# Measuring discursive influence across scholarship

**Aaron Gerow[a,1], Yuening Hu[b], Jordan Boyd-Graber[c,d,e,f], David M. Blei[g,h,i], and James A. Evans[j,k,1]**

[a]Department of Computing, Goldsmiths, University of London, New Cross, London SE14 6NW, United Kingdom; [b]Cloud AI, Google, Sunnyvale, CA 94809; [c]Department of Computer Science, University of Maryland, College Park, MD 20742; [d]University of Maryland Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742; [e]iSchool, University of Maryland, College Park, MD 20742; [f]Language Science Center, University of Maryland, College Park, MD 20742; [g]Department of Statistics, Columbia University, New York, NY 10027; [h]Department of Computer Science, Columbia University, New York, NY 10027; [i]Data Science Institute, Columbia University, New York, NY 10027; [j]Department of Sociology, University of Chicago, Chicago, IL 60637; and [k]Computation Institute, University of Chicago, Chicago, IL 60637

Edited by Peter S. Bearman, Columbia University, New York, NY, and approved February 12, 2018 (received for review November 18, 2017)

**Assessing scholarly influence is critical for understanding the collective system of scholarship and the history of academic inquiry. Influence is multifaceted, and citations reveal only part of it. Citation counts exhibit preferential attachment and follow a rigid "news cycle" that can miss sustained and indirect forms of influence. Building on dynamic topic models that track distributional shifts in discourse over time, we introduce a variant that incorporates features, such as authorship, affiliation, and publication venue, to assess how these contexts interact with content to shape future scholarship. We perform in-depth analyses on collections of physics research (500,000 abstracts; 102 years) and scholarship generally (JSTOR repository: 2 million full-text articles; 130 years). Our measure of document influence helps predict citations and shows how outcomes, such as winning a Nobel Prize or affiliation with a highly ranked institution, boost influence. Analysis of citations alongside discursive influence reveals that citations tend to credit authors who persist in their fields over time and discount credit for works that are influential over many topics or are "ahead of their time." In this way, our measures provide a way to acknowledge diverse contributions that take longer and travel farther to achieve scholarly appreciation, enabling us to correct citation biases and enhance sensitivity to the full spectrum of scholarly impact.**

scholarly influence | science of science | probabilistic modeling

Scholarship is a complex and chaotic process by which authors create, share, and promote new concepts, theories, methods, data, and findings. Subsequent research adopts scholarly innovations through direct contact with the original work or through follow-on research, review articles, seminars, and personal conversation. Researchers may acknowledge such influence by citing the work on which they draw or passively adopting the vocabulary, ideas, assumptions, approaches, or insights without explicit references. In this way, citations constitute an important but incomplete signal of influence, the full spectrum of which includes direct and indirect—acknowledged and ignored—influences from across the scholarly landscape. We hold that a key dimension of influence involves the ability to change scholarly discourse, which itself interacts with contextual features, such as the status of publication venues, authors' namesakes, and institutions that host the research. Here, we seek to detect discursive influence and the factors that shape it.

Research impact is commonly assessed by the number of explicit references to a document, author, or journal. A reference list represents a complex combination of considerations, including perceived influence of important arguments, methods, data, or findings but also, what authors, reviewers, and editors believe should be cited to appease readers or enhance their own status. Scientometrics provides measures of author, journal, and institutional impact (1), which tabulate functions of the citation distribution over articles and time. New "alt metrics" go beyond citations to assess online views, downloads, likes, and tweets or estimate readership, cocitation, and diversity (2, 3). Like citations, these metrics provide an informative but distorted portrait of influence: they exhibit rich-get-richer feedback (4–6) and manifest manipulation by authors and editors (7–9). Authors may cite their own work to drive traffic to prior articles (10, 11). Less nefariously, citations may certify membership in a community, reveal intellectual alliances, or reflect the status aspirations of a paper (12–14). Moreover, authors are often unable to cite all of their influences, selecting citations based on cognitive availability and space constraints (15, 16). Over time, citations of a paper tend to decay due to a preference for recentness (17), even when a paper continues to be influential. Further complicating the matter, citation cultures differ across disciplines and time periods (18, 19).

The primary artifact of scholarly work is not citations but text. Here, we offer a measure of influence based on text and context as they shape future discourse. We use scholarly discourse to reference the amalgam of language in academic publications, that traces a sequence of communication and scholarly argument. We adopt a broad notion of discourse as the overall state of scholarship expressed in published work. This approach avoids some challenges of discourse analysis while raising new ones. Analyzing specific aspects of scholarly conversation requires a selection process, the result of which may not generalize. Alternatively, analyzing a vast number of publications presents technical challenges. To model more general notions of discourse, we use a class of probabilistic techniques called topic models that extract patterns of term–document co-occurrence and yield semantically related word distributions called "topics."

Here, we develop a dynamic topic model designed to explain not only which documents are influential but how their influence is derived from content and its interaction with features, like authorship, affiliation, and publication venue. We build on models of lexical change (20, 21) and in particular, the document

## Significance

**Scientific and scholarly influence is multifaceted, shifts over time, and varies across disciplines. We present a dynamic topic model to credit documents with influence that shapes future discourse based on their content and contextual features. We trace discursive innovation in scholarship and identify the influence of particular articles along with their authors, affiliations, and journals. In collections of science, social science, and humanities research spanning over a century, our measure helps predict citations and reveals signals that recognize authors who make diverse contributions and whose contributions take longer to be appreciated, allowing us to compensate for bias in citation behavior.**

Author contributions: A.G., Y.H., J.B.-G., D.M.B., and J.A.E. designed research; A.G., Y.H., J.B.-G., and J.A.E. performed research; A.G., Y.H., J.B.-G., and J.A.E. contributed new reagents/analytic tools; A.G. and J.A.E. analyzed data; and A.G., Y.H., J.B.-G., D.M.B., and J.A.E. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.



Data deposition: The source code of the model's implementation has been deposited on GitHub and is available at https://github.com/gerowam/influence.

[1]To whom correspondence may be addressed. Email: a.gerow@gold.ac.uk or jevans@uchicago.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1719792115/-/DCSupplemental.

Published online March 12, 2018.

influence model (DIM) (22). In the DIM, documents receive a score based on how they change future topics, but the model offers no mechanism to explain the composition of influence. Our variant offers such an explanation by explicitly modeling document-level covariates alongside content. This allows for comparing different kinds of authors, institutions, and publication venues for each topic individually. We affirm the value of such features, as they help predict citations in a widely studied corpus of computational linguistics research (23). We also introduce a mechanism to assess a document's contribution over time by simulating future discourse without it. This dynamic measure of contribution helps address the limited scope of citations and provide a composite understanding of scholarly influence.

## Robust Model of Influence

We consider a document to be influential to the extent that it affects future discourse. Probabilistic topic models enable the analysis of such discourse in large text collections (24–26). Topics $\{k_0, \cdots, k_K\}$ are discovered from lexical co-occurrence in a set of documents and refer to probability distributions over observed words, the most likely of which typify a topic (27). In time-ordered texts, dynamic topic models derive topics from documents binned into periods $\{t_0, \cdots, t_T\}$, allowing topics to change over time (20). Each document is fit with a document–topic mixture, $\theta_d$, and each word is fit with a topic assignment $z_{d,k,n \in N}$ over the vocabulary $N$. The first (and to our knowledge, only) model to explicitly measure how individual documents change future topics is the DIM (22). DIM learns document influence in a topic, $\ell_{d,k}$, based on how it changes future topics but does not explain its composition.

Topic models have proven effective for analyzing scientific literature (21, 28, 29), although they make assumptions about the data. In particular, such models assume that a specified number of topics are present throughout. Statistically, words are assigned to topics as those topics semantically shift from "background" topics (composed of collection-wide words) to specific, coherent topics. Choosing the number of topics, $K$, is important and has empirical implications, but it can be interpreted as the number of retained discursive dimensions (27, 30). Choosing the number of topics was done in a theoretically motivated, data-driven fashion using static models of the same data fit with many topics to identify a reasonable threshold for how many topics the data used (*SI Appendix*). This approach approximates a Bayesian nonparametric search over possible numbers of topics (31, 32) and allows us to discover the topic complexity for each corpus.

Our variant adds an important explanatory mechanism to establish how influence arises. By incorporating a latent regression on document covariates, the model estimates the marginal effect of authorship, institution, and journal on influence. Although a modest technical innovation, this enables robust explanations of discursive influence and its origins. A document's influence in a particular topic, $\ell_{d,k}$, is a product of its content and its interaction with associated contexts traced in article metadata. The coefficients that solve the latent regression, denoted $\mu_k$, offer estimates of how covariates add or detract from influence in each topic. As we will see, this opens a range of insights into how scholarly influence unfolds.

## Results

The model produces three important results: (*i*) topics (word mixtures over time, $\hat{\beta}_{t,k}$, and topic mixtures within documents, $\theta_k$), (*ii*) variational estimates of influence ($\hat{\ell}_{d,k}$), and (*iii*) estimated marginal effects of covariates on influence ($\hat{\mu}_k$). A corpus of computational linguistics research (Association for Computational Linguistics Anthology Reference Corpus; ACL-ARC) (23) was used to compare our variant with DIM. Despite the additional complexity, our model performed comparably with DIM in terms of convergence in the likelihood bound and perplexity. Estimates of influence in DIM correlated with citation counts: $\rho = 0.22$ ($K = 10$), $\rho = 0.21$ ($K = 20$), $\rho = 0.21$ ($K = 50$), and $\rho = 0.18$ ($K = 75$). Our variant yielded significantly stronger and substantially larger correlations: $\rho = 0.28$ ($K = 10$), $\rho = 0.27$ ($K = 20$), $\rho = 0.23$ ($K = 50$), and $\rho = 0.22$ ($K = 75$; Fisher $z$ transform; all $p < 0.01$. This confirms that the exogenous features captured by our model have an observable effect on predicting citations. Furthermore, modeling these additional features adds explanation to our notion of discursive influence.

**Influence in Physics.** A collection of research published by the American Physical Society (APS) was used to estimate a static 500-topic model, discover an optimal 37-topic solution, and then, fit our 37-topic dynamic influence model. The APS collection provides a rich citation environment, where 90% of documents in the sample were cited by other APS publications. Most of the resulting topics typify subfields and allow us to trace the emergence of modern concepts in physics (*SI Appendix*). We found that document influence ($I_d$) (Eq. **8**) correlated with citation counts, $C_d$, at $\rho = 0.32$ ($p < 0.01$). Some anecdotal results affirm that discursive innovations are meaningful and substantive. For example, specialization is traced by authors' outsized marginal influence in a few, related topics (Fig. 1*A*). Ed Witten has driven advances in cosmology and string theory, Arno Penzias' radio telescope contributed greatly to understanding the big bang and the development of new cosmic detector arrays, and Philip Anderson cofounded spin-glasses as a model of magnetic phase transitions. Witten's byline lends a positive effect for influence in cosmology-related topics as does Penzias' byline in detectors and atomic physics and Anderson's byline in lattice quantum chromodynamics and superconductors among others. Nobel Laureates, overall, have a significantly more positive effect on their document's influence than those without a Nobel (Fig. 1*D*).

Discursive influence provides a way to measure impact. Fig. 2 lays out four kinds of papers based on their relative discursive influence and citation count. Each quadrant is characterized by high–low bins. Empirically, we define these quadrants relative to median influence and number of citations. Papers in the high/high and low/low quadrants participate in the standard scientific cycle of credit, where discursive contributions are cited by future work. Here, the signals from citations and discursive influence substitute for one another, with neither adding much information to the assessment of impact.

Papers in the off-diagonal quadrants, however, adhere less cleanly to this pattern. Citations and discourse here constitute complementary signals in at least three ways. First, citations exhibit preferential attachment, favoring authors who persist in narrow subfields of science, whereas discourse identifies greater influence for itinerant scientists whose careers span diverse topics. Second, citations tend to follow a skewed, log-normal distribution over time with rapid uptake followed by a diminishing tail of attention. As such, citation patterns favor articles that receive most of their references within this scientific news cycle. Document influence, however, credits papers that are enduringly influential as well as so-called "sleeping beauty" (SB) papers, which experience a delay before discovery and tend to have high document influence (Fig. 1*C*). Third, citations uniquely capture nondiscursive contributions, like the provision of new data or critiques, whereas influence captures both direct and indirect innovations.

We propose that scientists and scholars experience professional pressure to cite touchstone works of authors with an established presence in their field. Authors who contribute to more diverse topics post high discursive influence but are less likely to receive high citations. We measure author establishment through persistence (Eq. **11**), calculated as the inverse entropy of authors' sum of document–topic mixtures, scaled by their prolificness. Author persistence correlates with their most cited papers ($\rho = 0.57$, $p < 0.01$), much more than with the same author's most influential paper ($\rho = 0.34$, $p < 0.01$). More persistent authors receive more citations, but on average have less influential documents (Fig. 1*B*). Authors who remain productive in a narrow range of topics are more likely to have a highly cited paper than if they make more scattered contributions. Papers in

SOCIAL SCIENCES

APPLIED PHYSICAL SCIENCES

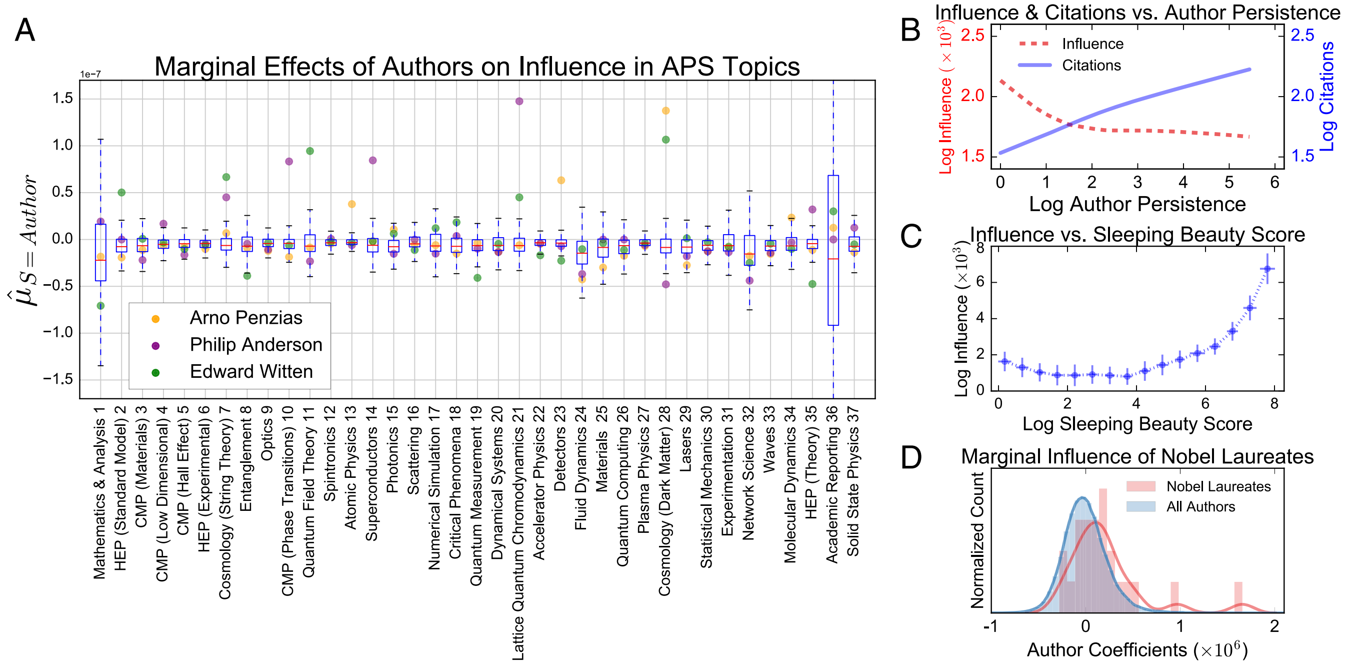


**Fig. 1.** (*A*) Box plots of author coefficients for each APS topic. Medians are shown as a red line within each box, the first quartiles are within the box, and the second and third quartiles are within bands. CMP, condensed matter physics; HEP, high energy physics. Note the wider distributions for more general topics 1 and 36. Overlaid are coefficients for three physicists. Positive values mean that an author's byline adds influence, whereas a negative value means it detracts. A positive coefficient does not necessarily mean that a document is highly influential itself, only that it was more influential than if it had it been written by the average author. (*B*) Locally weighted scatter plot smoothing (LOWESS) curves comparing document influence (dashed red line; left *y* axis) and citations (solid blue line; right *y* axis) with author persistence (Eq. **11**) (*x* axis). A consistent and statistically significant trend is established: more persistent authors tend to produce more highly cited but less influential documents, whereas less persistent authors have more influential but less cited documents. (*C*) LOWESS curve fit to the plot of documents' influence vs. their SB score. Error bars are $\pm 2$ SE of the mean in both dimensions. (*D*) Distribution of authors' marginal effect on influence for Nobel Laureates compared with all other authors.

the high citation/low influence bin include those from authors whose papers one "must cite," a self-reinforcing process where papers by persistent authors become required scholarly boundary markers. Authors of papers in this region are more persistent than those in any other quadrant (*SI Appendix*, Fig. S15). Conversely, papers in the low-citations/high-influence bin disproportionately credit itinerant authors who contribute to many fields.

We find that norms of citation encourage attention to recent work. We evaluate this by examining how articles make dynamic contributions to discourse over time. Although document influence is a static attribute representing a paper's effect on overall discourse, sometimes this influence occurs immediately, while in other cases, it takes time to foment. To establish the dynamics of discursive impact, we measure a document's topic contribution by estimating how different future topics would have been without it (Eq. **10**). Fig. 3 illustrates this dynamic contribution for two different kinds of SB papers. Felix Bloch's 1946 "Nuclear induction" (33) is typical in that it had a small impact on most topics but a sizeable one on a few. It is also typical in that it received a spike of citations shortly after publication, which then began to decay. Near the turn of the century, we see a significant spike in both the article's citations and its contribution of the Hall effect to the condensed matter physics topic. This is when transistors in microchips became small enough to be affected by the quantum Hall effect and the nuclear magnetic moments described by Bloch. A second paper, the top cited in our sample from 1947, is a "pure" SB, which garnered relatively few citations until a sudden spike in 2004. Philip Wallace's "The band theory of graphite" (34) described the structure of graphene, which at the time, was only observable on an iron film. Theoretically, graphene was exceptionally strong and "growable" in a block structure, but the technology to characterize and isolate it without a substrate was not discovered until 2004, which was awarded the 2010 Nobel Prize in Physics. Wallace's paper (34) contributed modestly to most topics and significantly to a few, but in 2001, its contribution to the materials topic jumps significantly.

With a sample of 500 articles from each percentile of document influence, we assessed the variance and half-life of citations compared with topic contribution. Half-life is the number of time steps after which the score is one-half what it was initially. Documents in higher percentiles of influence have longer half-lives, and in every percentile, citations have lower variance and shorter half-lives than topic contribution (*SI Appendix*). This suggests a conventional "statute of limitations" through which scholars no longer need to cite older articles with persistent influence, possibly because those ideas have entered popular consciousness or because their original authors are no longer around to claim them.

Other articles defy the scientific news cycle not because they are persistently influential, but because they garner little attention for a long period before a spike of citation attention, not unlike Wallace's paper (34) discussed above. The SB index (35) identifies such articles as a function of the convexity and time-to-maximum citations over time. An article that is dormant for a long time, after which it suddenly receives a burst of citations, has a high SB score, whereas the typical article, which receives a burst of citations on publication that decay thereafter, receives a low score. SB scores correlated four times more strongly with document influence ($r = 0.20$, $p < 0.01$) than with citations ($r = 0.05$, $p < 0.01$), suggesting that papers ahead of their time nevertheless receive fewer citations than their influence would predict. Such high-influence/low-citation articles violate the standard influence trajectory.

Finally, high citation/low influence papers contain important nontextual features, while high-influence/low-citations papers tend to have influence that may not be credited in citations. For example, the most cited paper in the ACL-ARC collection introduced the Penn Treebank (36), an important resource that accelerated research in parsing. Nevertheless, it was the resource and

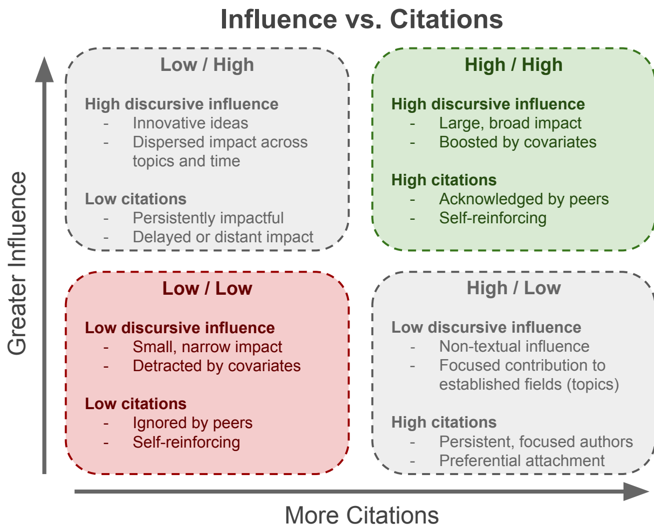


**Fig. 2.** A framework for scholarly impact: citations vs. discursive influence.

not conceptual innovations that made up its impact (22). When we mapped comment- and reply-type APS publications to the framework of Fig. 2, 54% were in the lower right quadrant, where they are highly cited but have low influence. Critical comments and replies kill existing lines of inquiry rather than birth new ones, offering corrections, rebuttals, or direct refutations without producing new concepts or terminology. As a result, they attract citations but do not stimulate discursive emulation. Only 6% were in the high influence/low citations region.

Many highly influential but undercited papers tended to be authored by early innovators whose discursive contributions became pervasive over generations. This is also artificially the case for articles in the first years of the corpus that channel ideas borrowed from documents published prior to our sample. We empirically defined the first year to the year during which mean document influence was within 2 SDs of the global mean as a burn-in period, up until 1942 for APS. Excluding these early papers, high citation/low influence papers are from authors who contributed innovations that are primarily nontextual in nature (data, methods, research-killing refutations, etc.), while high-influence/low-citation papers are from conceptual innovators whose language has become assumed within a field.

**Scholarly Reach in JSTOR Repository.** While physics forms a community with shared publishing habits, research in JSTOR offers a wider sample of academic traditions. After estimating a static, 500-topic model and discovering that using 53 topics best characterized the corpus, we fit the 53-topic dynamic model to full texts in JSTOR and found that citations were modestly correlated with document influence ($\rho < 0.17$, $p < 0.01$) over all topics. The diversity of JSTOR means that our sample includes documents from vastly different citation cultures. The correlation between citation counts and influence within individual topics was varied and weak. When citations were scaled by document–topic mixtures ($\theta_{d,k} C_d$), however, the correlations were all significantly more positive (*SI Appendix*, Table S6). This suggests that, while subjects have different citation habits, our measure of influence is sensitive to the subject variation captured by distinct topics. Using JSTOR's subject taxonomy, citations were grouped by their distance: zero, one, or two. Most citations are within subject (i.e., distance 0), while 10% have distance 1 and 5% have distance 2. Both outgoing and incoming citations show a preference for influential papers (Fig. 4 *B* and *C*). This shows that influential papers reach farther across disciplines in their references. Likewise, influential papers are cited from farther away. Influential papers are both more likely to cite work beyond their subject and to be cited from other subjects. Moreover, influential articles that receive citations from distant work have a higher ratio of influence to citations than works of more local interest. This suggests that the citation statute of limitations illustrated above may operate in not only time but also scientific space, releasing scholars from the obligation to cite influential work that is sufficiently old or topically distant from the influenced work, where the originating authors are not present to claim them.

Predicting citations is notoriously difficult (15, 17), in part because many papers are never cited. Looking at which papers receive any citations (29% in JSTOR), more influential papers were more likely to be cited outright (Fig. 4*A*). Document influence is also predictive in citation models: a logit model predicting citedness with document influence and publication date, $t_d$, of the form $C_d > 0 \sim I_d \times t_d + 1$ estimated a strong positive effect for document influence [$\beta(I_d) = 0.42$, $P > |z| = 0.000$]. Influence also helps predict actual citation counts. Using logistic link negative binomial regressions of the same form as above, the fully specified model estimated similar effects with $\beta(I_d) = 0.33$ (*SI Appendix*). Similar models fit to each topic produced comparable results. We also found topicwise variation in how citation is related to influence for authors. Authors' marginal impact on their documents' influence was more strongly correlated with their citation counts ($\hat{\mu}_{Auth} \times C_{Auth}$) in humanities and social sciences (e.g., philosophy, literary theory, and education topics). In mathematical and natural science fields (e.g., cell biology, physical chemistry, and various statistics topics), the correlation was considerably weaker and in many cases, nonexistent (*SI Appendix*, Table S7). This suggests that authors in "narratively" driven areas are much more likely to have citations conferred on them for having influenced the total flow of discourse than authors from empirical and formal fields, who may be more likely credited for more specific contributions.

The relationship between influence and citations in JSTOR is topic-specific, but some documents contribute to a variety of topics. We sampled some of the most cited documents in each percentile of the influence distribution that were published after the burn-in cutoff (1930 for JSTOR) (*SI Appendix*, Fig. S9). In the lowest percentile, the top-cited document was 1939's "A system for marking turtles for future identification" by Fred R. Cagle (37). The paper, which lives up to its title, remains highly cited in ecology and zoology research. By explaining a technique, "A system for marking turtles for future identification" offers little in the way of topic contribution: 9 topics were significantly changed by its publication, while the remaining 44 go unchanged. Cagle's

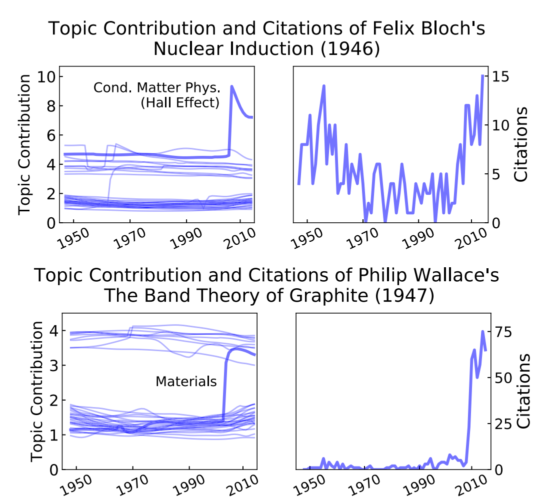


**Fig. 3.** Topic contributions (Eq. **10**) and citations for Felix Bloch's "Nuclear induction" (33) (*Upper*) and Philip Wallace's "The band theory of graphite" (34) (*Lower*). Both papers featured the typical contribution profile where they affect change in a few topics, which diminishes slowly over time. Each paper also exhibits a late spike in citations matched by a coincident spike in contribution to a specific topic (labeled).

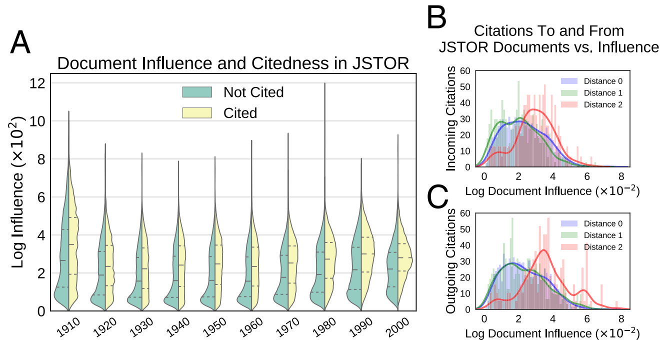


**Fig. 4.** (*A*) Violin plot of document influence (*y* axis; kernel density estimate bandwidth = 0.1) for cited and uncited documents in JSTOR grouped by decade; 2010 is omitted as incomplete. (*B* and *C*) Histograms and density estimates of (*B*) incoming and (*C*) outgoing citations among JSTOR documents grouped by distance in the subject tree.

paper (37) typifies low-influence/high-citation papers that make technical offerings adopted by future research, but that do not incite discursive shifts. In the 50th percentile, 1965's "Population fluctuations and clutch-size in the great tit" by C. M. Perrins (38) was the most cited (504 times within sample). Perrins' discursive contribution is spread over a range of topics and time. It immediately shifted discourse in biology- and ecology-related topics, fields where it continues to be cited today as a landmark paper. After publication, topics in statistics, group behavior, sexual health, medicine, and psychology adopted some of the conceptual terminology and remained changed to the end of the corpus. These contributions owe much to Perrin's use of terms about population dynamics, movement, procreation, and individual variation. Finally, the highest cited paper in the 90th percentile of influence was Sam Peltzman's 1976 "Toward a more general theory of regulation," (39) which had an influence $+3.5$ SD above the mean and 701 citations. Recall that influence is a latent variable fit within the model, while contribution is a post hoc estimate of how different topics would have been without a given document. Peltzman's 1976 paper (39) has a typical contribution profile—significantly influencing a few topics but only slightly altering most—although it sits near the top of the influence distribution. This is because much of its influence was derived extrinsically: it was published by an eminent researcher (who in 2013, *Wired* magazine listed as 1 of 28 top scientists without a Nobel Prize) in the high-impact *Journal of Law and Economics*, and it was published by the University of Chicago Press, a leading publisher of economics research. "Toward a more general theory of regulation" exemplifies a configuration to which many researchers can relate: good papers are often made more so with contextual boosts, like authorship, venue, and publisher.

## Discussion

Our measure of influence tracks changes in future discourse and explicitly identifies the content and context of previous documents that affect these discursive shifts. Document influence provides a direct measure of impact that allows us to disentangle many dimensions of influence across domains of science and scholarship. The model also estimates lasting contributions to topics over time and helps discover influential but undercited work. Our measures not only help predict citations better than previous similar models, but they provide an explanation of what drives influence. Most importantly, the model reveals how discursive innovations are adopted and credited. Alongside citations, discursive influence brings us closer toward a full-spectrum estimate of scholarly impact.

Assessing scholarly influence is a retrospective task, and it can be distorted by conflating citations with impact. Not only does our method enable previously impossible analyses of discursive influence, it could help authors decide what to cite and why. On the one hand, citations provide a simplified, censored, and synoptic trace of acknowledged influence—an important trace of impact. On the other hand, contributing change to scholarly discourse is another measurable kind of influence that can be explained over time, across individual subjects, and with respect to authors, institutions, and publication venues, all of which contribute to the complex evolution of scholarship.

## Materials and Methods

The APS collection contains 509,007 abstracts from 1913 to 2015 coded with type (article, review, commentary, etc.), journal, authors, and their institutional affiliations. To avoid spurious metadata, only papers with an author found twice and an affiliation that occurred three times were retained. This resulted in 74,459 covariates, $\tau$, over 251,382 documents dating from 1918 to 2015. Stop words, infrequent words, and statistically uninteresting words (by term frequency–inverse document frequency) were discarded. The final vocabulary contained 15,312 tokens. The model was fit with 37 topics, a value chosen by assessing topic use in a static, Bayesian nonparametric model with 500 topics (*SI Appendix*). Labels were given by three researchers with doctoral degrees in physics.

The JSTOR collection consists of over 2 million documents from 1894 to 2014; 28,861 covariates were coded in $\tau$ representing authorship, journal/venue, publisher, and discipline. Documents were excluded if they did not have at least one author with three or more documents or if they were classified in disciplines with a 20-y gap, representing subject instability or death (e.g., railroad science). The final sample contained 428,034 full-text articles. The vocabulary was processed similarly to APS and resulted in 20,155 tokens. A model with 53 topics was estimated, with the number of topics selected by fitting a 500-topic static model. Topics were labeled by the authors and assisted by Google Scholar. Discussion of the data, model specification, a closer inspection of the resulting topics, and the labeling process are in *SI Appendix*.

**The Generative Model.** We assume that influence is drawn from a Gaussian, the mean of which is given by a projection on $\tau_{t,d}$:

$$\ell_{d,k} \sim \mathcal{N}(\mu_k \tau_d, \sigma_\ell^2 \mathbb{I}), \qquad [1]$$

where $\mu$ is a $K \times S$ matrix of topic-specific coefficients. We assume that $\mu$ is drawn from a centered Gaussian of specified variance:

$$\mu_k \sim \mathcal{N}(0, \sigma_\mu^2 \mathbb{I}). \qquad [2]$$

The generative process for each time slice $t$ is as follows.

1) Draw topics $\beta_{t+1}|\beta_t, (w, \ell, z)_{t,1:D_t} \sim \mathcal{N}(\beta_t + \exp(-\beta_t) \sum_d \ell_d \sum_n w_{d,n} z_{d,n}, \sigma^2 \mathbb{I})$
2) Draw coefficients $\mu_k \sim \mathcal{N}(0, \sigma_\mu^2 \mathbb{I})$
3) For each document $d$ at time $t$:
   a) Draw $\theta_{t,d} \sim \mathrm{Dir}(\alpha)$
   b) For each word $w_{t,d,n}$
      *i*) Draw $z_{t,d,n} \sim \mathrm{Mult}(\theta_{t,d})$
      *ii*) Draw $w_{t,d,n} \sim \mathrm{Mult}(\pi(\beta_{t,z_{t,d,n}}))$
   c) Draw $\ell_{t,d} \sim \mathcal{N}(\mu \tau_{t,d}, \sigma_\ell^2 \mathbb{I})$,

where $\pi(\mathbf{x})$ maps the multinomial parameters $\mathbf{x}$ to their mean.

**Approximate Inference.** The model has latent variables for words, documents, time slices, topics, and covariate coefficients. Collapsed Gibbs sampling and direct expectation maximization are precluded by nonconjugacy in the topic parameters $\{\beta_1, \cdots, \beta_T\}$. Instead, the model estimates variational parameters that minimize the Kullback–Leibler (KL) divergence to the true posterior. Here, the topic parameters are a Gaussian chain governed by the variational parameters $\{\hat{\beta}_{k,1}, \cdots, \hat{\beta}_{k,T}\}$ that describe their means. The latent variable $\mu$ is also fit by variational estimation to the approximate $\hat{\mu}$. The variational distribution for a document's influence is given by the Gaussian of the mean influence and specified variance. This variational distribution, $q(\beta, \ell, \theta, \mathbf{z}, \mu|\hat{\beta}, \hat{\ell}, \gamma, \phi, \hat{\mu})$, is

$$\prod_{k=1}^{K} q(\beta_{k,1}, \cdots, \beta_{k,T}|\hat{\beta}_{k,1}, \cdots, \hat{\beta}_{k,T}) \times \prod_{k=1}^{K} q(\mu_k|\hat{\mu}_k) \times \prod_{t=1}^{T} \left( \prod_{d=1}^{D_t} q(\theta_{t,d}|\gamma_{t,d}) q(\ell_{t,d}|\hat{\ell}_{t,d}) \times \prod_{n=1}^{N_{t,d}} q(z_{t,d,n}|\phi_{t,d,n}) \right). \qquad [3]$$

The simplified objective is fit by expectation maximization. Two terms of the evidence lower bound are related to $\ell$, which requires expectation- and maximization-step updates that incorporate the projection in Eq. **1**. We define $X = \mathrm{Diag}(\exp(-\beta_{t,k}))(\mathbf{w}_t \circ \mathbf{z}_{t,k})$ and $\Delta\beta_{t,k} = \beta_{t+1,k} - \beta_{t,k}$. With $S$ covariates, $\tau_d$ is an $S$-length vector, and $\tau$ is an $S \times D$ matrix coding observed covariates across $D$. Thus, $\mu$ is a $K \times S$ matrix, the values of which will converge to an estimate of each covariate's effect on $\ell$. The lower bound of $\hat{\ell}_{t,d,k}$ is given by

$$\mathcal{L}_{\hat{\ell}_{t,k}} = \frac{1}{\sigma^2}\mathbb{E}_q\left[X^{\mathsf{T}}\Delta\beta_{t,k}\right]\hat{\boldsymbol{\ell}}_{t,k} - \frac{1}{2\sigma^2}\mathbb{E}_q\left[X^{\mathsf{T}}X\right]\hat{\boldsymbol{\ell}}^2_{t,k} - \frac{1}{2\sigma_d^2}(\hat{\ell}_{t,k} - \hat{\mu}_k\tau_t)^2. \quad [4]$$

This provides the E-step update for influence:

$$\hat{\boldsymbol{\ell}}_{t,k} \leftarrow \left(\mathbb{E}_q\left[X^{\mathsf{T}}X\right] + \frac{\sigma^2}{\sigma_d^2}\mathbb{I}\right)^{-1}\left(\mathbb{E}_q\left[X^{\mathsf{T}}\Delta\beta_{t,k}\right] + \hat{\mu}_k\tau_t\right). \quad [5]$$

The lower bound of $\hat{\mu}$ then is given by

$$\mathcal{L}_{\hat{\mu}_k} = \sum_t\sum_d\sum_k -\frac{1}{2\sigma_\ell^2}(\hat{\ell}_{t,d,k} - \hat{\mu}_k\tau_{t,d})^2 - \sum_k\frac{\hat{\mu}_k^2}{2\sigma_\mu^2}. \quad [6]$$

This yields the M-step update for the coefficients:

$$\hat{\mu}_k^{\mathsf{T}} \leftarrow \left(\sum_t\sum_d \tau_{t,d}\tau_{t,d}^{\mathsf{T}} + \frac{\sigma_\ell^2}{\sigma_\mu^2}\mathbb{I}\right)^{-1}\sum_t\sum_d \tau_{t,d}\hat{\ell}_{t,d,k}. \quad [7]$$

$\hat{\mu}$ is initialized by random draws from a centered, multivariate Gaussian, and its maximum likelihood estimation is done by variational expectation maximization.

**Model-Derived Metrics.** Document influence is the sum topic-proportional influence:

$$I_d = \sum_k \theta_{d,k}\hat{\ell}_{t,d,k}. \quad [8]$$

This offers a plausible interpretation: a document's influence is proportional to how much it changes the topics from which it draws.

Whereas $\hat{\ell}_{d,k}$ is a static feature, the topic contribution of a document over time is also accessible. An estimate of document contribution to a topic $k$ at time $t'$ can be computed by simulating the topic without document $d$:

$$\hat{\beta}^{(-d)}_{t',k} = \left\|\hat{\beta}_{t',k} - (w_d z_{d,k}\hat{\ell}_{d,k})\right\|_1. \quad [9]$$

This provides a conservative estimate, because it overlooks topic drift. The contribution can then be measured as the divergence between the topic with and without $d$:

$$Contribution(d,t') = KL\left(\hat{\beta}_{t',k}, \hat{\beta}^{(-d)}_{t',k}\right). \quad [10]$$

We define author persistence using the inverse independent entropy over their documents' topic mixtures and their prolificness in terms of documents and timespan. Given an author's set of documents $\mathbf{A}$, we define persistence as

$$Persistence(\mathbf{A}) = \left(1 - \frac{H\left(\|\sum_{d\in\mathbf{A}}\theta_d\|_1\right)}{\lceil H\rceil_K}\right)\cdot(|\mathbf{A}| + \Delta_t(\mathbf{A})), \quad [11]$$

where $H(p)$ computes the entropy of probability distribution $p$ as $\sum_i p_i\log(p_i)$ and $\lceil H\rceil_K$ is the maximum possible entropy for a $K$ vector. The first term–inverse entropy over an author's total document–topic mixture–is scaled up by the number of documents $|\mathbf{A}|$ and the number of years over which they were published, $\Delta_t(\mathbf{A})$. This measure is unbounded: authors can be ever more persistent given more time and documents in the same topics.

**ACKNOWLEDGMENTS.** A.G. was supported by a fellowship from the Izaak Walton Killam Foundation. A.G. and J.A.E. were supported by a grant from the Templeton Foundation (to the Metaknowledge Research Network). J.B.-G. is supported by National Science Foundation (NSF) Grant NCSE-1422492. J.A.E. is also supported by Air Force Office of Scientific Research Grant FA9550-15-1-0162 and NSF Grant SBE1158803.


1. Hirsch JE (2007) Does the h index have predictive power? *Proc Natl Acad Sci USA* 104:19193–19198.
2. Bergstrom CT, West JD, Wiseman MA (2008) The eigenfactor metrics. *J Neurosci* 28:11433–11434.
3. Fersht A (2009) The most influential journals: Impact factor and eigenfactor. *Proc Natl Acad Sci USA* 106:6883–6884.
4. Price DdS (1976) A general theory of bibliometric and other cumulative advantage processes. *J Am Soc Inf Sci* 27:292–306.
5. Wang M, Yu G, Yu D (2008) Measuring the preferential attachment mechanism in citation networks. *Phys A* 387:4692–4698.
6. Newman MEJ (2004) Coauthorship networks and patterns of scientific collaboration. *Proc Natl Acad Sci USA* 101:5200–5205.
7. Alberts B (2013) Impact factor distortions. *Science* 340:787–787.
8. Wilhite AW, Fong EA (2012) Coercive citation in academic publishing. *Science* 335:542–543.
9. Wenneras C, Wold A (2001) Nepotism and sexism in peer-review. *Women, Science, and Technology*, eds Wyer M, Barbercheck M, Cookmeyer D, Ozturk H, Wayne M (Routledge, London), pp 46–52.
10. Bartneck C, Kokkelmans S (2010) Detecting h-index manipulation through self-citation analysis. *Scientometrics* 87:85–98.
11. Zhivotovsky L, Krutovsky K (2008) Self-citation can inflate h-index. *Scientometrics* 77:373–375.
12. Adam D (2002) Citation analysis: The counting house. *Nature* 415:726–729.
13. Uzzi B, Mukherjee S, Stringer M, Jones B (2013) Atypical combinations and scientific impact. *Science* 342:468–472.
14. Whalen R et al. (2016) Citation distance: Measuring changes in scientific search strategies. *Companion to the 25th International World Wide Web Conference (WWW 2016)* (International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland), pp 419–423.
15. Stringer MJ, Sales-Pardo M, Amaral LAN (2008) Effectiveness of journal ranking schemes as a tool for locating information. *PLoS One* 3:e1683.
16. MacRoberts MH, MacRoberts BR (2010) Problems of citation analysis: A study of uncited and seldom-cited influences. *J Assoc Inf Sci Technol* 61:1–12.
17. Wang D, Song C, Barabási AL (2013) Quantifying long-term scientific impact. *Science* 342:127–132.
18. Althouse BM, West JD, Bergstrom CT, Bergstrom T (2009) Differences in impact factor across fields and over time. *J Am Soc Inf Sci Technol* 60:27–34.
19. Harzing AW, Alakangas S, Adams D (2014) hIa: An individual annual h-index to accommodate disciplinary and career length differences. *Scientometrics* 99:811–821.
20. Blei DM, Lafferty J (2006) Dynamic topic models. *Proceedings of the 23rd International Conference on Machine Learning (ICML 2006)* (International Machine Learning Society, La Jolla, CA), pp 113–120.
21. Wang X, McCallum A (2006) Topics over time: A non-Markov continuous-time model of topical trends. *Proceedings of the 12th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD 2006)* (ACM Press, New York), pp 424–433.
22. Gerrish S, Blei DM (2010) A language-based approach to measuring scholarly impact. *Proceedings of the 27th International Conference on Machine Learning (ICML 2010)* (International Machine Learning Society, La Jolla, CA), pp 375–382.
23. Bird S et al. (2008) The ACL Anthology Reference Corpus: A reference dataset for bibliographic research in computational linguistics. *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC 2008)* (European Language Resources Association, Paris), pp 1755–1759.
24. Blei DM, Ng AY, Jordan MI (2003) Latent Dirichlet allocation. *J Mach Learn Res* 3:993–1022.
25. Blei DM (2012) Probabilistic topic models. *Commun ACM* 55:77–84.
26. Griffiths TL, Steyvers M (2004) Finding scientific topics. *Proc Natl Acad Sci USA* 101:5228–5235.
27. Chang J, Gerrish S, Wang C, Boyd-graber JL, Blei DM (2009) Reading tea leaves: How humans interpret topic models. *Advances in Neural Information Processing Systems*, eds Bengio Y, Schuurmans D, Lafferty J, Williams C, Culotta A (Neural Information Processing Systems Foundation, La Jolla, CA), Vol 22, pp 288–296.
28. Wang C, Blei D, Heckerman D (2008) Continuous time dynamic topic models. *Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence (UAI 2008)* (AUAI Press, Corvallis, OR), pp 579–586.
29. Shwed U, Bearman PS (2010) The temporal structure of scientific consensus formation. *Am Sociol Rev* 75:817–840.
30. Wallach HM, Murray I, Salakhutdinov R, Mimno D (2009) Evaluation methods for topic models. *Proceedings of the 26th International Conference on Machine Learning (ICML 2009)* (Sage Publications, Los Angeles), pp 1105–1112.
31. Orbanz P, Teh YW (2011) Modern Bayesian nonparametrics. *Advances in Neural Information Processing 24*, eds Shawe-Taylor J, Zemel RS, Bartlett PL, Pereira F, Weinberger KQ (Neural Information Processing Systems Foundation, La Jolla, CA).
32. Orbanz P, Teh YW (2011) Bayesian nonparametric models. *Encyclopedia of Machine Learning*, eds Sammut C, Webb GI (Springer, Berlin), pp 81–89.
33. Bloch F (1946) Nuclear induction. *Phys Rev* 70:460–474.
34. Wallace PR (1947) The band theory of graphite. *Phys Rev* 71:622–634.
35. Ke Q, Ferrara E, Radicchi F, Flammini A (2015) Defining and identifying sleeping beauties in science. *Proc Natl Acad Sci USA* 112:7426–7431.
36. Marcus MP, Marcinkiewicz MA, Santorini B (1993) Building a large annotated corpus of English: The Penn treebank. *Comput Linguist* 19:313–330.
37. Cagle FR (1939) A System of marking turtles for future identification. *Copeia* 1939:170–173.
38. Perrins CM (1965) Population fluctuations and clutch-size in the great tit, Parus major L. *J Anim Ecol* 34:601–647.
39. Peltzman S (1976) Toward a more general theory of regulation. *J Law Econ* 19:211–240.


SOCIAL SCIENCES

APPLIED PHYSICAL SCIENCES