THE UNIVERSITY OF CHICAGO

Department of Statistics

STATISTICS COLLOQUIUM

JAMES A. EVANS

Department of Sociology
University of Chicago

"Modeling Discovery in Science and Technology"

MONDAY, November 28, 2016, at 4:00 PM
Eckhart 133, 5734 S. University Avenue
Refreshments following the seminar in Jones 111

ABSTRACT

In this talk I explore how computationally modeling and predicting scientific discovery and technological invention could create opportunities to improve them. I briefly review our prior work that analyzes this process through simple probabilistic models drawing on the position of scientific concepts such as problems, methods and physical entities within a graph inscribed by their co-presence in research papers and patents to predict successful future combinations. Then I describe a high-dimensional, stochastic (i.e., mixed membership) block model we use to characterize the composition of articles and patents, and a Hidden Markov Process over the time sequence of such models to predict the contents of future articles and patents. We use the improbability of new articles and patents in these models to predict success in terms of citations and awards. We find that taking the high-order combinatorial structure of articles and patents into account doubles our ability to predict their ultimate success, beyond modeling them as simple pairwise combinations. Moreover, we find that when we additionally model articles as combinations of the contexts they cite as well as the content they employ, we can predict 50% more variation in their ultimate success. Authors intensively cite concepts familiar to their audiences rather than those new to them, subtly weaving novelty into tradition, unlike inventors of patents who emphasize novelty. Nevertheless, novel combinations of cited contexts increase the likelihood that new concept combinations will be successful, suggesting that they have not previously been imagined or attempted by others. Together, these findings trace a scientific process of "humble innovation" where scientists are successful by searching broadly, but appearing to build on the shoulders of their audience.
