J PROD INNOV MANAG 2015;••(••):••–••

DOI: 10.1111/jpim.12294

A New Method for Identifying Recombinations of Existing Knowledge Associated with High-Impact Innovation*

Satyam Mukherjee, Brian Uzzi, Ben Jones, and Michael Stringer

How existing technologies and ideas are recombined into new innovations remains an important question, particularly as the store of prior technology, art, and work expands at an increasing rate. Yet, methodologies for identifying effective recombinations remain a nascent area of research. This paper extends our previous work, which developed a network methodology for assessing a scientific article's recombinations of prior work. The methodology uses information from the entire co-citation network of all papers recorded in the Web of Science to identify combinations of prior work that are conventional or atypical and then identifies the virtuous mix of conventional and atypical pairings associated with high impact work. Here, we summarize our prior method and findings, present new findings, and perform a case study application to the field of management science. First, the results show that despite an ever-increasing frontier of possible new combinations of prior work, atypical combinations of prior work are becoming increasingly rare with time, while the distribution of conventional pairings is increasing with time. Second, our analyses show that with time the atypical pairings found in hit papers have a relatively stable mean rate at which they become conventional pairing. Nevertheless, the variance around the mean is growing significantly, which indicates that there is a greater tendency over time for novel pairings either to be virtually never used again or to become conventional pairings.

Practitioner Points

• High-impact work is sharply elevated when combinations of prior work are anchored in substantial conventionality while mixing in a left tail of combinations that are rarely seen together.
• In business, companies and analysts could better predict the value of a future patent or product based on its combination of high conventionality and infusions of novelty.
• Consumer products like computer games that are able to incorporate conventional pairings with a small component of novelty may be more likely to be popular.

Introduction

Invention is spurred on when existing innovations or their components are assembled into original new designs (Becker, 1982; Guimera, Uzzi, Spiro, and Amaral, 2005; Jones, 2009; Jones, Wuchty, and Uzzi, 2008; Schilling, 2005; Schumpeter, 1939; Usher, 1954; Uzzi and Spiro, 2005; Weitzman, 1998). Nevertheless, it remains a mystery as to what recombined elements of existing innovations or components are likely to be promising and attractive to their intended audiences. In some creative fields, it is often lamented that it is impossible to know which recombinations will be fruitful (Committee on Facilitating Interdisciplinary Research, 2004). In filmmaking, it is accepted wisdom that "all hits are flukes," an idiom echoing Samuel Goldwin's quip that "nobody knows nothing" about what combination of previously used, or altogether new plot details, actors, settings, or direct references to past movies makes for a hit (Spitz and Horvát, 2014a, 2014b). As a consequence, many expensive failures must be suffered in the hope that one hit can recover the costs of experimentation. To some degree, similar uncertainty lies behind innovation in general (Börner et al., 2011; Evans and Foster, 2011; Fiore, 2008; Stokols, Hall, Taylor, and Moser, 2008; Wuchty, Jones, and Uzzi, 2007).

Nevertheless, innovative ideas can be difficult to absorb (Henderson and Clark, 1990) and communicate as well, leading creatives to intentionally display conventionality (familiar combinations used many times in previous work) side-by-side with atypical combinations. In his *Principia*, Newton presented his laws of gravitation using accepted geometry rather than his newly developed calculus despite the latter's importance in developing his

Address correspondence to: Satyam Mukherjee, Kellogg School of Management and Northwestern Institute on Complex Systems (NICO), 600 Foster Street, Evanston, IL 60208. E-mail: satyam.mukherjee@gmail.com. Tel: 847-491-2527.
* Sponsored by the Northwestern University Institute on Complex Systems (NICO), by the Army Research Laboratory under Cooperative Agreement Number W911NF-09-2-0053, and DARPA BAA-11-64, Social Media in Strategic Communication. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government.

insights (Whiteside, 1970). Similarly, Darwin devoted the first part of *On the Origin of Species* to conventional, well-accepted knowledge of the selective breeding of dogs, cattle, and birds. In commercial product design, similar dynamics between conventionality and novelty may occur. For example, e-books retain page-flipping graphics to remind the reader of physical books, and blue jeans were designed with a familiar watch pocket to look like conventional trousers. High gross movies like *Avatar* mix conventional storylines with a novel setup of computer graphics.

These patterns suggest that novelty and conventionality are not necessarily factors in opposition where one hand takes from the other; rather, innovations that mix novelty with conventionality may produce enough added value over the current state of the art while providing enough accepted knowledge to allow easy adoption.

BIOGRAPHICAL SKETCHES

Dr. Satyam Mukherjee is a postdoctoral fellow at the Kellogg School of Management and Northwestern Institute on Complex Systems (NICO). He obtained his doctorate in physics from the Indian Institute of Technology Madras, and his research interests include physics of complex systems, social network analysis, and sports analytics. Dr. Mukherjee's works have appeared in *Science*, *Scientific Reports*, *Interface,* and *Physical Review E* and have been featured in *The Economist*, *Times of India,* and *Wall Street Journal*.

Dr. Brian Uzzi is a globally recognized scientist, teacher, consultant, and speaker on leadership, social networks, and new media. He is the Richard L. Thomas Professor of Leadership and Organizational Change at the Kellogg School of Management, codirector of Northwestern Institute on Complex Systems (NICO), and faculty director of the Kellogg Architectures of Collaboration Initiative (KACI). He also holds professorships in Sociology at the Weinberg College of Arts of Sciences and in Industrial Engineering and Management Sciences at the McCormick School of Engineering. Dr. Uzzi has a Ph.D. in Sociology and M.S. from the State University of New York, Stony Brook. He also holds an M.S. in Organizational Psychology from Carnegie Mellon University and a B.A. in Business Economics from Hofstra University.

Dr. Ben Jones is the Gordon and Llura Gund Family Professor of Entrepreneurship, the professor of strategy, and the faculty director of the Kellogg Innovation and Entrepreneurship Initiative. He holds a Ph.D. from Massachusetts Institute of Technology. His research focuses largely on innovation and creativity, with recent work investigating the role of teamwork in innovation and the relationship between age and invention. His research has appeared in journals such as *Science, Quarterly Journal of Economics,* and *American Economic Review*, and has been profiled in media outlets such as the *Wall Street Journal,* the *Economist,* and *The New Yorker*.

Dr. Michael Stringer is co-founder and partner of Datascope Analytics, a Chicago data-driven consulting and design firm. Dr. Stringer received his B.S. in engineering physics from the University of Colorado and a Ph.D. in physics from Northwestern University. At Datascope Analytics, he has led and contributed to projects across a variety of industries for clients such as Procter & Gamble, Thomson Reuters, and other well-known companies.

In our prior work, a network-based methodology was developed for understanding innovation in science. Our focus in this paper is to examine the diversity of the conventional and atypical pairings of the content of the innovation. In scientific work, conventional and atypical pairing of the content of the innovation includes the knowledge put forth in prior published work and represented by the papers referenced in a publications bibliography. In our study of science, the focus is on "hit" innovations—those papers in the top 5% of the citation distribution for their field. To identify each article's pattern of conventional and atypical combinations of prior work, we developed a network analysis method that exploited the full co-citation network of the entire Web of Science (WoS), which includes 17.9 million papers that cite nearly three quarters of a billion papers over the period 1950–2000. The level of conventional and novel combinations of prior work is measured by coding all 17.9 million plus papers in the Web of Science across all subfields of science.[1] The relative conventionality and novelty of the prior work that an article combines is measured by examining the papers referenced in an article's bibliography (Small, 1973; Stringer, Sales-Pardo, and Amaral, 2010). The pairwise combinations of prior work in each article's bibliography allows us to assess how common or novel any pairwise combination of prior work was. To determine this, a full co-citation network is created and journal pairings are counted, using different journals as a proxy for different areas of knowledge (Itzkovitz, Milo, Kashtan, Ziv, and Alon, 2003; Small, 1973; Stringer et al., 2010). Having determined the observed frequency of each journal pairing, we considered the frequency distribution for each journal pairing that would have occurred by chance across the entire Web of Science by permuting the co-citation network using a variation of the Markov chain Monte Carlo (MCMC) algorithm. The more often a journal pairing occurred compared to chance, the more conventional the pairings of prior work, and the less frequently a journal pairing occurred compared to chance, the more novel the pairing. Further, because the null model algorithm randomizes the detailed article-level citation network structure of the

[1] The Web of Science is the largest known repository on the planet for recorded scientific papers worldwide, circa 1950–2000. We examined approximately 17.9 million scientific publications across 12,000 journals as indexed in the Thomson Reuters Web of Science (WoS) database, 1950–2005 period. According to each journal's subject area, the Institute for Scientific Information further defines three fields and constituent subfields: science and engineering (171 subfields), social sciences (54 subfields), and arts and humanities (27 subfields) with coverage for research publications in science and engineering since 1945, social sciences since 1956, and arts and humanities since 1975. The WoS records papers' citations, number of authors, and citation links to other papers in the database.

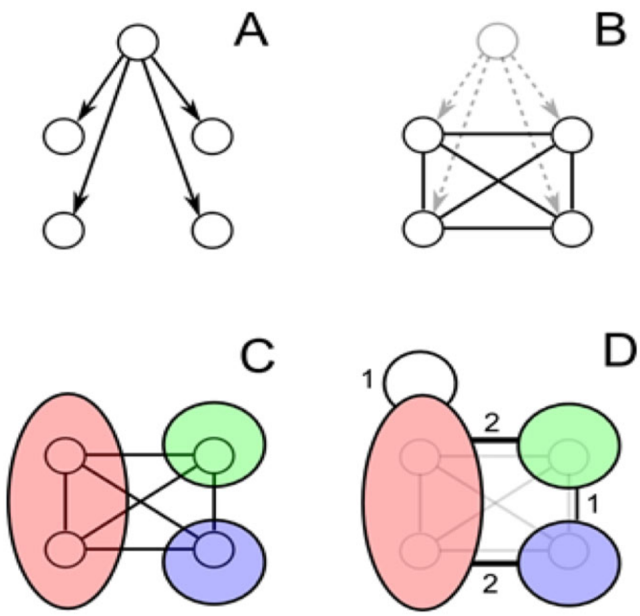

Figure 1. How the Null Model Relates Paper Pairs and Journal Pairs with a Monte Carlo Simulation

global citation network of the Web of Science, it preserves each article's in-degree and out-degree (i.e., backward and forward) citation counts on a constantly updated year-to-year basis, allowing each year of the Web of Science to be compared to every other year on a relative basis (i.e., all comparisons are made to the same third quantity—the null model), despite the growth in the size of prior art and the rate of new publications (Uzzi, Mukherjee, Stringer, and Jones, 2013). Because many innovations are based on elements of prior published work—patents, law, video games, or literature—the method is general.

Figure 1 presents a stylized example of how article pairs and journal pairs are drawn from the network structure of citations. In panel A, the circular nodes represent papers and the directed links exist when the top article cites the bottom four papers. In panel B, the circular nodes represent papers and the undirected co-citation links between papers are shown in black. A co-citation exists between each pair of papers that occurs in the reference list of the focal article. Here, there are four references and therefore six (i.e., four choose two) co-citation links. In panel C, article nodes are grouped by journal; the shaded ovals represent the three journals in which each of the cited papers is published. Finally, in panel D, the co-citation links between papers are mapped to the journal level, and the black links represent journal co-citations. Note that the total number of article-to-article co-citation links (six) is preserved at the journal co-citation level.

Figure 2 shows how citation links between papers are switched randomly but constrained to have the same origin year and target year. Thus in the left panel, switching links A and B is allowed, while switching links A and C is not allowed. The switching algorithm thus preserves for each article its (1) number of references, (2) citation count, (3) citation accumulation dynamics, and 4) the age distribution of referenced work. Performing QE switches (where Q = 10 and E is the number of edges in the graph) converges to a random graph from the configuration

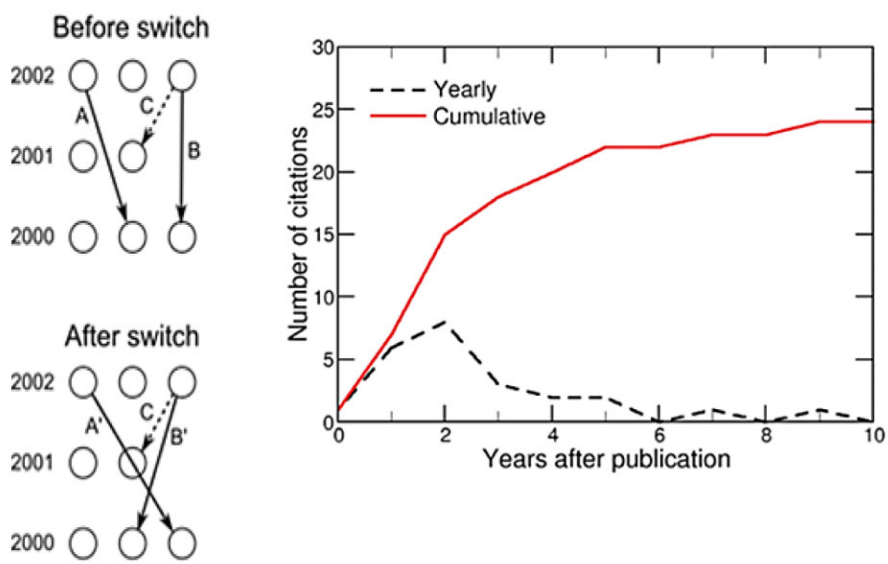

Figure 2. The Model Preserves the Co-Citation Network Backwards and Forwards in Time

model (26) where the number and dynamics of citations are preserved but the origin of the citations is randomized. Since each node is equally likely to be the originating node of any citation, given the constraints, it is known a priori that no disciplines exist in this randomized citation network. The middle panel above demonstrates the citation history of an article—the citation history of every article is exactly preserved under our null model, controlling for both the variation in magnitude and dynamics of citation accumulation to papers.

Comparing the observed frequency with the frequency distribution created with the randomized citation networks, a z-score for each journal pair is generated. This normalized measure describes whether any given pair appeared novel or conventional. Z-scores above zero indicate pairs that appear more often in the observed data than expected by chance, indicating relatively common or "conventional" pairings. Z-scores below zero indicate pairs that appear less often in the observed WoS than expected by chance, indicating relatively atypical or "novel" pairings.

The z-score measure has been used previously in network analysis to determine the degree to which observed linkages among nodes (in our case papers are nodes and links are citations among nodes) are due to chance (Saavedraa, Kathleen, and Uzzi, 2011; Wasserman and Faust, 1994; Watts and Strogatz, 1998). Viewed conceptually, our methodology uses the construct of cut ups—random pairings—to define novel combinations. The technique of "cut ups" was pioneered by poets to drive the process of innovation. In cut ups, a poem or text is written and then randomly cut into strips. The strips are randomly reassembled and new poem or text is written around random connections among the elements of the original text with the intention of creating an innovative, previously unimagined recombination of existing elements.

Our method assigns each article a distribution of journal pair z-scores based on the article's reference list. To characterize an article's tendency to draw together conventional and novel combinations of prior work, two summary statistics were taken. First, the central tendency of an article's combinations is characterized by the article's median z-score. The median allows us to characterize conventionality in the article's main mass of combinations. Second, the article's 10th percentile z-score is considered. The left tail allows us to characterize the article's more unusual journal combinations where novelty may reside.

Figure 3 uses a non-parametric categorization to summarize how papers are classified by their conventionality and novelty. An article is categorized as high or low

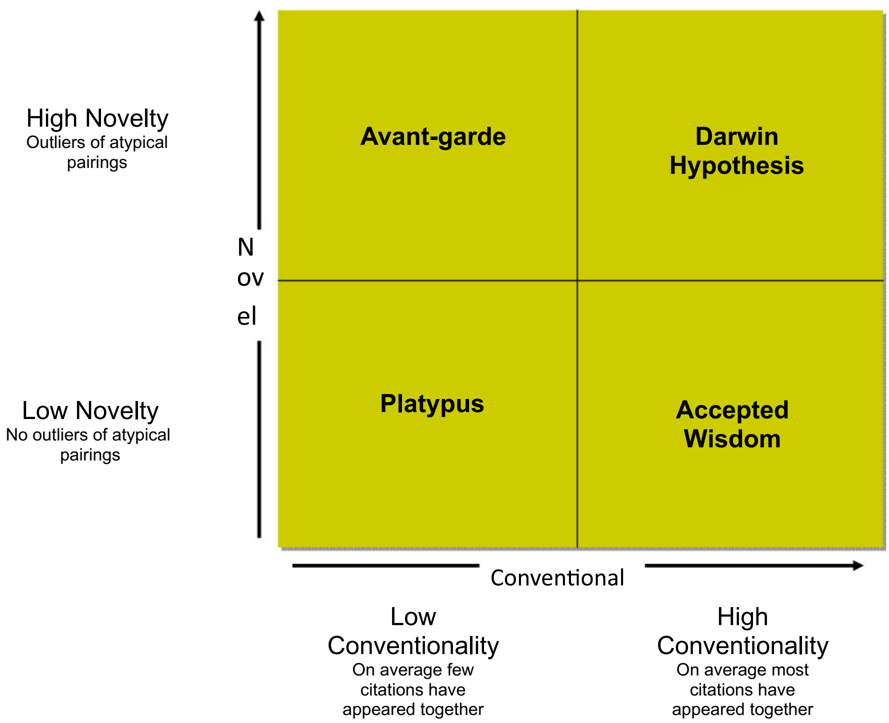

Figure 3. Categorization of Types of Scientific Papers Based on their Combinations of Conventional and Novel Pairings of Prior Published Work

"median conventionality" if its median z score is in the upper or lower half of all median z-scores, respectively, and having high or low "tail novelty" if the article's 10th percentile z-score is above or below zero, respectively. As shorthand, we term papers belonging to high-median conventionality and high tail novelty as *Darwin's Tower,* consistent with his approach to presenting his ideas in the *Origin of Species*. Papers with high median conventionality and low novelty are referred to as *Accepted Wisdom*; papers with high tail novelty and low conventionality are referred to as *Avant-garde*; and papers residing in the low conventionality and low novelty category are termed *Platypuses* because of their neither-nor-like quality.

Case Example for Management

Figure 4 illustrates the method for a single article in management, "A Dynamic Theory of Organizational Knowledge Creation," which proposes a theoretical framework for managing the dynamical aspects of innovation creation process (Nonaka, 1994). For example, our demonstration article (Nonaka, 1994) belongs to the category of high median conventionality and high tail novelty and also accrued 311 citations through the first eight years of publication (Uzzi et al., 2013). This paper cites 88 references and a subset of all possible observed journal frequency pairings of the article in Figure 4 is shown, including, for example, (i) Brown and Duigud *Organization Science* (1991) with Cohen, March and Olsen *Administrative Science Quarterly* (1972), (ii) Brown and Duigud *Organization Science* (1991) with Takeuchi and Nonaka *Harvard Business Review* (1986), and (iii) Takeuchi and Nonaka *Harvard Business Review* (1986) and Prahlad and Hamel *Harvard Business Review* (1990).

Our basic measurement question is to assess how much an article draws on conventional and atypical ideas by examining to what extent each pairwise combination of prior work listed in the reference section of the article is a common or atypical combination. To do this, we computed (1) the observed frequency of any given pairing of references in the WoS and (2) the frequency of a given pairing that would have occurred by chance. Combining (1) and (2) creates a normalized measure to describe whether any given pairing appears novel or conventional. To measure the observed frequency of any given pairing in the WoS, the following five steps are taken: (1) Take the references listed in a given article's bibliography. (2) Consider all pairwise combinations of the papers referenced in the bibliography of the article. (3) For each pairwise combination, record the two journals that were combined. (4) Repeat steps 1 to 3 for every article in the WoS. (5) Count the aggregate frequency of each journal pairing for all referenced pairs for a given publication period (e.g., by year or decade).

Having determined the observed frequency of each journal pairing, a null model is built to consider the

Journal Pairs	Observed	Expected	Z-score
ADMIN SCI QUART - ORGAN SCI	874	1.9	1245.857
HARVARD BUS REV - HARVARD BUS REV	1550	4.1	852.282
CALIF MANAGE REV - ORGAN SCI	95	0.6	118.0
CALIF MANAGE REV - COG PSYCHOL	3	3.599	0.333
Z-score of Zero means observed is as likely as chance			
HARVARD BUS REV - ORGAN SCI	1	1.49925	-0.62
COG PSYCHOL – HARVARD BUS REV	7	11.0063	-1.305

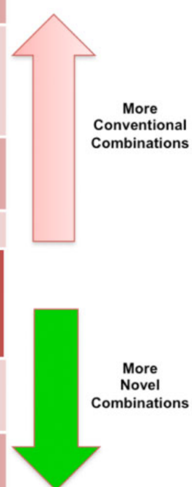

Figure 4. A Sample of the Journal Pair Frequencies, Expected Frequencies, and Z-score for an Illustrative Paper, "A Dynamic Theory of Organizational Knowledge Creation"

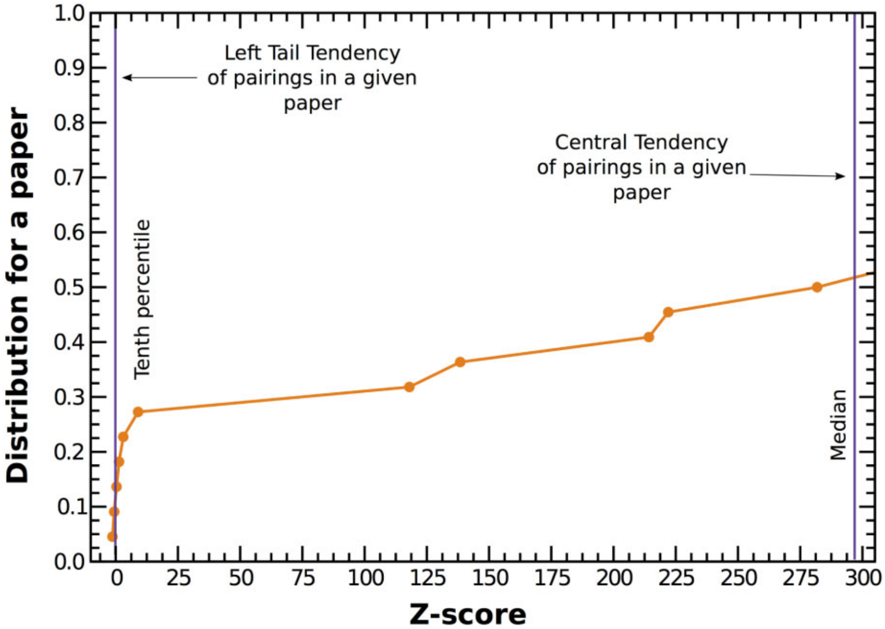

Figure 5. Distribution of Z-Scores of All Journal Pairs for the Paper, "A Dynamic Theory of Organizational Knowledge Creation"

underlying frequency distribution for each journal pairing that would have occurred by chance. The null model randomly reassigns the citation links between papers while preserving (1) the number of citations to and from each article and (2) the citing and cited year for each link using a variation of MCMC algorithm to randomly switch co-citations between papers while preserving the detailed article-level structure of the global citation network with its 17.9 million papers (nodes) and 600 million citations (edges).

Using this algorithm, 10 synthetic instances of the entire WoS are created, each with a different set of randomized citation links. For each instance of the WoS, steps 1 to 5 above are repeated, calculating the frequency of each co-referenced journal pair. Looking across all 10 randomized cases of the WoS, we generate a distribution of frequencies for each journal pair. Next, the z-score for each observed pair relative to what was expected by chance is evaluated as:

$$z = (\mathrm{obs} - \mathrm{exp})/\sigma$$

where *obs* is the observed frequency of the journal pair in the actual WoS, while *exp* is the mean and σ is the standard deviation of the number of journal pairs obtained from the 10 randomized simulations of the article-to-article citation network. Returning to each individual article, a z-score is assigned to each of the journal pairs in that article's reference list. Note that our method examines journal pairings, using journals to proxy for different areas of knowledge. The relative benefits of using journals and randomizations at the article level for defining areas of knowledge relative to other approaches has been demonstrated elsewhere (Uzzi et al., 2013).

Figure 4 displays the observed frequency, expected frequency, and z-score for several journal pairings in the article. Further, the observed frequency of pairings permits a distribution of journal pairings to be created for each article, which represents the collective prior knowledge of the papers in terms of conventionality and atypicality. Figure 5 summarizes the information of the distribution of the z-score for our example article from management. Two primary summary statistics: the median z-score for that article and the 10th percentile z-score are taken for that article. The first measure is a summary statistic that tells us about the central tendency of the combinations of journals that an article cites. The larger the median z-score for an article, the more the journals referenced in the article are combined compared to chance. The second measure is a summary statistic that tells us about the left tail for that article, that is, journal pairings that are relatively unusual among the set of journal pairings an article's reference list makes.

Key Results of the Method for Science in General

The overriding result of our analysis is that there are universal patterns of innovation in science expressed in the relationship between an article's impact and the degree to which it draws on a virtuous mix of conventional and atypical combinations of knowledge. Further, our analysis shows that the virtuous mix appears not to be a problem that is solved by individuals. Relatively speaking, teams show a much higher probability of hitting on the virtuous than do individuals. Moreover, at any mix of conventionality and atypicality, teams do more to assimilate combinations of conventional and atypical ideas in a single article.

First, hit papers are quantified as those papers in the top 5% of the citation distribution. Our findings indicate that papers that display a high level of median conventionality with a mix of high tail novelty (Darwin's Tower) have a likelihood of becoming a hit that is nearly twice that of the expected background rate of 5%. All other combinations have no higher expectations of becoming hits than that of the 5% background rate or have a rate lower than the background rate with the Platypus class of papers displaying a hit rate of just 2% (Uzzi et al., 2013).

Second, regression analysis with controls for other variables known to affect citation counts and fixed effects for year and each distinct field in the WoS shows that the above patterns hold for definitions of hit papers in the top 10%, 5%, and 1% levels of the distribution of citations and across time. Specifically, Figure 6 indicates that high tail novelty mixed with high median conventionality (Darwin's Tower) outperforms other categories in all decades from 1950 to 2000, regardless of whether a "hit" article is defined as the top 1%, 5%, or 10% by citations received. In all cases, Darwin's Tower sees hit rates approximately twice the background rate.

Third, Table 1 shows that the pattern of innovation found for high median conventionality and high tail novelty holds when regressions are run on a field-by-field basis. For each of the 252 fields in the WoS, the papers are categorized by ranking them according to their probability of producing hit papers. This analysis revealed that high tail novelty and high median conventionality (Darwin's Tower) are the highest impact papers in 64.38% of fields and either first or second in 85% of fields. And while the effect holds for 64.38% of the fields, it holds for the vast majority of all papers because it holds for the fields with the largest number of scientists in them. By contrast, low tail novelty and low median conventionality are the lowest or second lowest in about 87 of fields (Platypus) (see Uzzi et al., 2013, for additional details).

Fourth, our work shows that the process of finding the right balance between conventional and novel pairings favors teamwork over solo work. Solo authors and teams commonly write papers with the same level of median conventionality but teams are more likely to have left tail novelty than solo authors. Further, it is observed that at any combination of levels of conventionality and novelty, team papers are more likely to be hits than are solo papers. For example, taking a team and a solo article whose median levels of conventionality are at the 90 percentile of the median conventionality distribution and have high left tail novelty, the team article, on average, has nearly twice the probability of a hit. This indicates that teams not only reach for novelty more but appear to assimilate it better with conventional knowledge.

Here, as a further extension of the findings of Uzzi et al. (2013), we examined two dynamic processes related to innovation: (1) how the level of conventionality and the level of novelty in science are changing over time and (2) whether novel pairings are the seeds of future conventional pairings.

As mentioned above, scientific innovations draw heavily on conventional combinations of the past. Our analysis of the change in conventionality over time shows that science is becoming significantly more conventional. Figure 7A displays the cumulative distribution of median z-scores across all papers in the WoS published in the 1950s through 1990s by decade. This plot shows that there is a high propensity for conventionality in all years. Further, with respect to time, conventionality is systematically and increasingly dominant in scientific papers. In all decades, the cumulative distributions are moving significantly to the right. By the 1980s and 1990s, fewer than 4% of papers have median z-scores below 0 and more than 50% of papers have median z-scores above 64. Figure 7B considers the 10th percentile z-scores, which further suggest a propensity for conventionality. In all decades, the curves are systematically falling. With each decade that goes by, the level of novelty in scientific papers drops systematically. Only 41% of papers in the 1980s and 1990s have a 10th percentile z-score below 0. Overall, by these measures, science is increasingly becoming conventional and more rarely drawing on atypical pairings of prior work even as novel combinations remain critical to hits and the percentage of science done in teams has increased.

In regard to the changing distributions of conventionality and novelty for the field of management as defined in the WOS, our findings indicate that management reflects the pattern of most sciences. Figure 8A displays the cumulative distribution of median z-scores across all papers in the WoS published in the 1970s through 1990s by decade. Considering that a z-score below zero represents a journal pair that occurs less often than expected by chance, the analysis of median z-scores suggests very high propensity of conventionality in all decades. Further,

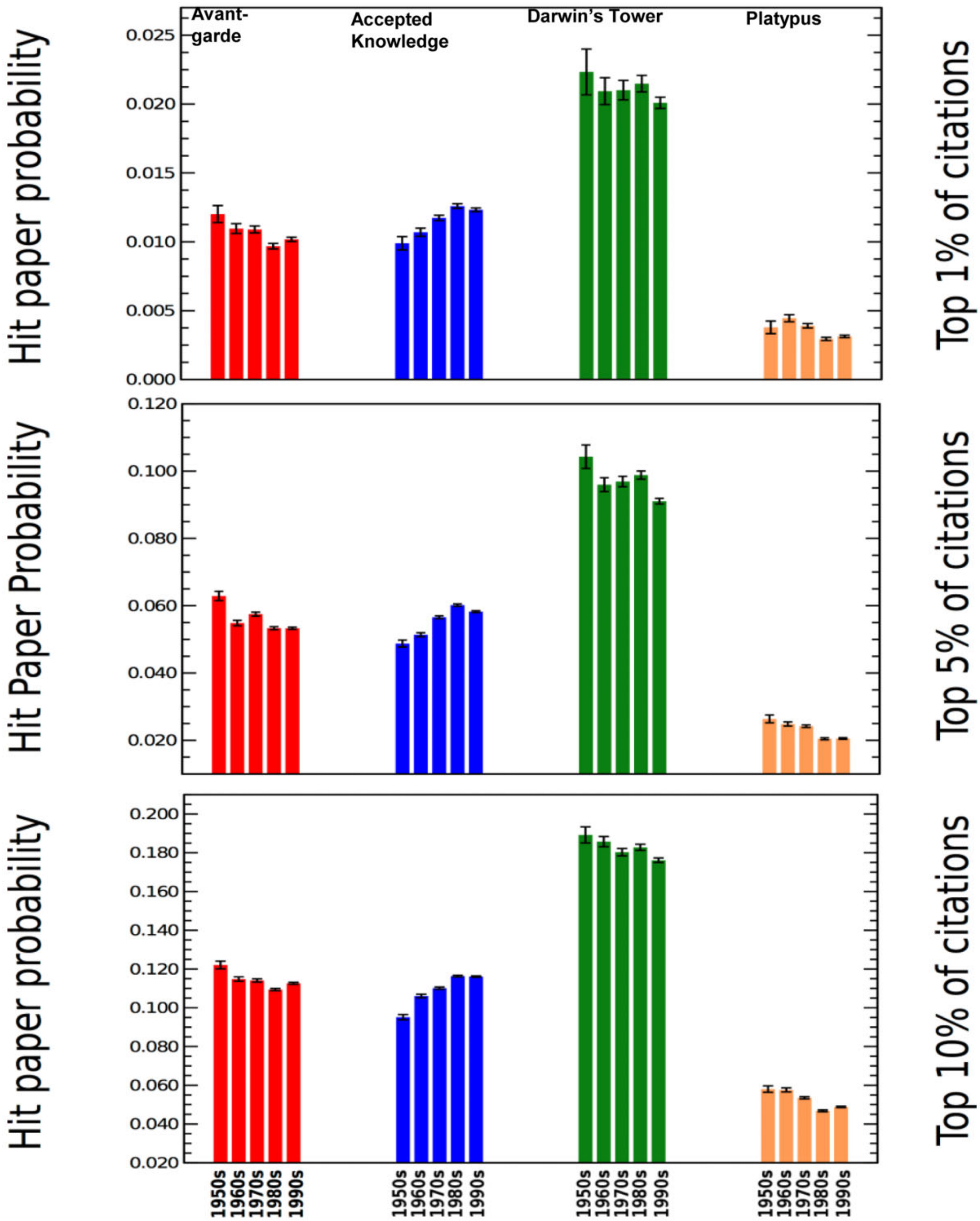

Figure 6. Fixed Effects Regressions for Probability of Impact Defined Multiple Ways and Controlling for Time and Field Fixed Effects Indicate a Robust Relationship between High Median Conventionality and High Tail Novelty in Scientific Papers

with respect to time, conventionality is systematically and increasingly dominant. In all decades, the cumulative distributions are moving significantly to the right. By the 1980s and 1990s, fewer than 5% of papers have median z-scores below 0 and more than 50% of papers have median z-scores above 32. In the 1990s, only 4.9% of papers had median z-scores below 0, while in the 1990s the percentage fell to 1.7%, indicating a persistent and prominent tendency for high conventionality. Figure 8B considers the 10th percentile z-scores, which further

Table 1. Darwin's Tower Dominates Innovation in Science as Measured on a Strict Field-by-Field Basis

	Rank			
	1st	2nd	3rd	4th
Avant-garde High tail novelty and low median conventionality	20.25%	44.53%	28.74%	6.48%
Accepted knowledge Low tail novelty and high median conventionality	9.71%	26.72%	50.61%	12.96%
Darwin's Tower High tail novelty and high median conventionality	64.38%	21.86%	3.64%	10.12%
Platypus Low tail novelty and low median conventionality	5.66%	6.89%	17.01%	70.44%

suggest a propensity for conventionality. In all decades, the curves are systematically falling. With each decade that goes by, the level of novelty in management papers drops. Only 19.6% of the papers in the 1970s, 17.9% of the papers in the 1980s, and 16.8% in the 1990s have a 10th percentile z-score below zero.

Finally, the sources of conventionality are investigated by examining whether the roots of conventional pairings originated in the novel pairings of high impact papers. Statistical power is gained by looking at all the fields of science simultaneously. First, the top 5% of highly cited papers published in 1980 is considered. We then examine whether the 10th percentile journal pairs in these papers, a combination known to be novel at the time, become increasingly conventional with time. Figure 9 quantifies the mean of the z-scores with time. The mean z-score of novel pairings found in hit papers in 1980 and followed through to 2000 show that novel pairings on average remain novel up to 20 years later. This suggests that novel pairings are not the seeds of conventional pairings but remain conventional. Nevertheless, it is observed that there is a significant increase in the variance around which novel pairings become conventional pairings in future. This suggests that over time, novel pairings are more likely to become conventional pairings. For example, in 1980, all the variance around the mean is negative (which is expected, as this is the first year the novel pairings appear). As time moves forward, novel pairings have an increased chance of becoming conventional pairings (i.e., there are now increased numbers of pairings over the mean and into the positive range of the z-score). While this evidence does not conclusively support or disprove this exploratory analysis, the prima facie evidence suggests that the mean is fairly stable but that some novel pairings become conventional, suggesting that these novel pairings hit on particularly virtuous combinations. Knowing, at the time, which combinations were virtuous was likely impossible given the broad variance in future outcomes among all novel pairings.

Discussion

Our analysis of 17.9 million papers across all scientific fields has implications for theories of innovation. Combinations of existing material are centerpieces in theories of creativity, whether in the arts, the sciences, or commercial innovation (Becker, 1982; Collins, 1998; Guimera et al., 2005; Jones et al., 2008; Schilling, 2005; Schumpeter, 1939; Usher, 1954; Uzzi and Spiro, 2005; Weitzman, 1998). Within the tradition of examining how recombinations of work may impact innovation, two lines of research have arisen. One line focuses on the ability of teams of diverse individuals to bring together novel combinations of preexisting innovations or their elements. In this view, each team member overcomes their tendency to narrowly view the range of possible recombinations that have been defined in their area of expertise or experience. If the members of the team have uncorrelated expertise or experience, chances are that unusual and new product designs could emerge (Page, 2007). Building on this approach is the idea that new ideas need to be embedded in a framework of accepted conventions, which increase the audience's ability to fluently evaluate innovativeness in an ever-expanding sea of knowledge (Einstein, 1949; Fleming, 2001; Jones, 2009). Yet, it remains unknown as to what mix of conventional and atypical ideas results in a virtuous balance of convention and invention, as well as how to measure convention and invention through time. Our method provides a means for measuring the degree to which a pairing of elements in an innovation are conventional or innovative and what mix of convention and atypical knowledge within an innovation is associated with high impact.

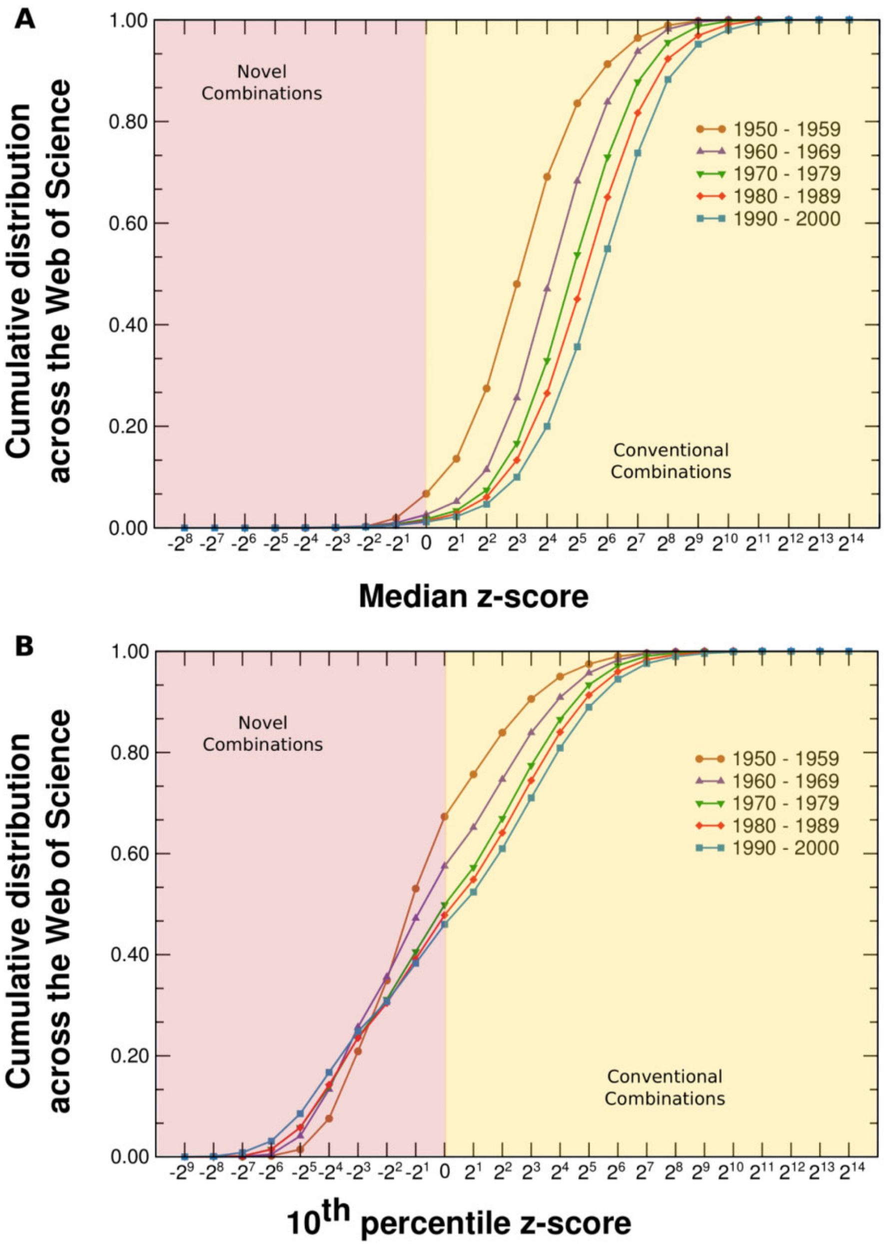

Figure 7. Temporal Evolution of Conventionality and Novelty for All Disciplines in Web of Science

Our results show that the propensity for high impact work is sharply elevated when combinations of prior work are anchored in substantial conventionality, while mixing in a left tail of combinations that are rarely seen together. Further, a link between the mix of conventional and atypical knowledge and teams is established. We find that teams are more likely to reach for novel combinations than are individuals, a fact that reinforces prior case study and lab research. Moreover, it is observed that at any mix of conventionality and novelty, teams do more with the mixture, i.e., they produce higher impact science at any mix of conventional and atypical knowledge. This suggests that teams are better at reaching for novel combinations and better at assimilating conventional ideas and novel ideas into new innovations.

We believe that our methods and conceptual approach can be applied to other domains. In business, links between novelty and conventionality in successful patents is a natural extension. Patents also draw on prior work and represent that prior work through citation patterns to prior art. If this is the case, then companies and analysts may be able to better predict the value of a future

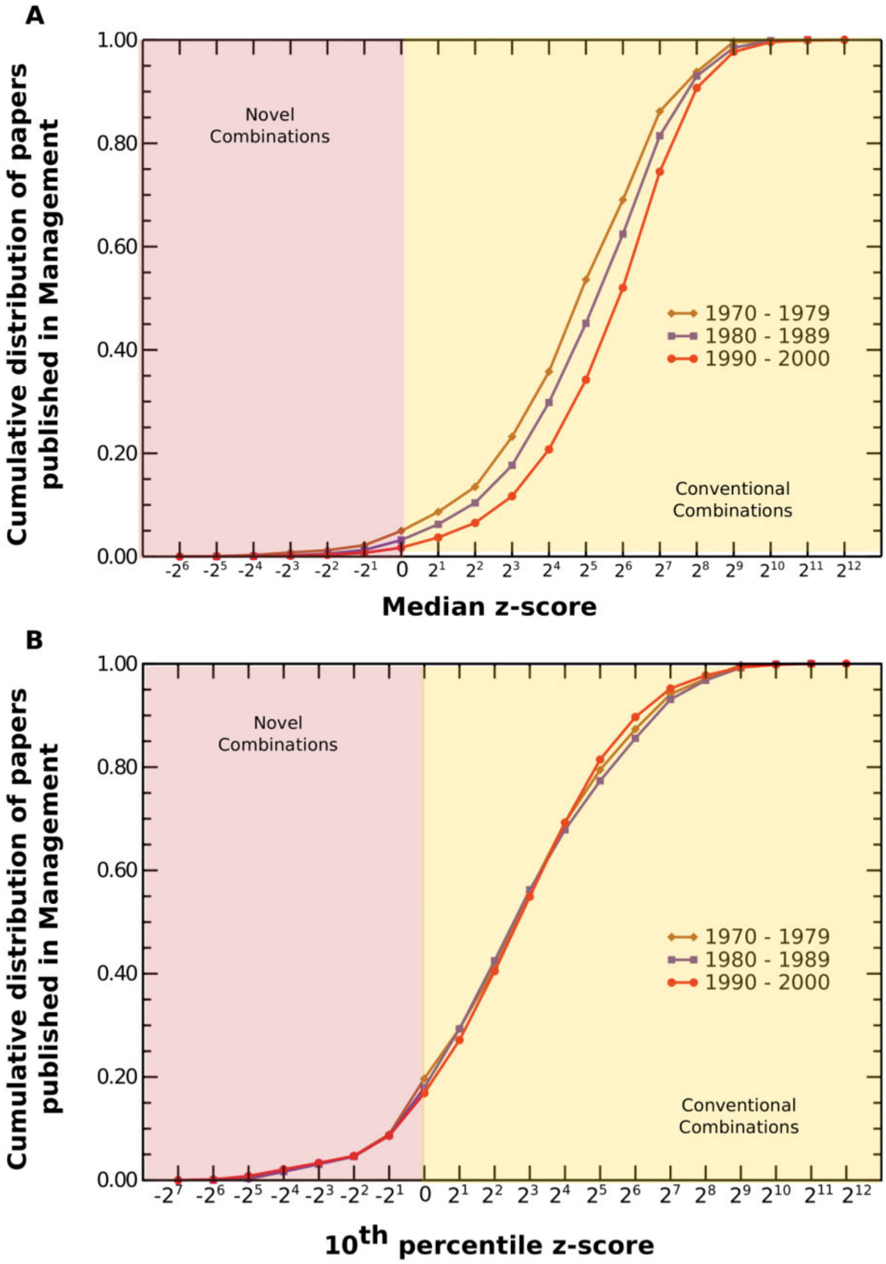

Figure 8. Temporal Evolution of Conventionality and Novelty in Management

patent based on its combination of high conventionality and infusions of novelty. Similarly, consumer product design might also benefit. Consider the development of new games or apps. In each case, there are established categories of story lines, settings, technical capabilities, and so on. For example, a computer game could be first-person or second-person shooter; online, computer specific, or both; any sports, racing, or fighting game; and within each of these categories there may be finer distinctions. In each case, these finer distinctions may provide a basis for quantifying what are conventional pairings of categories and novel pairings of categories in consumers' minds based on the frequency of pairings in prior games compared to chance. For example, those games that are able to incorporate conventional pairings with a small component of novelty may be more likely to be hits. In this way, the methodology and approach could be conceivably extended to diverse domains in technology, medicine, or consumer products. At root, our work suggests that creativity appears to be a nearly universal phenomenon of two extremes. At one extreme is conventionality and at the other is novelty. Curiously,

Figure 9. Z-Score of Novel Journal Pairs Found in High-Impact Papers Published in 1980 and Followed through the Year 2000 Show the Increasing Tendency to Become Conventional Pairings

notable advances appear most closely linked not with efforts along one boundary or the other but with efforts that reach toward both frontiers.

References

Becker, H. S. 1982. *Art worlds*. Oakland, CA: University of California Press.

Börner, K., N. Contractor, H. Falk-Krzesinski, S. Fiore, K. Hall, J. Keyton, T. Klein, B. Spring, D. Stokols, and W. Trochim. 2011. Mapping a research agenda for the science of team science. *Research Evaluation.* 20 (2): 145–58.

Brown, J. S., and P. Duigud. 1991. Organizational learning and communities-of-practice: Toward a unified view of working, learning, and innovation. *Organization Science* 2 (1): 40–57.

Cohen, M. D., J. G. March, and J. P. Olsen. 1972. Garbage can model of organization choice. *Administrative Science Quarterly* 17 (1): 1–25.

Collins, R. 1998. *The sociology of philosophies: A global theory of intellectual change*. Cambridge, MA: Harvard University Press.

Committee on Facilitating Interdisciplinary Research. 2004. *Facilitating interdisciplinary research*. Washington, DC: National Academy of Sciences, National Academy of Engineering, Institute of Medicine.

Einstein, A. 1949. *The world as I see it*. Secaucus, NJ: Citadel Press.

Evans, J., and J. Foster. 2011. Metaknowledge. *Science* 331 (6018): 721–25.

Fiore, S. M. 2008. Interdisciplinarity as teamwork: How the science of teams can inform team science. *Small Group Research* 39 (3): 251–77.

Fleming, L. 2001. Recombinant uncertainty in technological search management. *Science* 47 (1): 117–32.

Guimera, R., B. Uzzi, J. Spiro, and L. A. N. Amaral. 2005. Team assembly mechanisms determine collaboration network structure and team performance. *Science* 308 (5722): 697–702.

Henderson, R. M., and K. B. Clark. 1990. Architectural innovation: The reconfiguration of existing. *Administrative Science Quarterly* 35 (1): 9–30.

Itzkovitz, S., R. Milo, N. Kashtan, G. Ziv, and G. Alon. 2003. Subgraphs in random networks. *Physical Review E* 68: 026127 (1-8).

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

Jones, B., S. Wuchty, and B. Uzzi. 2008. Multi-university research teams: Shifting impact, geography, and stratification in science. *Science* 322 (5905): 1259–62.

Nonaka, I. 1994. A dynamic theory of organizational knowledge creation. *Organization Science* 5 (1): 14–37.

Page, S. E. 2007. *The difference: How the power of diversity creates better groups, firms, schools, and societies*. Princeton, NJ: Princeton University Press.

Prahlad, C., and G. Hamel. 1990. The core competence of the corporation. *Harvard Business Review* 68 (3): 79–91.

Saavedraa, S., H. Kathleen, and B. Uzzi. 2011. Synchronicity, instant messaging, and performance among financial traders. *Proceedings of the National Academy of Sciences* 108 (3): 5296–301.

Schilling, M. 2005. A "small world" network model of creative insight. *Creativity Research Journal* 17 (2–3): 131–54.

Schumpeter, J. 1939. *Business cycles*. New York: McGraw-Hill Book Company, Inc.

Small, H. 1973. Co-citation in scientific literature: A new measure of the relationship between two documents. *Journal of American Society for Information Science* 24 (4): 265–69.

Spitz, A., and E.-Á. Horvát. 2014a. Measuring long-term impact based on network centrality: Unraveling cinematic citations. *PLoS ONE* 9 (10): e108857.

Spitz, A., and E.-Á. Horvát. 2014b. A cookbook of cinematic delicacies that do not expire. *Leonardo* 47 (3): 271.

Stokols, D., K. L. Hall, B. K. Taylor, and R. P. Moser. 2008. The science of team science: Overview of the field and introduction to the supplement. *American Journal of Preventive Medicine* 35 (2): S77–89.

Stringer, M., M. Sales-Pardo, and L. A. N. Amaral. 2010. Statistical validation of a global model for the distribution of the ultimate number of citations accrued by papers published in a scientific journal. *Journal of American Society for Information Science* 61 (7): 1377–85.

Takeuchi, H. I., and I. Nonaka. 1986. The new product development game. *Harvard Business Review* 64 (1): 137–46.

Usher, A. P. 1954. *A history of mechanical invention*. Cambridge, MA: Harvard University Press.

Uzzi, B., and J. Spiro. 2005. Collaboration and creativity: The small world problem. *American Journal of Sociology* 111 (2): 447–504.

Uzzi, B., S. Mukherjee, M. Stringer, and B. Jones. 2013. Atypical combinations and scientific impact. *Science* 342 (6157): 468–72.

Wasserman, S., and K. Faust. 1994. *Social network analysis: Methods and applications (structural analysis in the social sciences)*. Cambridge, UK: Cambridge University Press.

Watts, D. J., and S. H. Strogatz. 1998. Collective dynamics of "small-world" networks. *Nature* 393 (6684): 440–42.

Weitzman, M. L. 1998. Recombinant growth. *Quarterly Journal of Economics* 113 (2): 331–60.

Whiteside, D. T. 1970. The mathematical principles underlying Newton's principia mathematica. *Journal for the History of Astronomy* 1 (2): 116–38.

Wuchty, S., B. Jones, and B. Uzzi. 2007. The increasing dominance of teams in the production of knowledge. *Science* 316 (5827): 1036–39.