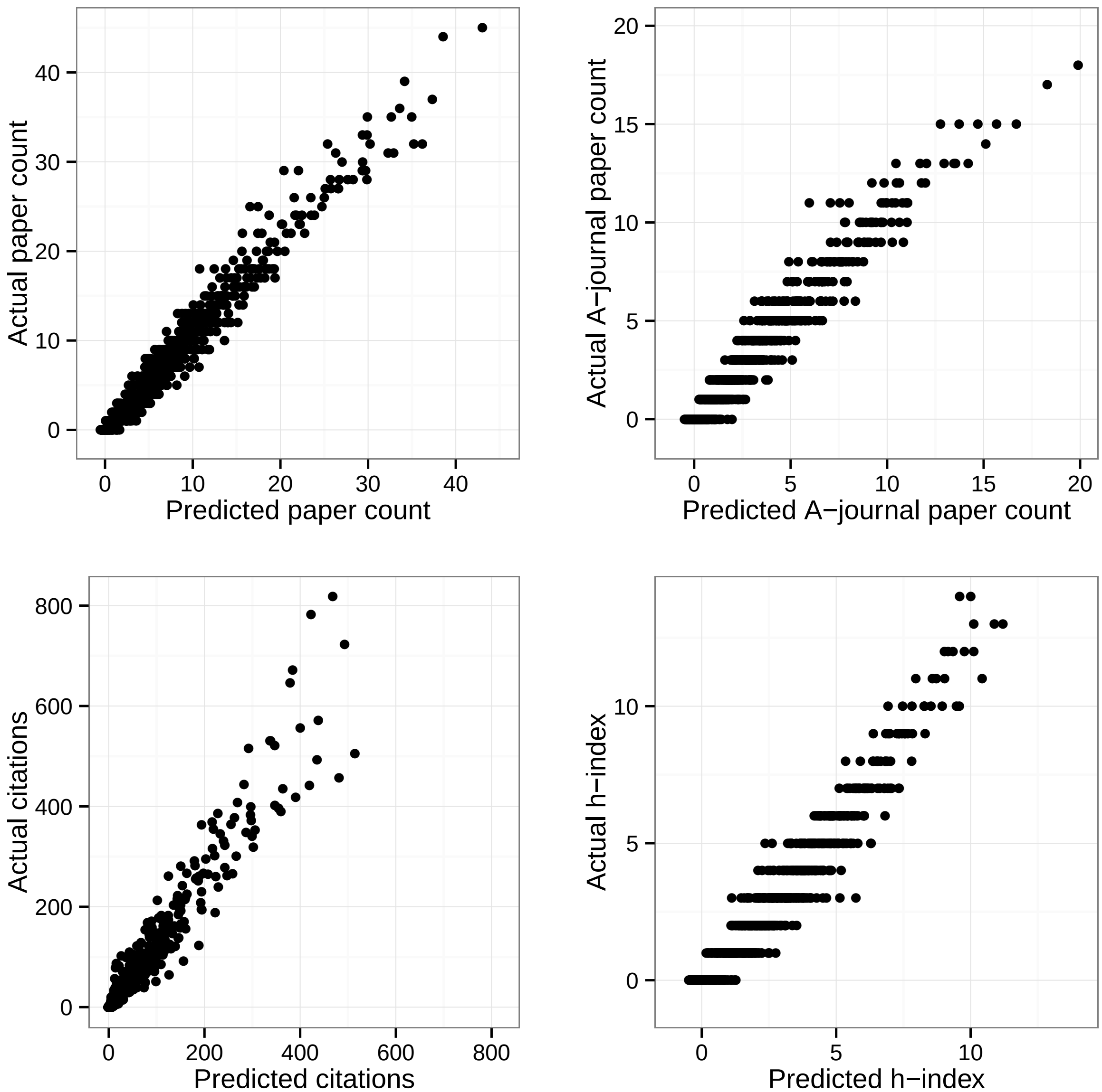

Figure 7 Final model fits for each of the publication metric prediction models.

Acknowledgments

The authors gratefully acknowledge the support of Thomson Reuters, both in providing the data used in this study and in funding the research. The authors also thank the reviewers of the paper, whose comments helped improve the work.

References

Acuna, Daniel, Stefano Allesina, Konrad Kording. 2012. Future impact: Predicting scientific success. *Nature* **489**(7415) 201–202.

Acuna, Daniel E., Orion Penner, Colin G. Orton. 2013. The future h-index is an excellent way to predict scientists' future impact. *Medical Physics* **40**(11) 110601–1–110601–3.

Azoulay, Pierre, Joshua S. Graff Zivin, Gustavo Manso. 2011. Incentives and creativity: evidence from the academic life sciences. *The RAND Journal of Economics* **42**(3) 527–554.

Bader, D. A., K. Madduri. 2006. Parallel algorithms for evaluating centrality indices in real-world networks. *Proceedings of the 35th International Conference on Parallel Processing (ICPP)*. 539–550.

Barabási, Albert-László. 2012. Network science. URL `http://barabasilab.neu.edu/networksciencebook`.

Bornmann, Lutz, Rüdinger Mutz, Hans-Dieter Daniel. 2008. Are there better indices for evaluation purposes than the h index? a comparison of nine different variants of the h index using data from biomedicine. *Journal of the American Society for Information Science and Technology* **59**(5) 830–837.

Breiman, Leo. 2001. Random forests. *Machine Learning* **45**(1) 5–32.

Brin, Sergey, Lawrence Page. 1998. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems* **30**(1) 107–117.

Burt, Ronald S. 1995. *Structural holes: The social structure of competition*. Harvard University Press, Cambridge, MA.

Burt, Ronald S. 2005. *Brokerage and Closure: An Introduction to Social Capital*. Oxford University Press, USA, New York.

Canty, Angelo, Brian Ripley. 2014. *boot: Bootstrap R (S-Plus) Functions*. Package version 1.3-13.

Currie, Russell R., Gurupdesh S. Pandher. 2011. Finance journal rankings and tiers: An active scholar assessment methodology. *Journal of Banking & Finance* **35**(1) 7–20.

Davison, A. C., D. V. Hinkley. 1997. *Bootstrap Methods and Their Applications*. Cambridge University Press, Cambridge.

Dorsey, E. Ray, Brian A. Raphael, Laura Balcer, Steven Galetta. 2006. Predictors of future publication record and academic rank in a cohort of neurology residents. *Neurology* **67**(8) 1335–1337.

Excellence for Research in Australia. 2010. ERA 2010 ranked journal list. URL `http://www.core.edu.au/images/journals/08sortrankalpha-ERA2010_journal_title_list.pdf`.

Freeman, Linton C. 1977. A set of measures of centrality based on betweenness. *Sociometry* **40**(1) 35–41.

Freeman, Linton C. 1979. Centrality in social networks conceptual clarification. *Social networks* **1**(3) 215–239.

Garfield, Eugene, Alfred Welljams-Dorof. 1992. Of Nobel class: A citation perspective on high impact research authors. *Theoretical Medicine* **13**(2) 117–135.

Goldenberg, Jacob, Gal Oestreicher-Singer, Shachar Reichman. 2012. The quest for content: How user-generated links can facilitate online exploration. *Journal of Marketing Research* **49**(4) 452–468.

Hirsch, J. E. 2005. An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences* **102**(46) 16569–16572.

Hirsch, J. E. 2007. Does the h index have predictive power? *Proceedings of the National Academy of Sciences* **104**(49) 19193–19198.

Hönekopp, Johannes, Julie Khan. 2012. Future publication success in science is better predicted by traditional measures than by the h index. *Scientometrics* **90**(3) 843–853.

Hult, G. Tomas M., Martin Reimann, Oliver Schilke. 2009. Worldwide faculty perceptions of marketing journals: Rankings, trends, comparisons, and segmentations. *globalEDGE Business Review* **3**(3) 1–23.

INFORMS. 2014. Award recipients. URL `https://www.informs.org/Recognize-Excellence/Award-Recipients/`.

Katona, Zsolt, Peter Pal Zubcsek, Miklos Sarvary. 2011. Network effects and personal influences: The diffusion of an online social network. *Journal of Marketing Research* **48**(3) 425–443.

Kuhn, Max. 2015. *caret: Classification and Regression Training*. URL `http://CRAN.R-project.org/package=caret`. R package version 6.0-52.

Liaw, Andy, Matthew Wiener. 2002. Classification and regression by randomforest. *R News* **2**(3) 18–22. URL `http://CRAN.R-project.org/doc/Rnews/`.

Mazloumian, Amin. 2012. Predicting scholars' scientific impact. *PLOS One* **7**(11) e49246.

National Science Foundation. 2014a. Faculty early career development (CAREER) program (NSF 14-532). URL `http://www.nsf.gov/pubs/2014/nsf14532/nsf14532.htm`.

National Science Foundation. 2014b. FY 2014 agency financial report.

Newman, M. E. J. 2003. The structure and function of complex networks. *SIAM Review* **45**(2) 167–256.

Newman, M. E. J. 2004. Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences* **101**(Suppl 1) 5200–5205.

Penner, Orion, Raj Pan, Alexander Petersen, Santo Fortunato. 2013. The case for caution in predicting scientists' future impact. *Physics Today* **66**(4) 8–9.

Pinheiro, Jose, Douglas Bates, Saikat DebRoy, Deepayan Sarkar, R Core Team. 2013. *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-113.

Podsakoff, Philip M., Scott B. MacKenzie, Nathan P. Podsakoff, Daniel G. Bachrach. 2008. Scholarly influence in the field of management: A bibliometric analysis of the determinants of university and author impact in the management literature in the past quarter century. *Journal of Management* **34**(3) 641–720.

R. Kelly Rainer, Jr., Mark D. Miller. 2005. Examining differences across journal rankings. *Communications of the ACM* **48**(2) 91–94.

Ritzberger, Klaus. 2008. A ranking of journals in economics and related fields. *German Economic Review* **9**(4) 402–430.

Theoharakis, Vasilis, Mary Skordia. 2003. How do statisticians perceive statistics journals? *The American Statistician* **57**(2) 115–123.

Valente, Thomas W. 1996. Network models of the diffusion of innovations. *Computational & Mathematical Organization Theory* **2**(2) 163–164.

Wasserman, Stanley, Katherine Faust. 1994. *Social network analysis: Methods and applications*. Cambridge University Press, New York and Cambridge, England.

Watts, Duncan J., Steven H. Strogatz. 1998. Collective dynamics of 'small-world' networks. *Nature* **393**(6684) 440–442.

Yang, Glen, Uwais Zaid, Bradley Erickson, Sarah Blaschko, Peter Carroll, Benjamin Breyer. 2011. Urology resident publication output and its relationship to future academic achievement. *Journal of Urology* **185**(2) 642–646.