MONEYBALL FOR ACADEMICS:
NETWORK ANALYSIS FOR PREDICTING RESEARCH IMPACT

Research-in-Progress

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

Erik Brynjolfsson
Sloan School of Management,
Massachusetts Institute of Technology
Cambridge, MA
erikb@mit.edu

Shachar Reichman
Sloan School of Management,
Massachusetts Institute of Technology
Cambridge, MA
shachar@mit.edu

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

Abstract

How are scholars ranked for promotion, tenure and honors? How can we improve the quantitative tools available for decision makers when making such decisions? Can we predict the academic impact of scholars and papers at early stages using quantitative tools?

Current academic decisions (hiring, tenure, prizes) are mostly very subjective. In the era of "Big Data," a solid quantitative set of measurements should be used to support this decision process.

This paper presents a method for predicting the probability of a paper being in the most cited papers using only data available at the time of publication. We find that highly cited papers have different structural properties and that these centrality measures are associated with increased odds of being in the top percentile of citation count.

The paper also presents a method for predicting the future impact of researchers, using information available early in their careers. This model integrates information about changes in a young researcher's role in the citation network and co-authorship network and demonstrates how this improves predictions of their future impact.

These results show that the use of quantitative methods can complement the qualitative decision-making process in academia and improve the prediction of academic impact.

Keywords: Citation analysis, Academic impact, Analytics, Networks

Introduction

How are scholars ranked for promotion, tenure and honors? How can we improve the quantitative tools available for decision makers when making such decisions? Can we predict the academic impact of scholars at early stages using quantitative tools?

In academia, some of the most important decisions facing personnel and funding committees concern young researchers. Personnel committee members 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 the research and support letters, which determines the final decision. The use of quantitative methods as part of these processes is usually very limited.

The impact of these decisions is not solely limited to scholars' careers, but also influences the ranking of departments and prestige of universities. 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 over $6 billion in research funding in 2012, including $150 million specifically for young researchers' awards.[1] With this at stake, we suggest that in the era of "Big Data," a solid quantitative set of measurements should be used to support the decision process.

In sports, when management considers signing an athlete, they first need to evaluate the contribution of this athlete to their team, clearly a subjective judgment. However, this decision is supported by a wide set of quantitative methods (e.g., baseball Sabermetrics), evaluating a wide range of variables, from the athlete's physical characteristics, her performance over the last year in different scoring measurements to overall career statistics. The analytics of the athlete's statistics generate a full picture of her abilities, her past performance and a prediction of her future success, which becomes a widespread technique for making these decisions (Davenport and Harris, 2007). In the business domain, a recent work, (Brynjolfsson et al., 2011) found that firms that adopt data-driven decision-making outperformed firms that make their decisions in the traditional way of experience and intuition. We propose that the academia, like other

[1] From the NSF Agency Financial Report - http://www.nsf.gov/pubs/2013/nsf13002/.

business fields, could benefit from the use of analytics as part of the decision-making process

In this work, we develop and evaluate tools to support the academic decision making at early stages of a paper's publication and at early stages of a researcher's career. In particular, we leverage the structure of the citation network, the graph with published papers as vertices and citations as edges, to predict the future impact of a paper based on data available at the time of publication. We expand the analysis to predict researchers' future successes by aggregating the structural roles of their early career papers. In particular, we use the early career information to predict if a researcher will receive a major career award.

Theory and Literature

In recent years, there have been significant research efforts for predicting the future impact of a paper based on data available at early stages of its lifetime. Initial interest in the problem of predicting citations came from the KDD Cup 2003 competition, in which contestants predicted citation counts for papers in the next three months (Gehrke et al., 2003). In early approaches to the problem of the early prediction of paper outcomes, authors have relied on simple information describing a new paper like properties of the author and journal and words appearing in the title, abstract, and article text (Fu and Aliferis, 2008, Honekopp and Kleber, 2008, Ibanez et al., 2009, Yogatama et al., 2011). Other approaches have improved predictions by including some information available soon after publication, such as paper ratings by experts (Lokker et al., 2008), citations soon after publication (Adams, 2005), the number of paper downloads soon after publication (Brody et al., 2006), or the number of Twitter mentions soon after publication (Eysenbach, 2011). Some work has investigated whether structural information about the citation network predicts future citations (Shibata et al., 2007), and a recent approach started to leverage citation network information available at publication time to predict a paper's future impact (Livne et al., 2013). In this work, we demonstrate the predictive improvements obtained by including a citation network structure into predictions of the future paper's impact for a large set of papers in the management, information systems and operations research literature.

Considering the importance of academic decisions, it is not surprising that the measurement of scholars' impact has also received extensive attention in the literature. Most noticeably, 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). Egghe (2006) proposed the g-index which is based on the distribution of citations received by a given researcher's publications. Harzing (2012) suggested the m-quotient, which incorporates the length of the researcher's academic career into 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.).

In the information systems (IS) field, as IS research stems from and interacts with many other academic disciplines, even the ranking of journals (excluding MISQ and ISR) is not well accepted. Hence, the ranking of scholars is even more complicated. Venkatesh (2009) presented a ranking of both scholars and departments based on the number of publications in different sets of journals. Dean et al. (2011) employs this framework to examine how tenure decisions are affected by the inclusion of more journals in the journal "basket," and the effect of publishing in highly rated non-IS business journals on the tenure decision.

With the wide selection of scholar ranking, it is somewhat surprising that the predictions of academic outcome has received limited attention in the literature. Garfield and Welljams-Dorof (1992) correlated citation counts with Nobel Prizes, revealing the power of citation analysis to forecast Nobel Prize winners and "of-Nobel class" scholars. Hirsch (2007) and Acuna et al. (2012) found that the h-index is a good predictor of future achievement when compared with the total number of citations, papers or citations per paper.

While many ranking algorithms have been proposed, they all account for the number of citations as the primary part of the ranking method. In this paper, we examine the network position and specifically, the scholar's centrality in the co-authorship network and the aggregated network centrality of her papers in the citation network. We integrate these centrality measures into the scholars' future impact prediction algorithm.

To summarize, this paper contributes to the existing literature by presenting a method to predict future paper's and author's impact in the field of management, information systems and operations research, using citation network and co-authorship network measures, which have not previously been well studied. The paper presents a method to predict the probability of a paper being in the most cited papers using only data on paper available at (or very close to) time of publication.

Using information available early in an author's career, we predict the future impact of researchers, building what we believe to be the first prediction model of practical use to personnel and grant committees. This model integrates information about changes in a young researcher's role in the citation network and co-authorship network. It further demonstrates how this improves predictions of their future impact.

Data and Measures

We collected data from Thomson-Reuters Web of Knowledge on over 700,000 papers and 2,200,000 citations that were published in management, information systems and operational research journals from 1975 to 2012.

Using this data, we are able to examine the evolution of two types of networks—the citation networks and co-authorship networks, and to conduct an in-depth analysis of how the changes of these network structures can assist in predicting the future success and impact of papers and scholars.

Additionally, we used three datasets of distinguished academic awards: 1) 57 Association of Information Systems (AIS) Fellows Recipients, 2) 15 recipients of the INFORMS Information Systems Society Distinguished Fellows awards, and 3) 292 INFORMS fellows. The first two awards recognize individuals who had made outstanding contributions to the IS discipline, while the third is an award that is given to outstanding lifetime achievement in operations research and management science.

Citations and Co-Authors Network Analysis

We study how network position and specifically the network centrality of papers in the citation network and authors in the co-authorship network can be integrated into a future impact prediction algorithm. The idea to include network indexes into prediction methods stem from the fact that a citation represents a flow of information. A research idea that was presented in one paper is built upon in another paper. Recent literature relates network structure properties to information dissemination in networks (e.g., Valente, 1996, Mayzlin, 2002, Trusov et al., 2009, Katona et al., 2011). In the co-authorship network, centrality of an author may indicate better access to new information, better opportunities of new collaborations and even may reflect the multidisciplinary levels of the authors (Newman, 2004). Additionally, structural importance of an author (e.g., higher centrality) may also indicate a unique role in the network, which allows her to effect the flow of information due to the fact that they separate non-redundant sources of information (Burt, 2005, 2009).

Typically, network analysis research treats the network as a static current state, which is the most recent citation and co-authorship networks that include all the papers published to date. Since citation and collaboration networks 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, using their references as the links to the former citation network. The co-authorship network snapshots are generated in a similar way, including any collaboration instance up to each year. This method creates a better representation of the centrality role of both the papers and the scholars over the years.

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):

1. *Closeness Centrality* of a node is a measure of the average minimal distance (number of hops) between this node and any other node in the network. In a citation network, closeness represents the average number of papers one needs to follow through the references of other papers in order to navigate from a single paper to any other paper in the network.

2. *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). The formal definition is:

   $$BC(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

   where σ_{st} is the number of shortest paths from node s to node t, and $\sigma_{st}(v)$ is the number of shortest paths from s to t that pass through node v, and V is the set of all nodes.

3. A *Clustering Coefficient* of a node quantifies how close a node's neighbors are to being a clique (Watts and Storgatz, 1998). In other words, the clustering coefficient reflects the extent to which the neighbors of an actor (node) are also neighbors of each other. In the context of a citation network, the clustering coefficient enables us to identify a closed set of authors and papers that tend to mostly cite other members of that community (e.g., a sub-topic with a small number of researchers). The formal definition of a clustering coefficient is:

   $$CC_i = \frac{2L_i}{k_i(k_i - 1)}$$

   where L_i is the number of links between the k_i neighbors of node i (Barabási, 2012).

4. The PageRank of a node measures the node's relative importance within the set of nodes in the network (Brin and Page, 1998). The formal definition of the PageRank of node i is:

$$PR(i) = \sum_{v \in B_i} \frac{PR(v)}{L(v)} \quad , \quad \sum_i PR(i) = 1$$

where B_i is the set of all nodes linking to node i, $PR(v)$ is the *PageRank* of node v and $L(v)$ is the number of outgoing links of node v.

For each yearly snapshot of the citation network, we calculate the centrality measures of all the papers that are included in the network (papers published before or in that year). Similarly, we calculate the centrality measures of any author in each snapshot of the co-authorship network. While co-authorship and citation networks could be treated as a single network, we chose to use each type separately to emphasize the role of a node (a paper or an author) where all the other nodes are of the same type.

Our conjecture is that these measures may indicate the importance of a paper in the flow of new research ideas—the citation network, and therefore may affect its academic impact. The co-authorship network reflects patterns of collaboration within the academic community (Newman, 2004). Thus, centrality of an author may indicate her role in this community and influence her academic success.

Results

Citation Network Analysis

We performed a network analysis of a snapshot of the citation network at each year from 1975-2012. We calculated the centrality measures described above for each paper in our dataset. Specifically, we looked at the centrality of papers at early stages, zero to five years after publication. Year zero indexes measure the centrality at the time of publication. Year one is the first network that captures citation patterns over a whole year period, and by year five, the citation patterns of a paper were established. In general, these time periods are still considered early stages in a publication's lifetime.

We then computed the ratio of these centrality measures between highly cited papers and the rest of the papers. Table 2 and Figure 1 present the results of this comparison.

These results are quite surprising. It is clear that highly cited papers have a significantly higher Betweenness Centrality (BC), even at early stages after publication. This may indicate a unique reference structure of these papers. They tend to connect between areas that are less connected in the citation network and thus, their betweenness is higher. In other words, these papers play an important role in increasing network connectivity. We also see a difference in the average Clustering Coefficient at time zero, demonstrating lower clique tendency of highly cited papers. Over the years, this difference increases significantly, and by year five, the clique tendency of highly cited papers is less than half of that of the papers that were not highly cited.

Table 1. A comparison of the normalized Betweenness Centrality (Norm_BC) and Clustering Coefficient (CC) indexes of the top 1% for cited papers vs. the rest of the papers (Average 1975-2007)

Average Ratio	Year zero-time of publication	One year after publication	Two years after publication	Five years after publication
$\frac{\text{Norm\_BC}_{\text{highly cited}}}{\text{Norm\_BC}_{\text{Not highly cited}}}$	7.88	14.86	17.87	21.95
$\frac{\text{ClusteringCo}_{\text{highly cited}}}{\text{ClusteringCo}_{\text{Not highly cited}}}$	0.86	0.77	0.69	0.44

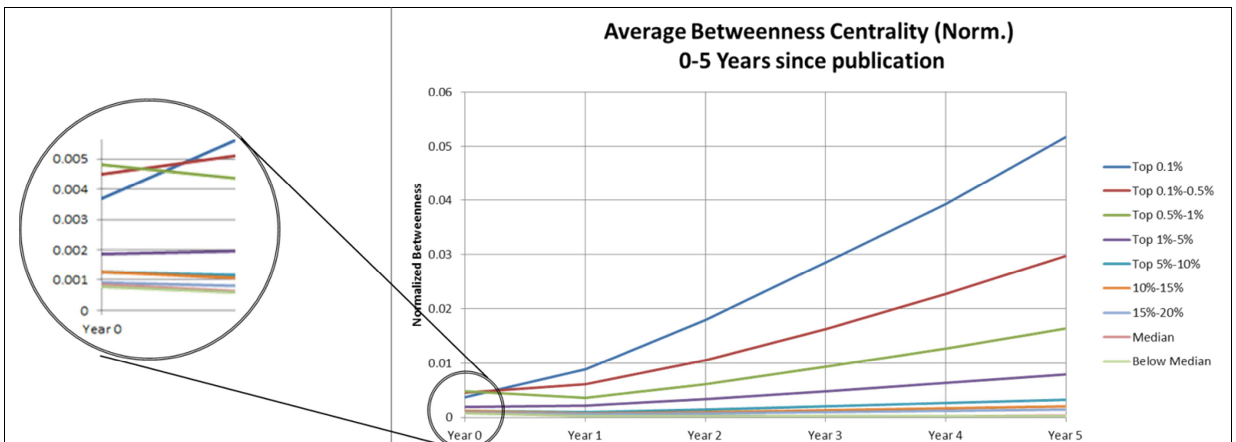

Figure 1. Betweenness Centrality of the highly cited papers vs. the rest of the papers, at time of publication (left figure) and 0-5 years after publication

Taken together, this implies that highly cited papers have different link structures and different roles in the citation network, from the day of publication. In the next steps of our analysis, we use these differences to predict the outcome of a paper at early stages.

Prediction of Paper Outcomes

This section focuses on predicting the impact of a paper near its time of publication. We start the analysis with time zero data, data available at time of publication. We used the citation network snapshot at the time of publication and calculated the aforementioned centrality measures for each of the papers. We measure impact as the total number of citations the paper received to the year of 2012. We calculate the citation count percentile at 2012 of all the papers that were published in the same publication year. The comparison with papers published in the same year allows us to control for the papers' age as well as the differences in network characteristics of different years. We used a logistic regression model to classify a given paper as a highly cited (in the top 0.1%, 0.5%, or 1% in number of citations at 2012):

$$ln\frac{p(y)}{1-p(y)} = \beta_0 + \beta_{BC_0}X_{BC_0} + \beta_{CC_0}X_{CC_0} + \beta_{PR_0}X_{PR_0} + \beta_{CL_0}X_{CL_0} \quad (1)$$

where X_{BC_0} is the betweenness centrality of the paper at time zero; X_{CC_0} is the clustering coefficient of the paper at time zero; X_{PR_0} is the PageRank of the paper at time zero and X_{CL_0} is the closeness of the paper at time zero.

We randomly selected 70 percent of the papers as a training set and 30 percent of the data as a validation set. The results of the regression (presented in Table 3) show that even at time zero, the network centrality measures are associated with higher odds of a paper being highly cited. In particular, betweenness centrality has a positive effect on the odds ratio, meaning that papers that act as better brokers in the citation network are more likely to be at the top of cited papers. Clustering coefficient on the other hand, has a negative effect, indicating that an increase in the paper's references being a clique is correlated with a decrease in its odds of being a top cited paper.

Table 2. A logistic regression estimation of paper being in the top 0.1%, 0.5% and 1% of highly cited papers of the same year of publication						
	Top 0.1%		Top 0.5%		Top 1%	
	β	Std.	β	Std.	β	Std.
(Intercept)	-7.06 ***	0.14	-5.35 ***	0.06	-4.67***	0.04
Betweenness Centrality (time zero)	0.00006 ***	0.00	0.00007 ***	0.00	0.00009 ***	0.00
Clustering Coefficient (time zero)	-2.05 *	1.08	-1.07 **	0.37	-0.3	0.2
PageRank (time zero)	-6.03	271.6	10150	6685	-4.13	5.35
Closeness (time zero)	2241	91860	-10680	177700	-34630	166600
*** denotes significance at 0.001 level; ** significance at 0.01 level; * significance at 0.05 level						

To demonstrate the effectiveness of the predictions of these metrics, we used Receiver Operating Characteristic (ROC) curve to evaluate the performance of the model. Our empirical results shown in Figure 2 show that the model can differentiate between random true positive and true negative (AUC - The area under the ROC curve) 78 percent of time for the top 0.1%, 74 percent of the time for the top 0.5% and 70 percent of the time for the top 1% of cited papers.

These findings indicate that the citation network centrality measures at time of publication are associated with the impact of the paper as measured by future citation count. The more central a paper is in the network and the less its references are clustered, the likelihood of it being a top cited paper increase.

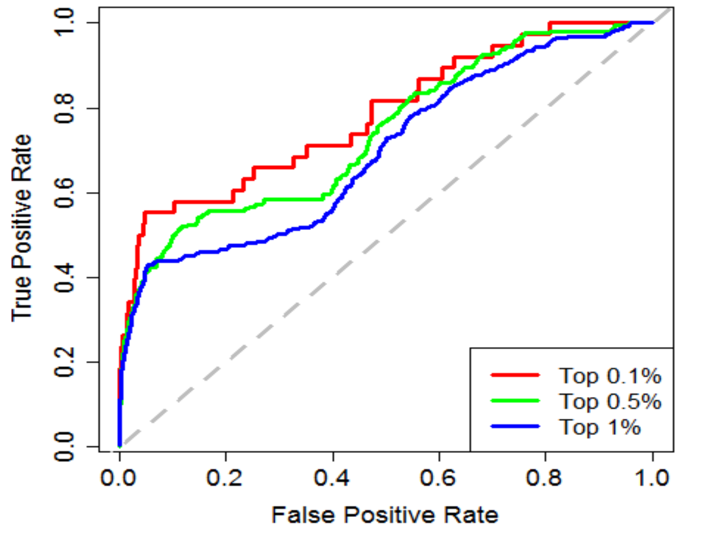

Figure 2. Receiver operating characteristic (ROC) curves of top cited papers based on data available at publication time

Prediction of Distinguished Academic Awards

The next step of our analysis study scholars' impact, where we look at three career awards—AIS Fellows, INFORMS IS Society Distinguished Fellows and INFORMS Fellows, as a measure of a distinguished academic outcome. We used only early career data, from the first five years of the scholars' careers, as input to the model.

We used logistic regression to predict which scholars will receive an award using four different models. We define the total number of citations as the baseline model against which we compare the results of the other prediction models.

First, we examine how scholars' positions in the co-authorship networks over the years may influence the probability of receiving career awards (model 1). We then looked at the aggregate network position of the scholar's publications in the citation networks over the first five years (model 2), studying if it may impact the probability of receiving these distinguished academic awards. The third model combines the first two models and includes all available data, including citation counts, co-authorship network centrality and citation network centrality.

We used the following logistic regression models to classify a given scholar as a recipient of an award:

1. Model 1, which uses only the sum of citations (X_{cit}.) at year j (1 to 5).

$$ln\frac{p(y)}{1-p(y)} = \beta_0 + \sum_j \beta_{cit_j} X_{cit_j} \tag{2}$$

2. Model 2, which uses only co-authorship centrality measures ($X_{conet-BC}$., $X_{conet-CC}$, $X_{conet-PR}$., $X_{conet-CL}$).

$$ln\frac{p(y)}{1-p(y)} = \beta_0 + \sum_j (\beta_{conet-BC_j} X_{conet-BC_j} + \beta_{conet-CC_j} X_{conet-CC_j} + \beta_{conet-PR_j} X_{conet-PR_j} + \beta_{conet-CL_j} X_{conet-CL_j}) \tag{3}$$

3. Model 3, which combines the sum of citations (X_{cit}) and co-authorship centrality measures ($X_{conet-BC}$, $X_{conet-CC}$, $X_{conet-PR}$., $X_{conet-CL}$).

$$ln\frac{p(y)}{1-p(y)} = \beta_0 + \sum_j (\beta_{cit_j} X_{cit_j} + \beta_{conet-BC_j} X_{conet-BC_j} + \beta_{conet-CC_j} X_{conet-CC_j} + \beta_{conet-PR_j} X_{conet-PR_j} + \beta_{conet-CL_j} X_{conet-CL_j}) \tag{4}$$

4. Model 4, which uses the average centrality of the scholars' papers in the first five years (X_{cit_net-BC},X_{cit_net-CC},X_{cit_net-PR}, X_{cit_net-CL})

$$ln\frac{p(y)}{1-p(y)} = \beta_0 + \sum_j (\beta_{cit\_net-BC_j} X_{cit\_net-BC_j} + \beta_{cit\_net-CC_j} X_{cit\_net-CC_j} + \beta_{cit\_net-PR_j} X_{cit\_net-PR_j} + \beta_{cit\_net-CL_j} X_{cit\_net-CL_j}) \tag{5}$$

5. Model 5 combines all the data, the sum of citations, co-authorship centrality and the average citation network centrality of the scholars' papers.

$$\begin{aligned} ln\frac{p(y)}{1-p(y)} = \beta_0 + \sum_j (&\beta_{cit_j} X_{cit_j} + \beta_{conet-BC_j} X_{conet-BC_j} + \beta_{conet-CC_j} X_{conet-CC_j} \\ &+ \beta_{conet-PR_j} X_{conet-PR_j} + \beta_{conet-CL_j} X_{conet-CL_j} + \beta_{cit\_net-BC_j} X_{cit\_net-BC_j} \\ &+ \beta_{cit\_net-CC_j} X_{cit\_net-CC_j} + \beta_{cit\_net-PR_j} X_{cit\_net-PR_j} + \beta_{cit\_net-CL_j} X_{cit\_net-CL_j} \end{aligned} \tag{6}$$

We collected data on scholars who were announced as recipients of these awards from 1999 (the first year of the earliest award—the AIS Fellows award) to 2013. We omitted scholars for whom we had no publication data or when the author's name was ambiguous. Our data set includes 250 recipients of the awards and over 130,000 who did not receive these awards. We randomly selected 70 percent of the data set as a training set and validated the prediction models on the rest of the data set.

We used Receiver Operating Characteristic (ROC) curve to compare the performance of the five models. We also calculated the area under the ROC curve (AUC), which is equivalent to the probability that the method will rank a randomly chosen positive instance higher than a randomly chosen negative instance.

Our results in Figure 3 demonstrate that adding co-authorship network centrality and citation network centrality of the first five years of scholars' careers performs better than the baseline model, which uses only the citation count. We find that adding the centrality measures increases the AUC by 5 percent (0.8 for the citations count only and 0.85 for the citations + co-authorship + citation network centrality).

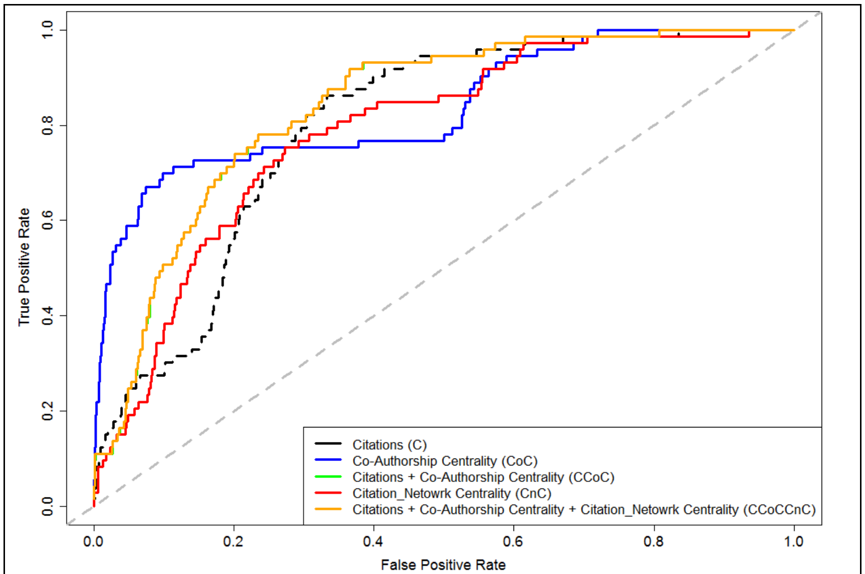

Figure 3. Receiver operating characteristic (ROC) curves of predictions of distinguished awards' recipients based on scholars' early career data

Conclusion and Future Directions

In academia, some of the most important decisions facing personnel and funding committees concern young researchers. These decisions have important implications on science, business and education. Nevertheless, the decision-making process is mainly based on the subjective evaluation of experts.

We present methods to predict a paper's future impact and an author's future impact in the field of management, information systems and operations research, using centrality data of both the citation network and co-authorship network. In particular, we focus on data available at the time of publication for papers' impact predictions and data from the first five years of one's career to predict her future academic impact.

An analysis of the citation network and the centrality of different papers in these networks revealed interesting patterns. Highly cited papers are more central in the citation network from the time of publication. We find that these different structural properties are associated with an increase in the odds ratio of a paper being at the top percentile in the number of citations.

Looking at three distinguished career awards, we find that using scholars' co-authorship centrality and aggregated centrality of her papers performed better than predictions when using only citations.

These results support our argument that improving quantitative methods can complement the qualitative decision-making process in academia.

We will extend the research to include additional variables that can be analyzed from published papers, including content analysis, an author's affiliations, author ordering, journal ranking, etc. An additional set of variables will include close to real time measures, such as search volumes and social media appearance and influence as part of the prediction model.

The overall vision for this project is to create an academic dashboard that will include a suite of measures and prediction methods that could supplement the current subjective tools. In accordance with findings in other business areas, our conjecture is that the use of a data-driven process in academic decisions would yield better predictions of future scholars' achievements.

References

Acuna, D. E., Allesina, S., and Kording, K. P. 2012. "Predicting scientific success," *Nature* (489:7415), pp 201-202.

Adams, J. 2005. "Early citation counts correlate with accumulated impact," *Scientometrics* (63:3), pp 567-581.

Barabási A.L. 2012. "Network Science Book Project," http://barabasilab.neu.edu/networksciencebook/

Bornmann, L., Mutz, R., and Daniel, H. D. 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), pp 830-837.

Brin, S., and Page, L. 1998. "The anatomy of a large-scale hypertextual Web search engine," *Computer networks and ISDN systems* (30:1), pp 107-117.

Brody, T., Harnad, S., and Carr, L. 2006. "Earlier web usage statistics as predictors of later citation impact," *Journal of the American Society for Information Science and Technology* (57:8), pp 1060-1072.

Brynjolfsson, E., Hitt, L., and Kim, H. 2011. "Strength in Numbers: How does data-driven decision-making affect firm performance?,").

Burt, R. S. 2005. *Brokerage and Closure: An Introduction to Social Capital: An Introduction to Social Capital*, (OUP Oxford.

Burt, R. S. 2009. *Structural holes: The social structure of competition*, (Harvard University Press.

Davenport, T. H., and Harris, J. G. 2007. *Competing on analytics: the new science of winning*, (Harvard Business School Press.

Dean, D. L., Lowry, P. Benjamin and Humpherys, S.(2011) "Profiling the research productivity of tenured information systems faculty at U.S. Institutions". *MIS Quarterly*, (30: 1) pp. 1-15.

Egghe, L. 2006. "Theory and practise of the g-index," *Scientometrics* (69:1), pp 131-152.

Eysenbach, G. 2011. "Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact," *Journal of medical Internet research* (13:4).

Freeman, L. C. 1977. "A set of measures of centrality based on betweenness," *Sociometry*), pp 35-41.

Fu, L. D., and Aliferis, C. 2008. "Models for predicting and explaining citation count of biomedical articles," AMIA Annual Symposium Proceedings, American Medical Informatics

Association, 2008, p. 222.

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

Gehrke, J., Ginsparg, P., and Kleinberg, J. 2003. "Overview of the 2003 KDD Cup," *ACM SIGKDD Explorations Newsletter* (5:2), pp 149-151.

Harzing, A.-W. 2012. "Reflections on the h-index," *Business & Leadership* (1:9), pp 101-106.

Hirsch, J. E. 2005. "An index to quantify an individual's scientific research output," *Proceedings of the National academy of Sciences of the United States of America* (102:46), p 16569.

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

Hönekopp, J., and Kleber, J. 2008. "Sometimes the impact factor outshines the H index," *Retrovirology* (5:1), p 88.

Ibáñez, A., Larrañaga, P., and Bielza, C. 2009. "Predicting citation count of Bioinformatics papers within four years of publication," *Bioinformatics* (25:24), pp 3303-3309.

Katona, Z., Zubcsek, P. P., and Sarvary, M. 2011. "Network effects and personal influences: The diffusion of an online social network," *Journal of Marketing Research* (48:3), pp 425-443.

Livne, A., Adar, E., Teevan, J., and Dumais, S. "Predicting Citation Counts Using Text and Graph Mining,").

Lokker, C., McKibbon, K., McKinlay, R. J., Wilczynski, N. L., and Haynes, R. B. 2008. "Prediction of citation counts for clinical articles at two years using data available within three weeks of publication: retrospective cohort study," *BMJ* (336:7645), pp 655-657.

Mayzlin, D. 2002. "The influence of social networks on the effectiveness of promotional strategies," *Yale School of Management Working Paper*).

Newman, M. E. J. 2003. "The structure and function of complex networks," *SIAM review* (45:2), pp 167-256.

Newman, M. E. J. 2004. "Coauthorship networks and patterns of scientific collaboration," *Proceedings of the National Academy of Sciences of the United States of America* (101) APR 6 2004, pp 5200-5205.

Podsakoff, P. M., MacKenzie, S. B., Podsakoff, N. P., and Bachrach, D. G. 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* (34:4), pp 641-720.

Shibata, N., Kajikawa, Y., and Matsushima, K. 2007. "Topological analysis of citation networks to discover the future core articles," *Journal of the American Society for Information*

Science and Technology (58:6), pp 872-882.

Valente, T. W. 1996. "Network models of the diffusion of innovations," *Computational & Mathematical Organization Theory* (2:2), pp 163-164.

Venkatesh V. 2009. "Rankings Based on AIS Senior Scholars' Basket of Journals". Available at: http://www.vvenkatesh.com/ISranking/Credits.asp

Wasserman, S., and Faust, K. 1994. *Social network analysis: Methods and applications*, (Cambridge university press.

Watts, D. J., and Strogatz, S. H. 1998. "Collective dynamics of 'small-world'networks," *nature* (393:6684), pp 440-442.

Yogatama, D., Heilman, M., O'Connor, B., Dyer, C., Routledge, B. R., and Smith, N. A. Year. "Predicting a scientific community's response to an article," Proceedings of the Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics2011, pp. 594-604.