

NETWORK NEURO SCIENCE

an open access  journal



Citation: Betzel, R. F., Fukushima, M., w He, Ye, Zuo, Xi-Nian, Sporns, O. (2016) Dynamic fluctuations coincide with periods of high and low modularity in resting-state functional brain networks *Network Neuroscience*, 1

DOI:
<http://dx.doi.org/10.1162/NETN-00001>

Supporting Information:
<http://dx.doi.org/10.7910/DVN/PQ6ILM>

Received: 20 October 2016
Accepted: 7 November 2016
Published: 26 January 2016

Competing Interests: The authors have declared that no competing interests exist.

Corresponding Author:
George Chacko
netelabs@nete.com

Handling Editor:
Xi-Nian Zuo

Copyright: © 2019
Massachusetts Institute of Technology
Published under a Creative Commons
Attribution 4.0



The MIT Press

RESEARCH ARTICLE

Co-citations in context: disciplinary heterogeneity is relevant

James Bradley¹, Sitaram Devarakonda², Avon Davey², Dmitriy Korobskiy², Siyu Liu², Djamil Lakhdar-Hamina², Tandy Warnow³ and George Chacko²

¹Raymond A. Mason School of Business, College of William and Mary, Williamsburg, VA, USA

²Netelabs, NET ESolutions Corporation, McLean, VA 22102, USA

³Department of Computer Science, University of Illinois at Urbana-Champaign, Champaign, IL 61820, USA

Keywords: co-citation analysis, bibliometrics, random graphs

ABSTRACT

Citation analysis of the scientific literature has been used to study and define disciplinary boundaries, to trace the dissemination of knowledge, and to estimate impact. Co-citation, the frequency with which pairs of publications are cited, provides insight into how documents relate to each other and across fields. Co-citation analysis has been used to characterize combinations of prior work as conventional or innovative and to derive features of highly cited publications. Given the organization of science into disciplines, a key question is the sensitivity of such analyses to frame of reference. Our study examines this question using semantically-themed citation networks. We observe that trends reported to be true across the scientific literature do not hold for focused citation networks, and we conclude that co-citation analysis requires a contextual perspective.

INTRODUCTION

Citation and network analysis of scientific literature reveals information on semantic relationships between publications, collaboration between scientists, and the practice of citation itself (de Solla Price, 1965; Garfield, 1955; Newman, 2001; Patience, Patience, Blais, & Bertrand, 2017; Shi, Leskovec, & McFarland, 2010; Stigler, 1994). Co-citation, the frequency with which two documents are cited together in other documents provides additional insights, including the identification of semantically related documents, fields, specializations, and new ideas in science (Boyack & Klavans, 2010; Marshakova-Shaikevich, 1973; Small, 1973; Zuckerman, 2018).

Uzzi, Mukherjee, Stringer, and Jones (2013) used a novel approach for co-citation analysis to characterize a subset of highly cited articles with respect to both novel and conventional combinations of prior research. The frequency with which references were co-cited in 17.9 million articles and their cited references from the Web of Science (WoS) was calculated and expressed as journal pair frequencies (observed co-citation frequencies). Expected co-citation values were generated from randomized networks using Monte Carlo simulations under a random graph model. Observed frequencies were then normalized (shifted and scaled) to averaged expected values from ten randomized networks and termed as *z-scores*. Consequently, every article was associated with multiple *z-scores* corresponding to co-cited journal pairs in its references. For each article, positional statistics of *z-scores* were calcu-

lated and evaluated to set thresholds for a binary classification of conventionality using the median z-score of an article and novelty using the tenth percentile of z-scores within an article.

Thus, LNHC would denote low novelty (LN) and high conventionality (HC), with all four combinations of LN and HN with LC and HC being possible. The authors observed that HNHC articles were twice as likely to be highly cited compared to the background rate, suggesting that novel combinations of ideas flavoring a body of conventional thought were a feature of impact.

Key to Uzzi *et al.* is their random graph model and its underlying assumptions. The citation switching algorithm used to generate expected values by random substitutions of references is designed to preserve the number of publications, the number of references in each publication, and the year of publication of both publications and references. Importantly, disciplinary origin does not affect the probability that a reference is selected to replace another one. For example, a reference in quantum physics can be substituted, with equal probability, by a reference in quantum physics, quantum chemistry, classical literature, entomology, or anthropology. Such substitutions poorly model the disciplinary nature of scientific endeavor and citation behavior (Garfield, 1979; Klavans & Boyack, 2017; Moed, 2010; Wallace, Lariviere, & Gingras, 2012). In addition, under this random model, a reference cited many times in a given year is selected with the same probability as a reference cited only once, which appears inconsistent with the power law or lognormal citation distributions described in the literature (Perline, 2005; Stringer, Sales-Pardo, & Amaral, 2010). Accordingly, model misspecification is likely to arise on account of the simulated values not reflecting the empirical data very well.

A follow-up study by Boyack and Klavans (2014) explored the impact of discipline and journal effects on these definitions of conventionality and novelty. While their study had some methodological differences in the use of Scopus data rather than WoS data, a smaller dataset, and a χ^2 calculation rather than Monte Carlo simulations to generate expected values of journal pairs, Boyack and Klavans noted strong effects from disciplines and journals. While they also reported the trend that HNHC is more probable in highly cited papers, they observed that “only 64.4% of 243 WoS subject categories” in the Uzzi *et al.* study met the criterion of having the highest probability of hit papers in the HNHC category. Further, they observed that journals vary widely in terms of size and influence and that 20 journals accounted for nearly 15% of co-citations in their measurements. Lastly, they noted that three multidisciplinary journals accounted for 9.4% of all atypical combinations.

Despite different methods used to generate expected values, both of these key preceding studies measured co-citation frequencies across the scientific literature (using either WoS or Scopus) and normalized them without disciplinary constraints before subsequently analyzing disciplinary subsets. We hypothesized instead that modifying the normalization to constrain substitution references to be drawn only from the citation network being studied (the “local network”) rather than all of WoS (the “global network”) would reduce model misspecification by limiting substitutions from disciplinarily ectopic references. Consequently, we used keyword searches of the scientific literature to construct exemplar citation networks themed around academic disciplines of interest: *applied physics*, *immunology*, and *metabolism*. Within these disciplinary frameworks, we calculated observed and expected co-citation frequencies using a refined random graph model and an efficient Monte Carlo simulation algorithm.

Our analyses, using multiple techniques, provides substantial evidence that a constrained model where reference substitutions are limited to a local (disciplinary) network reduces model misspecification compared to the unconstrained model that uses the global network (WoS). Furthermore, re-analyses of these three semantically-themed citation networks under the improved model reveals strikingly different trends. For example, while Uzzi *et al.* reported that highly cited articles are more likely than expected to be both HC and HN and that this trend largely held across all disciplines, we find that these trends vary with the discipline so that universal trends are not apparent. Specifically, HC remains highly correlated with highly cited articles in the immunology and metabolism datasets but not with applied physics, and HN is highly correlated with highly cited articles in applied physics but not with immunology and metabolism. Thus, disciplinary networks are different from each other, and trends that hold for the full WoS network do not hold for even large networks (such as metabolism). Furthermore, we also found that the categories demonstrating the highest percentage of highly cited articles (e.g., HC, HN, etc.) are not robust with respect to varying thresholds for high citation counts or for highly novel citation patterns. Overall, our study, although limited to three disciplinary networks, suggests that co-citation analysis that inadequately considers disciplinary differences may not be very useful at detecting universal features of impactful publications.

MATERIALS AND METHODS

Bibliographic data

We have previously developed ERNIE, an open source knowledge platform into which we parse the Web of Science (WoS) Core Collection (Keserci, Davey, Pico, Korobskiy, & Chacko, 2018). WoS data stored in ERNIE spans the period 1900-2019 and consists of over 72 million publications. For this study, we generated an analytical dataset from years 1985 to 2005 using data in ERNIE. The total number of publications in this dataset was just over 25 million publications (25,134,073), which were then stratified by year of publication. For each of these years, we further restricted analysis to publications of type Article. Since WoS data also contains incomplete references or references that point at other indexes, we also considered only those references for which there were complete records (Table 1). For example, WoS data for year 2005 contained 1,753,174 publications, which after restricting to type Article and considering only those references described above resulted in 916,573 publications, 6,095,594 unique references (set of references), and 17,167,347 total references (multiset of references). Given consistent trends in the data (Table 1), we analyzed the two boundary years (1985 and 2005) and the mid-point (1995). We also used the number of times each of these articles was cited in the first 8 years since publication as a measure of its impact.

We constructed three disciplinary datasets in areas of our interest based on the keyword searches: immunology, metabolism, and applied physics. For the first two, rooted in biomedical research, we searched Pubmed for the term ‘immunology’ or ‘metabolism’ in the years 1985, 1995, and 2005 (Table 2). Pubmed IDs (pmids) returned were matched to WoS IDs (wos.ids) and used to retrieve relevant articles. For the applied physics dataset, we directly searched traditional subject labels in WoS for ‘Physics, Applied’. While applied physics and immunology represent somewhat small networks (roughly 2-6% of WoS), metabolism represents approximately 22% of WoS, making them interesting and meaningful test cases. We also examined publications in the five major research areas in the WoS: life sciences & biomedicine, physical sciences, technology, social sciences, and arts & hu-

Table 1. Summary of base WoS Analytical Dataset. Only publications of type Article with at least two references and references with complete publication data were selected for this data set. The number of unique publications of type Article, unique references (ur), total references (tr), and the ratio of total references to unique references increases monotonically with each year indicating that both the number of documents and citation activity increase over time.

| Year | Unique Publications | Unique References (ur) | Total References (tr) | tr/ur |
|------|---------------------|------------------------|-----------------------|-------|
| 1985 | 391,860 | 2,266,584 | 5,588,861 | 2.47 |
| 1986 | 402,309 | 2,316,451 | 5,708,796 | 2.46 |
| 1987 | 412,936 | 2,427,347 | 5,998,513 | 2.47 |
| 1988 | 426,001 | 2,545,647 | 6,354,917 | 2.50 |
| 1989 | 443,144 | 2,673,092 | 6,749,319 | 2.52 |
| 1990 | 458,768 | 2,827,517 | 7,209,413 | 2.55 |
| 1991 | 477,712 | 2,977,784 | 7,729,776 | 2.60 |
| 1992 | 492,181 | 3,134,109 | 8,188,940 | 2.61 |
| 1993 | 504,488 | 3,278,102 | 8,676,583 | 2.65 |
| 1994 | 523,660 | 3,458,072 | 9,255,748 | 2.68 |
| 1995 | 537,160 | 3,680,616 | 9,875,421 | 2.68 |
| 1996 | 663,110 | 4,144,581 | 11,641,286 | 2.81 |
| 1997 | 677,077 | 4,340,733 | 12,135,104 | 2.80 |
| 1998 | 693,531 | 4,573,584 | 12,728,629 | 2.78 |
| 1999 | 709,827 | 4,784,024 | 13,280,828 | 2.78 |
| 2000 | 721,926 | 5,008,842 | 13,810,746 | 2.76 |
| 2001 | 727,816 | 5,203,078 | 14,261,189 | 2.74 |
| 2002 | 747,287 | 5,464,045 | 15,001,390 | 2.75 |
| 2003 | 786,284 | 5,773,756 | 16,024,652 | 2.78 |
| 2004 | 826,834 | 6,095,594 | 17,167,347 | 2.82 |
| 2005 | 886,648 | 6,615,824 | 19,036,324 | 2.88 |

manities, using the extended subcategory classification of 153 sub-groups to categorize disciplinary composition of cited references in the datasets we studied.

Table 2. Disciplinary Datasets. PubMed and WoS were searched for articles using search terms, ‘immunology’, ‘metabolism’, and ‘applied physics.’ Counts of publications are shown for each of the three years analyzed and expressed in parentheses as a percentage of the total number of publications in our analytical WoS dataset (Table 1) for that year. Note that Applied Physics and Immunology each represent about 4% of the publications in WoS, but Metabolism occupies nearly one-fourth of WoS.

| Year | Applied Physics | Immunology | Metabolism |
|------|-----------------|---------------|-----------------|
| 1985 | 10,298 (2.7%) | 21,606 (5.5%) | 78,998 (20.2%) |
| 1995 | 21,012 (3.9%) | 29,320 (5.5%) | 121,247 (22.6%) |
| 2005 | 35,600 (4.0%) | 37,296 (4.2%) | 200,052 (22.6%) |

Monte Carlo simulations, normalization of observed frequencies, annotations, and ‘hit’ papers

We performed analyses on publications from 1985, 1995, and 2005. Building upon prior work (Uzzi et al., 2013), all $\binom{n}{2}$ reference pairs were generated for each publication, where n is the number of cited references in the publication. These reference pairs were then mapped to the journals they were published in using ISSN numbers as identifiers. Where multiple ISSN numbers exist for a journal, the most frequently used one in the WoS was assigned to the journal. In addition, publications containing fewer than two references were discarded. Journal pair frequencies were summed across the dataset to create observed frequencies (F_{obs}).

For citation shuffling, we developed a performant citation switching algorithm, *runtime enhanced permuting citation switcher (repcs)* (Korobskiy, Davey, Liu, Devarakonda, & Chacko, 2019), that randomly permuted citations grouped by year of publication to switch citations

while preserving the year of publication for both articles and references, the number of publications and the number of references in each dataset, and the disciplinary composition of the references in each dataset. Our approach is different from previous studies in these ways: (i) we sampled citations in proportion to their citation frequency (equivalently from a multiset rather than a set) in order to better reflect citation practice, (ii) we permitted a substitution to match the original reference in a publication when the random selection process dictated it rather than attempting to enforce that a different reference be substituted, and (iii) we introduced an error correction step to delete any publications that accumulated duplicate references during the substitution process. As a benchmark, we used the citation switching algorithm of Uzzi et al. (2013), henceforth referred to as *umsj* (as also done in Boyack and Klavans (2014)), using code kindly provided by the authors. A single comparative analysis showed that while 10 simulations of the WoS 1985 dataset (391,860 publications) completed in 2,186 hours using the *umsj* algorithm, it completed in less than one hour using our implementation of the *repcs* algorithm on a Spark cluster. We also tested *repcs* under comparable conditions to *umsj* and estimated a runtime advantage of at least two orders of magnitude. Using either the *repcs* or *umsj* algorithms for 10 simulations resulted in expected value coverage of 75% of the observed journal pair frequencies. Subsequent rounds of development and testing resulted in 99% coverage of observed journal pairs when 1,000 simulations were conducted. The runtime advantage was significant enough that we chose to use the *repcs* algorithm in our study and generated expected values averaged from 1,000 simulations for improved coverage of every dataset we analyzed.

Averaging the result of 1,000 simulations for each dataset studied, z-scores were then calculated for each journal-pair using the formula $(F_{obs} - F_{exp})/\sigma$ where F_{obs} is the observed frequency, F_{exp} is the averaged simulated frequency, and σ is the standard deviation of the simulated frequencies for a journal pair. As a result of these calculations, each publication becomes associated with a set of z-scores corresponding to the journal pairs derived from pairwise combinations of its cited references. Positional statistics of z-scores were calculated for each publication, which was then labeled according to conventionality and novelty: (i) HC if the median z-score exceeded the median of median z-scores for all publications and LC otherwise, and (ii) HN if the tenth percentile of z-scores for a publication was less than zero, and LN otherwise. We also analyzed the effect of defining high novelty using the first percentile of z-scores.

To consider the relationship between citation impact, conventionality, and novelty we calculated percentiles for the number of accumulated citations in the first 8 years since publication for each article we studied and stratified. We investigated multiple definitions of hit articles, with hits defined as the 1%, 2%, 5%, and 10% top-cited articles.

RESULTS

Model Misspecification and the Attributes of Disciplinary Context

A source of misspecification arises from not accounting for disciplinary heterogeneity by treating all eligible references within WoS as equiprobable substituents when studying a disciplinary network. Under this model (Uzzi et al., 2013), the probability of selecting a reference from a discipline is identical to the proportion of the articles in WoS in that discipline for a given year. If the global model accurately reflects citation practice, the expected proportion of references within papers published in a given discipline D would be approximately equal to the proportion of references in D , and conversely, the degree to which the

proportion deviates from the expected value would reflect the extent of model misspecification.

To study the disciplinary composition of references in our custom datasets, we first used the high level WoS classification of five major research areas: life sciences & biomedicine, physical sciences, social sciences, technology, and arts & humanities. The two largest of these research areas are physical sciences and life sciences & biomedicine, which contribute on average approximately 35.1% and 62.8%, respectively, of the references in WoS over the three years of interest. Under the unconstrained model, we would expect close to 35% of the references cited by the publications in any large network to be drawn from the physical sciences and close to 63% of the references to be drawn from life sciences and biomedicine. Yet the empirical data present a very different story: roughly 80% of the references cited in physical sciences publications are from the physical sciences and 90% of the references cited in life sciences & biomedicine publications are from the life sciences & biomedicine. In other words, the empirical data shows a strong tendency of publications to cite papers that are in the same major research area rather than in some other research area. Thus, there is a strong bias towards citations that are *intra-network*. Our observations are in agreement with Wallace et al. (2012) who found that, often, a majority of an article's citations are from the specialty of the article, even though that percentage varied among disciplines in the eight specialties they investigated (from approximately 39% to 89% for 2006). Furthermore, these findings argue that a discipline-indifferent random graph model would exhibit misspecification in deviating substantially from the empirical data, and supports the concern about definitions of innovation and conventionality that are based on deviation from expected values.

We also analyzed disciplinary composition at a deeper level using all 153 Subjects in the WoS extended classification and examining the consequences of citation shuffling within a disciplinary set or all of the Web of Science. References in publications belonging to these three datasets were summarized as a frequency distribution of 153 WoS Subjects as classes. A single shuffle of the references in the disciplinary datasets and in the corresponding WoS year slice was performed, using either the *repcs* or *umsj* algorithms, after which subject frequencies were computed again. The fold change in subject frequencies of references before and after shuffling was calculated for these groups using all 153 subject categories and summarized in the box plots in Fig 1. As an example, the applied physics dataset contained one reference labeled Genetics and Heredity, but after the shuffle (using the WoS background), acquired 1496 references labeled Genetics and Heredity. Similarly, the metabolism dataset contained one reference labeled Philosophy, but after a single shuffle (again using the WoS background) it had 661 occurrences with this label. The data show convincingly that a publication's disciplinary composition of references in a network is preserved when citation shuffling is constrained to the network, but is significantly distorted when the WoS superset is used as a source of substitution. A second inference is that the two algorithms, *repcs* and *umsj*, have equivalent effects in this experiment (and so are only distinguishable for running time considerations).

We then tested the conjecture that model misspecification would be reduced by constraining the substitutions to disciplinary networks by examining the Kullback-Leibler (K-L) Divergence (Kullback & Leibler, 1951) between observed and predicted citation distributions, restricted to the set of journals in a given disciplinary network. The results (Table 3) confirm this prediction: simulations under the constrained model (where the background network is the local disciplinary network) consistently have a lower K-L divergence com-

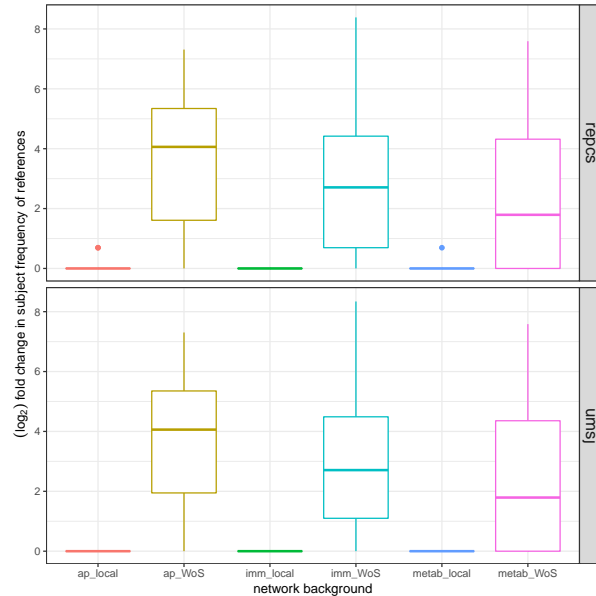


Figure 1. Citation shuffling using the local network preserves the disciplinary composition of references within networks, but using the global network does not. Publications of type Article belonging to the three disciplinary networks (ap=applied physics, imm=immunology, and metab=metabolism) were subject to a single shuffle of all their cited references using either the local network (i.e., the cited references in these networks, denoted bg.local) or the global network (i.e., references from all articles in WoS, denoted bg.WoS) as the source of allowed substitutions, where “bg” indicates the disciplinary network. Citation shuffling was performed using either our algorithm (*repcs*, top row) or that of Uzzi et al. (*umsj*, bottom row). The disciplinary composition of cited references before and after shuffling was measured as frequencies for each of 153 sub-disciplines (from the extended subject classification in WoS) and expressed as a fold difference between citation counts grouped by subject for original (o) and shuffled (s) references using the formula ($\text{fold_difference} = \text{ifelse}(o > s, o/s, s/o)$) and rounded to the nearest integer. A fold difference of 1 indicates that citation shuffling did not alter disciplinary composition. Data are shown for articles published in 1985. All eight boxplots are generated from 153 observations each. Null values were set to 1. Note y-axis: \log_2 scale.

Table 3. Model misspecification is reduced by constraining substitutions to the local disciplinary networks. We computed Kullback-Leibler (K-L) divergences between empirical and simulated journal pair frequencies using two different background networks (local versus global) for each disciplinary network (applied physics, immunology, and metabolism) for the years 1985, 1995, and 2005. K-L divergence was calculated using the R seewave package (Sueur et al., 2008). For every disciplinary network, there is a smaller K-L divergence between simulated and observed data when using the local network (i.e., the disciplinary network) as compared to the global network (all of WoS). Put differently, model misspecification is reduced in the constrained model compared to the unconstrained model.

| Disciplinary Network | Year | Background Network | K-L Divergence | Ratio |
|----------------------|------|--------------------|----------------|-------|
| Applied Physics | 1985 | local | 1.21 | 1.96 |
| | 1985 | global | 2.37 | |
| | 1995 | local | 0.86 | 2.77 |
| | 1995 | global | 2.37 | |
| | 2005 | local | 0.95 | 2.47 |
| | 2005 | global | 2.35 | |
| Immunology | 1985 | local | 0.75 | 2.24 |
| | 1985 | global | 1.68 | |
| | 1995 | local | 0.78 | 2.19 |
| | 1995 | global | 1.70 | |
| | 2005 | local | 0.73 | 2.63 |
| | 2005 | global | 1.92 | |
| Metabolism | 1985 | local | 1.11 | 2.02 |
| | 1985 | global | 2.24 | |
| | 1995 | local | 1.07 | 2.17 |
| | 1995 | global | 2.33 | |
| | 2005 | local | 1.19 | 2.18 |
| | 2005 | global | 2.60 | |

pared to simulations under the unconstrained model (where the background network is all of WoS). Furthermore, the K-L divergence for the unconstrained model is generally twice as large as the K-L divergence for the constrained models, with ratios that range from 1.96 to 2.77, and are greater than 2.0 in eight out of nine cases. These results clearly demonstrate that constraining reference substitutions to the given local disciplinary network better fits the observed data, and hence reduces model misspecification.

Calculation of Novelty and Conventionality using the constrained model

Since the constrained model better fits the observed data, we evaluated the distribution of highly cited articles (i.e., “hit articles”) in the four categories (HNHC, HNLC, LNHC, LNLC), for different thresholds for hit articles. Figure 2, Panels (a) and (b), compares hit rates for the four categories among the immunology, metabolism, applied physics, and WoS datasets for 1995, where the hit rate is defined as the number of hit articles in each category divided by the number of articles in the category. The calculation for the hit rates for the WoS dataset (bottom row, Figure 2) mirrors Uzzi et al.’s results, whereby the largest hit rates were for the HNHC category, despite our methodological changes in sampling citations in proportion to their frequency. However, the trends for all three disciplinary networks are different from those for WoS. Specifically, the highest hit rates for the 1995 immunology and metabolism datasets are in the LNHC category for the top 1% of cited articles (and

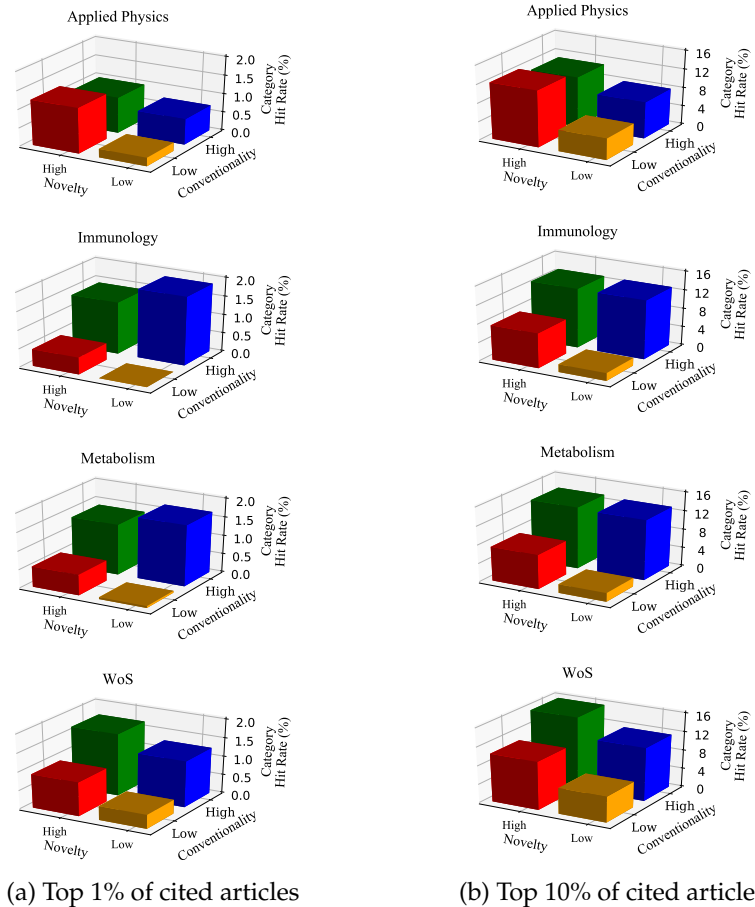


Figure 2. Effect of using the improved model on categorical hit rates for Immunology, Applied Physics, and WoS for 1995. Panels (a) and (b) show hit rates for the LNL, LNH, HNLC, and HNHC categories for the applied physics, immunology, metabolism, and WoS datasets when hit articles are defined as the top 1% and top 10% of articles, respectively. Novelty in both panels is defined at the 10th percentile of articles' z-score distributions. The results for the WoS data set also show that the highest hit rate is for the HNHC category. Results for the three disciplinary networks all differ from the overall WoS results: the highest hit rates for the immunology and metabolism datasets are in the LNH category and the highest hit rate for the applied physics datasets are in the HNLC category. The number of data points in the applied physics, immunology, metabolism, and WoS data sets are 18,305, 21,917, 97,405, and 476,288, respectively.

tied between LNHC and HNHC for the top 10%), and the highest hit rates for the 1995 applied physics datasets are in the HNLC category for both the top 1% and top 10% of all cited articles. Thus, the category exhibiting the highest hit rate among highly cited papers depends on the specific disciplinary network and to some extent on the threshold for being highly cited.

Furthermore, the categories displaying the greatest hit rate vary to some extent with the year. For example, when the 10% top-cited articles are deemed to be hits and novelty is defined at the 10th percentile of z-scores, the category with the highest hit rate in applied physics for 1995 is in HNLC (12.3% versus 10.9% for HNHC), while the hit rate for HNHC is greater than for HNLC in 1985 and 2005 (13.2% versus 10.9%, and 11.4% versus 10.7%, respectively).

We evaluated the statistical significance of the categorical hit rates using multiple methods. Our first test was based on the null hypotheses that hits were distributed randomly among the four categories with uniform probability in proportion to the number of articles in each category. Rejecting the null hypothesis, using a Chi-Square Goodness of Fit test, supports a non-uniform dispersion of hits with some of the four categories being associated with higher or lower than expected hit rates. The null hypothesis was rejected at a $p < 0.001$ in all cases in Figure 2, with the exception of the immunology and applied physics datasets where hit articles are designated as the top 1% of articles: valid tests were not possible in those instances due to too few expected hits. The null hypothesis was rejected with $p < 0.001$ for all valid tests for all parameter settings, all datasets, and all years: hypotheses tests were valid in 73 of 96 instances. We conclude that it is likely that the distribution of hits among categories is not uniform and that, instead, hit rates vary among the categories in all disciplinary datasets.

We also tested the explanatory power of each framework dimension by classifying articles as LN or HN and, separately, as LC or HC. We tested the null hypothesis that hits are distributed between LN and HN (LC and HC) in proportion to the total number of articles assigned to those categories. That null hypothesis was rejected for the WoS data along both dimensions. Consistent with prior findings, hit articles were overrepresented in the HC category in every instance of WoS data at a $p < 0.001$ and also overrepresented in the HN category at a $p < 0.001$ in all but two cases: the p-values in those exceptions were 0.002 and 0.007. Hits in the immunology and metabolism data were overrepresented in the HC category with the same statistical significance as for WoS. The relationship of novelty with hits in the immunology and metabolism dataset differed dramatically from the WoS, however, with statistically significant findings of hit articles being sometimes overrepresented in the LN category, and sometimes being underrepresented. Consistent with the WoS, hit articles in applied physics were positively related with HN with a statistical significance of at least $p < 0.10$ in all 12 parameter sets, and at $p < 0.05$ in 10 of 12 cases. To the contrary, a strong positive relationship was found between LC and hit articles in applied physics in 5 of 12 instances with $p < 0.10$. These results suggest that (1) both conventionality and novelty are strongly related to hits in the WoS, (2) the conventionality dimension is strongly related with hits in immunology and metabolism and novelty is not, and (3) novelty is more strongly related with hits in applied physics than is conventionality. More generally, we find that the dimensions most strongly related with hit articles vary between disciplinary and broad data sets, and also among disciplines.

We described concerns with model misspecification along two general dimensions: the background dataset and sampling methodology for the random graph. The differences we found from prior research in terms of which categories demonstrated the highest hit rates were caused both by using disciplinary datasets and our sampling methodology, *repcs*, through the article z-score distributions. When z-scores are shifted downward using one algorithm versus another, for example, then the former algorithm can result in an increased percentage of HN articles. We therefore examined the extent to which each of our methodological differences contributed to our observations. We found that z-scores changed sign more as a consequence of background network (local network or all of WoS) and much less as a consequence of sampling algorithm (*umsj* or *repcs*). For example, on the immunology dataset, 28.6% of the journal pairs changed signs with our sampling algorithm (*repcs*) as the background network is changed from global (all of WoS) to local, and only 2.8% of z-scores changed signs in the WoS dataset depending on whether *umsj* or *repcs* was used.

We conclude that the choice of background datasets is the source of a majority of differences we observed in the categories demonstrating the highest hit rates, although our sampling approach, most notably sampling from a multiset so as to reflect the observed frequencies of individual citations as well as their associated journals and disciplines, can also create material differences.

DISCUSSION

The principal difference between the two models is a single parameter—the set of references that can be used as substituents during the substitution process. We provided several lines of evidence that showed that changing this one parameter from a global network to the local disciplinary network reduces model misspecification. Furthermore, using the constrained model (which allows substitutions only within the local network) instead of the unconstrained model (which allows substitutions in the entire WoS network) produces different trends in terms of conventionality and novelty, depending on the network. In particular, when using the unconstrained model, highly cited papers were most likely to be in the HNHC category but this trend does not consistently hold when using the constrained model. Instead, we find that conventionality flavored with novelty is *not* generally a feature of impactful research. Further, high “novelty” is not always indicative of impactful research (as reflected in citation counts), as shown by both Immunology and Metabolism having their highest hit rates for low novelty (Figure 2).

More generally, these results show that the trends approaching universality in highly cited papers are not robust to changes in thresholds for defining high impact or high novelty articles, or with time, and may be the consequence of using a random model that has a poor fit to the observed data. On the other hand, while the constrained model reduces model misspecification compared to the unconstrained model, this does not imply that the constrained model is reasonable nor that trends observed under the constrained model convincingly explain scientific practice. Indeed, there are significant challenges in using random models to understand human behavior, of which citation practice is one example.

Our work has shown that the use of local networks enables simulations that are more consistent with research citation patterns. Further work might explore additional constraints on random assignment of citations to publications to better align benchmarks with citation practice. For example, the proximity of possible citations as defined in a co-author network (Wallace et al., 2012) might be considered when defining probabilities for citation

substitutions. Thus, evaluation in context (Hicks, Wouters, Waltman, de Rijcke, & Rafols, 2015) and increased consideration of the disciplinary nature of the scientific enterprise is likely to result in improved models that yield further insight.

ACKNOWLEDGMENTS

We thank the authors of Uzzi et al. (2013) for sharing their simulation code. We are grateful to Kevin Boyack and Dick Klavans for constructively critical discussions. Research and development reported in this publication was partially supported by Federal funds from the National Institute on Drug Abuse, National Institutes of Health, US Department of Health and Human Services, under Contract Nos. HHSN271201700053C (N43DA-17-1216) and HHSN271201800040C (N44DA-18-1216). The content of this publication is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. TW receives funding from the Grainger Foundation. All the code used in this study is freely available from a Github site (Korobskiy et al., 2019). Access to the bibliographic data analyzed in this study requires a license from Clarivate Analytics, which had no role in funding, experimental design, review of results, and conclusions presented.

AUTHOR CONTRIBUTIONS

This study was designed by GC, JB, SD, and TW. Simulations and analysis were performed by AD, DLH, GC, JB, and SD. Infrastructure and workflows used to generate data used in this study were developed by AD, DK, SL, SD, and GC. All authors reviewed and commented on the manuscript, which was written by GC, JB, and TW.

REFERENCES

- Boyack, K., & Klavans, R. (2010). Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately? *Journal of the American Society for Information Science and Technology*, 61(12), 2389–2404. doi: 10.1002/asi.21419
- Boyack, K., & Klavans, R. (2014). Atypical combinations are confounded by disciplinary effects. In *International conference on science and technology indicators* (pp. 49–58). Leiden, Netherlands: CWTS-Leiden University.
- de Solla Price, D. J. (1965). Networks of Scientific Papers. *Science*, 149(3683), 510–515. doi: 10.1126/science.149.3683.510
- Garfield, E. (1955). Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas. *Science*, 122(3159), 108–111. doi: 10.1126/science.122.3159.108
- Garfield, E. (1979). *Citation Indexing-Its Theory and Application in Science, Technology, and Humanities* (1st ed.). The address: John Wiley and Sons, ISI Press.
- Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature News*, 520(7548), 429. doi: 10.1038/520429a
- Keserci, S., Davey, A., Pico, A. R., Korobskiy, D., & Chacko, G. (2018). ERNIE: A data platform for research assessment. *bioRxiv*. doi: 10.1101/371955
- Klavans, R., & Boyack, K. W. (2017). Research portfolio analysis and topic prominence. *Journal of Informetrics*, 11(4), 1158–1174. doi: 10.1016/j.joi.2017.10.002
- Korobskiy, D., Davey, A., Liu, S., Devarakonda, S., & Chacko, G. (2019). *Enhanced Research Network Informatics Environment (ERNIE)* (Github Repository). NET ESolutions Corporation. Retrieved from <https://github.com/NETESOLUTIONS/ERNIE>
- Kullback, S., & Leibler, R. A. (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22(1), 79–86. doi: 10.1214/aoms/1177729694
- Marshakova-Shaikovich, I. (1973). System of document connections based on references. *Nauch-Tekhn.Inform, Ser.2*, 6(4), 3–8. doi: 10.1002/asi.4630240406
- Moed, H. F. (2010). Measuring contextual citation impact of scientific journals. *Journal of informetrics*, 4(3), 265–277.
- Newman, M. E. J. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, 98(2), 404–409. doi: 10.1073/pnas.98.2.404
- Patience, G. S., Patience, C. A., Blais, B., & Bertrand, F. (2017). Citation analysis of scientific categories. *Heliyon*, 3(5), e00300.
- Perline, R. (2005). Strong, Weak and False Inverse Power Laws. *Statistical Science*, 20(1), 68–88.
- Shi, X., Leskovec, J., & McFarland, D. A. (2010). Citing for high impact. In *Proceedings of the 10th annual joint conference on digital libraries* (pp. 49–58). New York, NY, USA: ACM. doi: 10.1145/1816123.1816131

- Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science*, 24(4), 265–269. doi: 10.1002/asi.4630240406
- Stigler, S. (1994). Citation patterns in the journals of statistics and probability. *Statistical Science*, 9(1), 94–108.
- Stringer, M. J., Sales-Pardo, M., & Amaral, L. A. N. (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 the American Society for Information Science and Technology*, 61(7), 1377–1385. doi: 10.1002/asi.21335
- Sueur, J., Aubin, T., & Simonis, C. (2008). Seewave: a free modular tool for sound analysis and synthesis. *Bioacoustics*, 18, 213–226.
- Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. (2013). Atypical combinations and scientific impact. *Science (New York, N.Y.)*, 342(6157), 468–472. doi: 10.1126/science.1240474
- Wallace, M. L., Lariviere, V., & Gingras, Y. (2012). A Small World of Citations? The Influence of Collaboration Networks on Citation Practices. *PLOS One*, 7, e33339. doi: 10.1371/journal.pone.0033339
- Zuckerman, H. (2018). The Sociology of Science and the Garfield Effect: Happy Accidents, Unanticipated Developments and Unexploited Potentials. *Frontiers in Research Metrics and Analytics*, 3, 20. doi: 10.3389/frma.2018.00020