MIT IDE RESEARCH BRIEF VOL. 2016. 11

MONEYBALL FOR PROFESSORS: MODELS FOR PREDICTING RESEARCH IMPACT

Dimitris Bertsimas, Erik Brynjolfsson, Shachar Reichman, John Silberholz

Tenure decisions are the key human resource (HR) judgments made by academic institutions in the U.S. The impact of these decisions is not solely limited to scholars' careers, but also influences the ranking of departments, the funding of universities, and the strength of scientific research in general.

We developed a method based on quantitative metrics that can predict research performance better than earlier predictive models.[1] This "Moneyball for professors" also outperforms traditional tenure committee decisions.

The implications of these decisions are widespread. A tenured faculty member at a prestigious university will receive millions of dollars in career compensation, and will occupy a faculty spot for decades. On a broader scope, the National Science Foundation provided $5.8 billion in research funding in 2014, including $220 million specifically for young researchers at top universities.

Despite the importance, however, these decision-making processes rely mainly on subjective assessments of candidates by personnel and funding committees that typically consider only research output following the candidate's doctoral degree graduation.

The use of quantitative methods in these processes is usually very limited. Given the stakes, and the boom of predictive analytics in the HR industry, we think it is time for a "Moneyball moment" in academia. Models predicting future outcomes can be used to support tenure decisions for early-career faculty. Moreover, our research finds that there's strong potential for data-driven models to be used as decision aids for academic and financial committees that will improve selections.

1.Dimitris Bertsimas, Erik Brynjolfsson, Shachar Reichman, John Silberholz (2015) OR Forum—Tenure Analytics: Models for Predicting Research Impact. Operations Research 63(6):1246-1261 http://dx.doi.org/10.1287/opre.2015.1447

2.Hirsch JE (2005) An index to quantify an individual's scientific research output. Proc. Natl. Acad. Sci. 102(46):16569–16572.

3.Podsakoff PM, MacKenzie SB, Podsakoff NP, Bachrach DG (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. J. Management 34(3): 641–720.

MIT INITIATIVE ON THE DIGITAL ECONOMY

Researchers have previously examined scholarly work and tenure appointment, but the dial hasn't moved much and traditional methods, which essentially count total numbers of citations, largely remain in place. Most notably, in 2005, J.E. Hirsch[2] presented the case for the *h-index*. For this metric, a scientist is assigned an index of *h*, if *h* of her *N* papers contains at least *h* citations, and the other *N h* papers have no more than *h* citations each. Several later studies extended, modified, and offered alternatives to the *h-index*, such as including years in the field, graduate school attended, editorial board memberships, etc.[3]

For our study, we used a large-scale database containing 198,310 papers published during 1975–2012 in the field of Operations Research (OR) to test the whether a scholar would perform well on a number of future success metrics.

THE USE OF QUANTITATIVE METHODS IN THESE PROCESSES IS USUALLY VERY LIMITED. GIVEN THE STAKES, AND THE BOOM OF PREDICTIVE ANALYTICS IN THE HR INDUSTRY, WE THINK IT IS TIME FOR A "MONEYBALL MOMENT" IN ACADEMIA.

IN THIS RESEARCH BRIEF

- Current academic tenure and funding decision processes are too often based on subjective assessments of candidates, not quantitative methods. We feel it is time for a "Moneyball moment" in academia in which models predicting future outcomes are used to support tenure decisions for early-career faculty.
- The data-based models significantly outperform simple predictive models based on citation counts alone. By using a data set of 54 scholars, these statistical models made a different decision than the tenure committees for 16 (30%) of the candidates and the outcomes from our models were better.
- There's strong potential for data-driven models and analytics to complement the tenure decision-making process in academia and improve predictions about performance and scholarly research.
- Using a bibliometric database of Operations Research papers,and we found that a scholar's publications early in their career can be used to predict later-career success.

Dimitris Bertsimas, Erik Brynjolfsson, Shachar Reichman, John Silberholz

We used statistical models based on data from the scholar's first five years of publication, a subset of the information available to tenure committees.

Our models use a concept called "network centrality" to measures how connected a given scholar is in the citation network, the co-authorship network, and a dual network combining the two (see figure 1.)

We found this approach significantly outperformed simple predictive models based on citation counts alone. In addition, by using a data set of 54 scholars who obtained a Ph.D. after 1995, and held an assistant professorship at a top-10 OR program in 2003 or earlier, these statistical models made a different decision than the tenure committees for 16 (30%) of the candidates.

Specifically, these new criteria yielded a set of scholars with significantly better A-journal paper count, and better citation counts, as well as better *h-indexes*, compared with the scholars actually selected by tenure committees (see figure 2). A-journals are defined as publications in Management Science, Mathematical Programming, Mathematics of Operations Research, or Operations Research.

Perhaps even more importantly, these results also show that analytics can complement the tenure decision-making process in academia and improve predictions about performance and scholarly research. In turn, businesses and the public are better served by the academic community.

METHODOLOGY

To evaluate whether prediction models for future academic success could be useful to tenure committees, we built a data set of OR scholars who obtained a Ph.D. since 1996.

To obtain a more homogenous set of scholars, we limited our analysis to those who held an assistant professorship at a top-10 university for OR, as determined by the number of INFORMS fellows at the university.

PERHAPS EVEN MORE IMPORTANTLY, THESE RESULTS ALSO SHOW THAT ANALYTICS CAN COMPLEMENT THE TENURE DECISION-MAKING PROCESS IN ACADEMIA AND IMPROVE PREDICTIONS ABOUT PERFORMANCE AND SCHOLARLY RESEARCH.

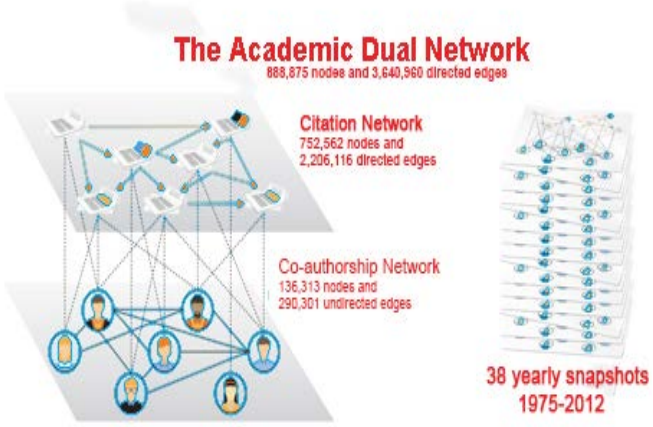

FIGURE 1: NETWORKS USED TO COMPUTE CENTRALITY MEASURES IN PREDICTION OF ACADEMIC IMPACT. IN THE CITATION NETWORK, NODES ARE PAPERS, AND DIRECTED EDGES ARE CITATIONS BETWEEN PAPERS. IN THE CO-AUTHORSHIP NETWORK, NODES ARE AUTHORS, AND EDGES INDICATE CO-AUTHORSHIP OF AT LEAST ONE PAPER. THE DUAL NETWORK COMBINES THESE TWO NETWORKS, ADDING AN EDGE BETWEEN AN AUTHOR AND EACH OF THEIR PAPERS. SOURCE: OPERATIONS RESEARCH

The set of universities used in this analysis were Carnegie Mellon University, Columbia University, Cornell University, Georgia Institute of Technology, Massachusetts Institute of Technology, University of Michigan, Princeton University, Stanford University, University of California, Berkeley, and University of California, Los Angeles.

We identified scholars who held a position in one of the target universities and we gathered a set of 41,103 presentation records from this time period from the INFORMS.org website. We then narrowed the set of 15,178 records further by filtering records to one of our target universities. We manually reviewed this set of presentation records, obtaining a set of 685 scholars. For each of the 685, we searched publicly available information to obtain the year range of each tenure-track position that scholar has held, as well as the year they obtained their Ph.D.

Although 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. The 203 scholars in the limited data set included members from computer science, public health, and chemical engineering so we removed any scholars with fewer than half of their journal articles present in our database of OR publications, limiting our analysis to 75 scholars.

To obtain the publications of each scholar in our publication data set, we manually linked publications to papers.

Finally, we identified the 61 scholars based on whether they had received tenure at their first top-10 institution: 35 scholars had achieved this. For the remaining 26, we determined if they either left because tenure was not granted (or was not going to be granted) or for other reasons, like a personal or family purpose.

We measured short-term and medium-term success in publication using a scholar's paper count, number of A-journal publications, *h-index*, and citation count nine and 16 years after either first publishing or becoming an assistant professor.

We evaluated tenure decisions using a number of characteristics of interest to tenure committees, but that need to be collected manually for each scholar such as research or teaching award.

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. 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.

THE PAST AS A PRELUDE TO 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. For this purpose, we defined statistical models to predict a set of metrics using only centrality measures available within five years of an author's first publication.

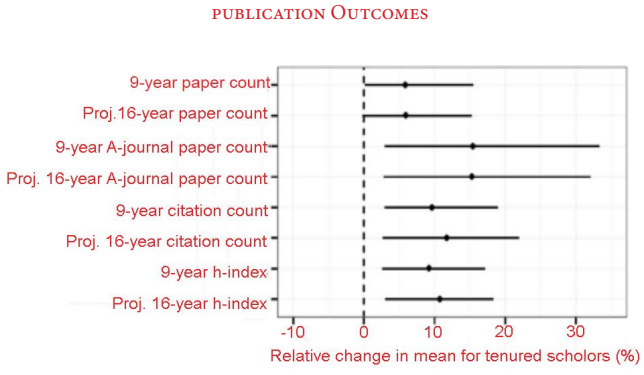

FIGURE 2: RELATIVE CHANGE IN MEAN PUBLICATION OUTCOMES FOR SCHOLARS SELECTED BY OUR MODELS COMPARED WITH SCHOLARS TENURED BY UNIVERSITIES. OUTCOMES ARE MEASURED NINE AND 16 YEARS AFTER SCHOLARS WERE HIRED AS AN ASSISTANT PROFESSOR. SOURCE: OPERATIONS RESEARCH

We also 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.

It remains to be seen whether these models can be useful to tenure committees, as committees have access to information not available to the models, including forthcoming papers, the text of published papers, teaching evaluations, and letters of support.

Our data sets sought to compensate for this information. To compare the tenure decision-making process currently being used by universities to the proposals made by the network centrality models we ranked the scholars by their predicted value for each of the publication metrics, using publication information from five years after assistant professorship as the independent variables for each scholar.

Among the 54 scholars in the data set, 35 (65%) were tenured at a top-10 university. so we labeled the top-ranked 35 scholars as the "tenure selections" of the network centrality models. The network centrality models agreed with tenure committees on 38 (70%) of the scholars, granting tenure to eight scholars not selected by the committees and not granting tenure to eight scholars selected by the committees.

SIGNIFICANT FINDINGS

A key finding was that using our methods we were able to predict later-career success better than earlier, simpler predictive models or tenure committees. Furthermore, in 30% of cases, our model recommended a different tenure decision than tenure committees, and the candidates selected by the model had better research performance than those given tenure by the committees. This 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. Although the models proposed in this work rank scholars based on predictions of various measures of future research productivity, they do not account for other important considerations for tenure, such as a scholar's service to their university, teaching ability, or personality.

Like research output, some of these other criteria can also be quantified, and we demonstrated that the those 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 distribu tion of subfield, gender, or typical number of coauthors.

However, other criteria, such as personality or creativity[4] 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 data set with identical five-year research productivity values (paper count, A-journal paper count, citation count, and *h-index*), one pair 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 has other limitations, too. 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 those selected for tenure by the models presented in this work.

Furthermore, 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.

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, such limits don't necessarily exist.

4. Azoulay P, Zivin JSG, Manso G (2011) Incentives, and creativity: Evidence from the academic life sciences. RAND J. Econom. 42(3): 527-554.

The models described in this work could be expanded a number of ways. First, the data sources 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. Although we also believe the proposed models could be useful for other research fields and for programs outside the top-tier, the only way to confirm the broader effectiveness of the proposed methodology is to test it in other settings.

In addition, we only considered models for the tenure deci sion. 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. More experimentation is needed to evaluate the usefulness of predictions of future research impact in making these decisions.

One way the models can be implemented more widely is for them to be developed and distributed as a complementary service to an existing bibliometric database like Google Scholar or the Thomson Reuters Web of Science. Models also 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, we expect that candidates might change their publication behavior to boost their centrality in citation and co-authorship networks, prompting further recalibration of the proposed model.

For the prediction models to be useful to tenure committees, they need to be implemented and separately calibrated for a broad range of academic dsciplines using a large-scale bibliometric database. Though broader evaluation is needed, we're encouraged by the findings.The demonstrated effectiveness of these models in the field of OR suggests great potential for data-driven models as decision aids to academic personnel committees.

PERHAPS THE MODELS COULD BE DEVELOPED ANDDISTRIBUTEDASACOMPLEMENTARYSERVICE TO AN EXISTING BIBLIOMETRIC DATABASE LIKE GOOGLE SCHOLAR OR THE THOMSON REUTERS WEB OF SCIENCE. MODELS ALSO WOULD NEED TO BE UPDATED PERIODICALLY, AS PATTERNS OF PUBLICATION CHANGE OVER TIME.

MONEYBALL FOR PROFESSORS: MODELS FOR PREDICTING RESEARCH IMPACT

Dimitris Bertsimas, Erik Brynjolfsson, Shachar Reichman, John Silberholz

MIT INITIATIVE ON THE DIGITAL ECONOMY

The MIT Initiative on the Digital Economy brings together internationally recognized researchers seeking solutions to how people can – and will – thrive in a digital world. Drawing on MIT's strengths in technology and innovation, IDE explores the profound impact of a rapidly advancing digital economy, and how it's changing the ways we live and work.

SUPPORT THE MIT IDE

Foundations, private donors and corporate members are critical to the success of the IDE. Their support fuels cutting-edge research by MIT faculty and graduate students, and enables new faculty hiring, curriculum development, events, and fellowships. Contact Christie Ko (cko@mit.edu) to learn how you or your organization can support the IDE.

TO LEARN MORE ABOUT THE IDE, INCLUDING UPCOMING EVENTS, VISIT MITSLOAN.MIT.EDU/IDE

Dimitris Bertsimas is the Boeing Leaders for Global Operations Professor of Management, a Professor of Operations Research, and the Co-director of the Operations Research Center at MIT.

Erik Brynjolfsson is the Director of the MIT Initiative on the Digital Economy, the Schussel Family Professor at the MIT Sloan School, and Chairman of the MIT Sloan Management Review.

Shacher Reichman is an Assistant Professor at the Coller School of Management at Tel Aviv University and a Research Affiliate at MIT Sloan School of Management.

John Silberholz is a Lecturer in the MIT Sloan School of Management.

The full paper can be found in Operations Research here