# Modeling Scientific Impact with Topical Influence Regression

**James Foulds** **Padhraic Smyth**
Department of Computer Science
University of California, Irvine
{jfoulds, smyth}@ics.uci.edu

## Abstract

When reviewing scientific literature, it would be useful to have automatic tools that identify the most influential scientific articles as well as how ideas propagate between articles. In this context, this paper introduces *topical influence*, a quantitative measure of the extent to which an article tends to spread its topics to the articles that cite it. Given the text of the articles and their citation graph, we show how to learn a probabilistic model to recover both the degree of topical influence of each article and the influence relationships between articles. Experimental results on corpora from two well-known computer science conferences are used to illustrate and validate the proposed approach.

## 1 Introduction

Scientific articles are not created equal. Some articles generate entire disciplines or sub-disciplines of research, or revolutionize how we think about a problem, while others contribute relatively little. When we are first introduced to a new area of scientific study, it would be useful to automatically find the most important articles, and the relationships of influence between articles. Understanding the impact of scientific work is also crucial for hiring decisions, allocation of funding, university rankings and other tasks that involve the assessment of scientific merit. If scientific works stand on the shoulders of giants, we would like to be able to find the giants.

The importance of a scientific work has previously been measured chiefly through metrics derived from citation counts, such as impact factors. However, citation counts are not the whole story. Many citations are made in passing, are relevant to only one section of an article, or make no impact on a work but are referenced out of "politeness, policy or piety" (Ziman, 1968). In reality, scientific impact has many dimensions. Some articles are important because they describe scientific discoveries that alter our understanding of the world, while some develop essential tools and techniques which facilitate future research. Other articles are influential because they introduce the seeds of new ideas, which in turn inspire many other articles.

In this work we introduce *topical influence*, a quantitative metric for measuring the latter type of scientific influence, defined in the context of an unsupervised generative model for scientific corpora. The model posits that articles "coerce" the articles that cite them into having similar topical content to them. Thus, articles with higher topical influence have a larger effect on the topics of the articles that cite them. We model this influence mechanism via a regression on the parameters of the Dirichlet prior over topics in an LDA-style topic model. We show how the models can be used to recover meaningful influence scores, both for articles and for specific citations. By looking not just at the citation graph but also taking into account the content of the articles, topical influence can provide a better picture of scientific impact than simple citation counts.

## 2 Background

Bibliometrics, the quantitative study of scientific literature, has a long history. One example of a widely-used bibliometric measure of interest is the *impact factor* of a publication venue for a given year, defined to be the average number of times articles from

that venue, published in the previous two years, were cited in that year. However, the quality of articles in a given publication venue can vary wildly, and it is difficult to compare impact factors between different disciplines of study. The number of citations an article receives is an indication of importance, but this is confounded by the unknown function of each citation. Measures of importance such as PageRank (Brin and Page, 1998) can be derived recursively from the citation graph. Such graph-based measures do not in general make use of the textual content of the articles, although it is possible to apply them to graphs where the edges between articles are determined based on the similarity of their content instead of the citation graph (Lin, 2008).

A variety of methods have previously been proposed for analyzing text and citation links together, such as modeling connections between words and citations Cohn and Hofmann (2001), classifying citation function (Teufel et al., 2006), and jointly modeling citation links and document content (Chang and Blei, 2009). However, these methods do not directly measure article importance or influence relationships between articles given their citations.

More closely related to the present work, Dietz et al. (2007) proposed the citation influence model (CIM). Building on the latent Dirichlet allocation (LDA) framework, CIM assumes that each word is drawn by first selecting either (a) the distribution over topics of a cited article (with probability proportional to the influence weight of that article on the present article) or (b) a novel topic distribution, and drawing a topic from the selected distribution, then finally drawing the word from the chosen topic.[1] In their approach, every word is assigned an extra latent variable, namely the cited article whose topic distribution the topic was drawn from. For the model proposed in this paper, we do not need to introduce these additional latent variables, which leads to a simpler latent representation and fewer variables to sample during inference. Dietz et al. (2007) also assume that the citation graph is bipartite, consisting of one set of citing articles and one set of cited articles—in contrast, our proposed models can handle arbitrary citation graphs in the form of directed acyclic graphs (DAGs). While both the CIM and our approach can identify the influence of specific citations between articles, our model can also infer how influential each article is overall, and provides a flexible modeling framework which can handle different assumptions about influence.

Another related method is due to Shaparenko and Joachims (2009), who propose a mixture modeling approach for the detection of novel text content. Nallapati et al. (2011) introduced TopicFlow, a PLSA-based model for the flow of topics in a document network. In their model, citing articles "vote" on each cited article's topic distribution in retrospect, via a network flow model. Since this voting occurs in time-reversed order, it does not describe an influence mechanism and is not a generative model that can simulate or predict new documents.

Finally, the document influence model of Gerrish and Blei (2010) can be viewed as orthogonal to this work, in that it models the impact of documents on *topics* over time (specifically, how topics change over time) rather than how articles influence the specific *articles* that cite them.

## 3 Topical Influence Regression

Scientific research is seldom performed in a vacuum. New research builds on the research that came before it. Although there are many aspects by which the importance of a scientific article can be judged, in this work we are interested in the extent to which a given article has or will have subsequent articles that build upon it or are otherwise inspired by its ideas. We begin by defining *topical influence*, a quantitative measure for this type of influence.

### 3.1 Topical Influence

It is not immediately obvious how one might quantify such a notion of "idea-based" influence. However, the mechanism used in the scientific community for giving credit to prior work is citation. The presence of a citation from article $b$ to article $a$ therefore indicates that article $b$ may have been influenced by the ideas in article $a$, to some unknown extent. We hypothesize that the extent of this influence manifests itself in the language of $b$. Using latent Dirichlet allocation (LDA) topics as a concrete proxy for

[1] A somewhat similar model was also proposed by He et al. (2009)

the vague notion of "ideas", we define the *topical influence* of $a$ to be the extent to which article $a$ coerces the documents which cite it to have similar topic distributions to it. Topical influence will be made precise in the context of a generative model for scientific corpora, conditioned on the citation graph, called *topical influence regression* (TIR).

The proposed model extends the LDA framework of Blei et al. (2003). In LDA, each word $w_i^{(d)}$ of each document $d$ is assigned to one of $K$ latent topics, $z_i^{(d)}$. Each topic $\Phi^{(k)}$ is a discrete distribution over words. Document $d$ has a distribution over topics $\theta^{(d)}$, which can be viewed as a "location in topic space" summarizing its thematic content. The $\theta^{(d)}$'s have a Dirichlet prior distribution with parameters $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_K]^\intercal$. Although the $\alpha_k$'s are often set to be equal, representing a relatively uninformative prior over the $\theta$'s, a unique $\alpha^{(d)}$ for each document can also be used to encode prior information such as the effect of other variables on the topics of that document (Mimno and McCallum, 2008). In our case, we want to model the influence that a document has on the topic distributions of the documents that cite it. A natural way to encode such influence, then, is to allow documents to affect the value of $\alpha^{(d)}$ for each document $d$ that cites them.

Accordingly, we model each article $d$ as having a latent, non-negative "topical influence" value $l^{(d)}$. Let $n^{(d)}$ be number of words in article $d$, $n_k^{(d)}$ be the number of words assigned to topic $k$, and let $C^{(d)}$ be the set of articles that $d$ cites. We model $\alpha^{(d)}$ as

$$\alpha^{(d)} = \sum_{c \in C^{(d)}} l^{(c)} \bar{z}^{(c)} + \alpha \,, \qquad (1)$$

where $\bar{z}^{(c)} = \frac{1}{n^{(c)}}[n_1^{(c)}, \ldots, n_K^{(c)}]^\intercal$ is the normalized histogram of topic counts for document $c$, and $\alpha$ is a constant for smoothing. Since the $\bar{z}^{(c)}$'s sum to one, the topical influence $l^{(c)}$ of article $c$ can be interpreted as the number of words of precision that it adds to the prior of the topic distributions of each document that cites it. As we increase $l^{(c)}$, the articles that cite $c$ become more likely to have similar topic proportions to it. Thus, $l^{(c)}$ encodes the degree to which article $c$ influences the topics of each of the articles that cite it.

From another perspective, marginalizing out $\theta^{(d)}$, we can view the topic counts (in the standard LDA model) for document $d$ as being drawn from a Polya urn scheme with $\alpha_k^{(d)}$ (possibly fractional) balls of each color $k \in \{1, \ldots, K\}$ initially in the urn. For each word, a ball is drawn randomly from the urn and the topic assignment is determined according to its color $k$. The ball is replaced in the urn, along with a new ball of color $k$. In our model, for each article $c$ cited by article $d$ we place $l^{(c)}$ balls, with colors distributed according to $\bar{z}^{(c)}$, into article $d$'s urn initially. Thus, article $d$'s topic assignments are more likely to be similar to those of the more influential articles that it cites. The total number of balls that $d$ added to other articles' urns,

$$T^{(d)} \triangleq \sum_{b: d \in C^{(b)}} l^{(d)} = l^{(d)} \left| \{b : d \in C^{(b)}\} \right| \qquad (2)$$

measures the total impact (in a topical sense) of the article. We refer to this as *total topical influence*.

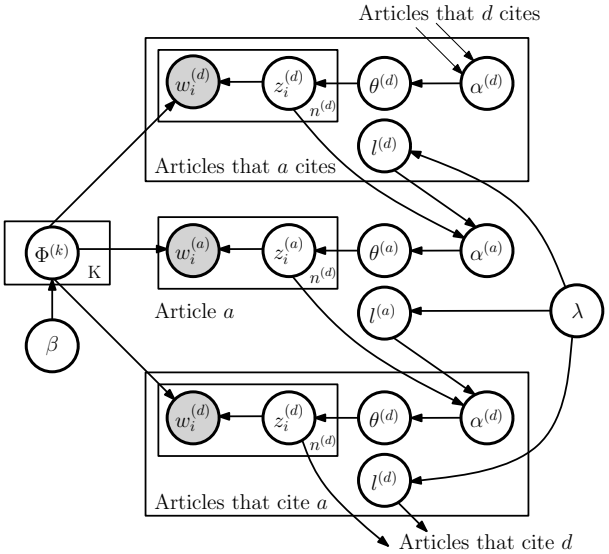


Figure 1: The graphical model for the portion of the TIR model connected to article $a$ (the links from the $z$'s and $l$'s to the $\alpha^{(d)}$'s are deterministic).

### 3.2 Generative Model for Topical Influence Regression

The full assumed generative process for articles in this model begins with a directed acyclic citation graph $G = \{V, E\}$. Intuitively, citation graphs are typically DAGs because articles can normally only cite articles that precede them in time. We assume that $G$ is a DAG so that influence relationships are

consistent with some temporal ordering of the articles, and so that the resulting model is a Bayesian network. Here, each vertex $v_i$ corresponds to an article $d_i$, edge $e = (v_1, v_2) \in E$ IFF $d_1$ is cited by $d_2$, and vertices (articles) are numbered in a topological ordering with respect to $G$. Such an ordering exists because $G$ is a DAG. We model each article $d$'s word vector $w^{(d)}$ as being generated in topological sequence, similarly to LDA but with its prior over topic distribution being Dirichlet($\alpha^{(d)}$), as given by Equation 1. Note that each $\alpha^{(d)}$ is a function of the topics of the documents that it cites, parameterized by their topical influence values. We therefore call this model *topical influence regression* (TIR).

The TIR model provides us with topical influence scores for each article, but it does not tell us about topical influence relationships between specific pairs of cited and citing articles. To model such relationships, we can consider a hierarchical extension to TIR, with edge-wise topical influences $l^{(c,d)}$ for each edge $(c, d)$ of the citation graph, $l^{(c,d)} \sim$ TruncGaussian($l^{(c)}, \sigma, l^{(c,d)} \geq 0$). In this case,

$$\alpha^{(d)} = \sum_{c \in C^{(d)}} l^{(c,d)} \bar{z}^{(c)} + \alpha \, . \tag{3}$$

This hierarchical setup allows us to continue to infer article-level topical influences, and provides a mechanism for sharing statistical strength between influences associated with one cited article. We shall refer to the model with influences on just the nodes (articles) as TIR, and the hierarchical extension with influences on the edges as TIRE. The graphical model for TIR is given in Figure 1, and the generative process is detailed in the following pseudocode:

- For each topic $k$
  - Sample the topic $\Phi^{(k)} \sim$ Dirichlet($\beta$)
- For each document $d$, in topological order
  - Sample an influence weight, $l^{(d)} \sim$ Exponential($\lambda$)
  - If using the TIRE model
    - For each cited document $c \in C^{(d)}$
      - Draw edge influence weight, $l^{(c,d)} \sim$ TruncGauss($l^{(c)}, \sigma, l^{(c,d)} \geq 0$)
  - Assign a prior over topics via $\alpha^{(d)} = \sum_{c \in C^{(d)}} l^{(c)} \bar{z}^{(c)} + \alpha$ (TIR), or $\alpha^{(d)} = \sum_{c \in C^{(d)}} l^{(c,d)} \bar{z}^{(c)} + \alpha$ (TIRE)
  - Sample a distribution over topics, $\theta^{(d)} \sim$ Dirichlet($\alpha^{(d)}$)
  - For each word $i$ in document $d$
    - Sample a topic $z_i^{(d)} \sim$ Discrete($\theta^{(d)}$)
    - Sample a word $w_i^{(d)} \sim$ Discrete($\Phi^{(z_i^{(d)})}$)

### 3.3 Relationship to Dirichlet-Multinomial Regression

The TIR model can be viewed as an adaption of the Dirichlet-multinomial regression (DMR) framework of Mimno and McCallum (2008) to model topical influence. DMR also endows each document with its own unique $\alpha^{(d)}$, but with $\alpha_k^{(d)} = \exp(x^{(d)\intercal} \lambda_k)$ being a function of the observed feature vector $x^{(d)}$ parameterized by regression coefficients $\lambda$. The DMR model can also be applied to text corpora with citation information, by setting the feature vectors to be binary indicators of the presence of a citation to each article. TIR differs in that the functional form of the regression is parameterized in a way that directly models influence, and also differs in that the regression takes advantage of the content of the cited articles via their topic assignments.

Because an article's prior over topic distributions depends on the topic assignments of the articles that it cites, TIR induces a network of dependencies between the topic assignments of the documents. Specifically, if we collapse out $\Theta$, the dependencies between the $z$'s of each document form a Bayesian network whose graph is the citation graph. In contrast, DMR treats the documents as conditionally independent given their citations, and does not exploit their content in the regression.

To illustrate this, Figure 2 shows an example citation graph and the resulting Bayesian network. In the figure, an edge in (a) from $c$ to $d$ corresponds to a citation of $c$ by $d$. Conditioned on the topics, the dependence relationships between $z$ nodes in (b) follow the same structure as the citation graph.

## 4 Inference

We perform inference using a Markov chain Monte Carlo technique. We use a collapsed Gibbs sampling approach analogous to Griffiths and Steyvers (2004), integrating out $\Theta$ and

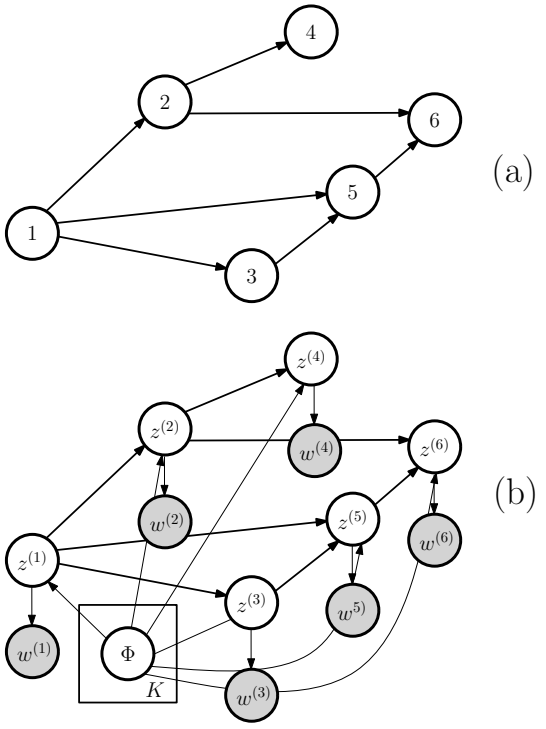


Figure 2: (a) An example citation network. (b) Graphical model for TIR on the example network, collapsing out $\Theta$ but retaining topics $\Phi$. Influence variables and hyperparameters not shown for simplicity.

$\Phi$. The update equation for the topic assignments is

$$Pr(z_i^{(d)} = k | z^{-(d,i)}, \ldots) \propto (n_k^{(d)-(d,i)} + \alpha_k^{(d)}) \frac{n_k^{(w_i^{(d)})-(d,i)} + \beta_{w_i^{(d)}}}{n_k^{-(d,i)} + \sum_w \beta_w} \times \prod_{d':d \in C(d')} \text{Polya}(z^{(d')} | \alpha^{(d')} : z_i^{(d)} = k, z^{-(d,i)}, l) \quad (4)$$

where the $n_k$'s are the counts of the occurrences of topic $k$ over all of the entries determined by the superscript. The $-(d,i)$ superscript indicates excluding the current assignment for $z_i^{(d)}$. The update equation is similar to the update equations of Griffiths and Steyvers, but with a different $\alpha$ for each document $d$, and with multiplicative weights for each document that cites it. These weights $\text{Polya}(z^{(d)}|\alpha^{(d)})$ are the likelihood for a multivariate Polya (a.k.a. Dirichlet-multinomial) distribution,

$$\text{Polya}(z^{(d)}|\alpha^{(d)}) = \frac{\Gamma(\sum_k \alpha_k^{(d)})}{\Gamma(n^{(d)} + \sum_k \alpha_k^{(d)})} \prod_k \frac{\Gamma(n_k^{(d)} + \alpha_k^{(d)})}{\Gamma(\alpha_k^{(d)})}.$$

In the case of TIR, in the collapsed model the full conditional posterior for the topical influence values $l$ is $Pr(l|z,\lambda) \propto Pr(z|l)Pr(l|\lambda)$. Here, $Pr(z|l) = \prod_{d=1}^{D} \text{Polya}(z^{(d)}|l^{C^{(d)}}, z^{C^{(d)}})$. The topical influence values $l$ can be sampled using Metropolis-Hastings updates, or slice sampling. An alternative is to perform stochastic EM, optimizing the likelihood or the posterior probability of $l$, interleaved within the Gibbs sampler, as in Mimno and McCallum (2008) and Wallach (2006). In experiments on synthetic data we found that maximum likelihood updates on $l$, obtained via gradient ascent, resulted in the lowest L1 error from the true $l$, so we use this strategy for the experimental results in this paper. The derivative of the log-likelihood with respect to the topical influence $l^{(a)}$ of article $a$ is

$$\frac{\mathrm{d}Pr(z|l)}{\mathrm{d}l^{(a)}} = \sum_{d:a \in C(d)} \Big( \Psi(\sum_k \sum_{c \in C^{(d)}} l^{(c)} \bar{z}_k^{(c)} + K\alpha) - \Psi(\sum_k \sum_{c \in C^{(d)}} l^{(c)} \bar{z}_k^{(c)} + K\alpha + n^{(d)}) \Big) + \sum_{d:a \in C(d)} \sum_{k=1}^{K} \bar{z}_k^{(a)} \Big( \Psi(\sum_{c \in C^{(d)}} l^{(c)} \bar{z}_k^{(c)} + \alpha + n_k^{(d)}) - \Psi(\sum_{c \in C^{(d)}} l^{(c)} \bar{z}_k^{(c)} + \alpha) \Big),$$

where $\Psi(.)$ is the digamma function. For TIRE, the likelihood decomposes across documents and we can optimize the incoming edge weights for each document separately. We have

$$\frac{\mathrm{d}Pr(z^{(d)}|l)}{\mathrm{d}l^{(a,d)}} = \Psi(\sum_k \sum_{c \in C^{(d)}} l^{(c,d)} \bar{z}_k^{(c)} + K\alpha) - \Psi(\sum_k \sum_{c \in C^{(d)}} l^{(c,d)} \bar{z}_k^{(c)} + K\alpha + n^{(d)}) + \sum_{k=1}^{K} \bar{z}_k^{(a)} \Big( \Psi(\sum_{c \in C^{(d)}} l^{(c,d)} \bar{z}_k^{(c)} + \alpha + n_k^{(d)}) - \Psi(\sum_{c \in C^{(d)}} l^{(c,d)} \bar{z}_k^{(c)} + \alpha) \Big).$$

We optimize the node-level $l$'s in TIRE via the least squares estimate (LSE), $\hat{l}^{(a)} = \frac{1}{|\{d:a \in C^{(d)}\}|} \sum_{d:a \in C^{(d)}} l^{(a,d)}$. Although the LSE for the mean of a truncated Gaussian is biased, it is widely used as it is more robust than the MLE (A'Hearn, 2004).

## 5 Experimental Analysis

In this section we experimentally investigate the properties of TIR and TIRE. We consider two scientific corpora: a collection of 3286 of articles from the Association for Computational Linguistics (ACL) conference[2] (Radev et al., 2009) published between 1987 and 2011, and a corpus of articles from the Neural Information Processing Systems (NIPS) conference[3] containing 1740 articles from 1987 to 1999. The corpora both contained a small number (53, and 14, respectively) of citation graph loops due to insider knowledge of simultaneous publications. Some loops were removed by manual deletion of "insider knowledge" edges, and others were removed by deleting edges in the loop uniformly at random. For computational efficiency, we performed approximate Gibbs updates where we drop the multiplicative Polya likelihood terms in Equation 4. This corresponds to only transmitting influence information downward in the citation DAG, but not transmitting "reverse influence" information upwards. Preliminary experiments on synthetic data indicated that this did not significantly impact the ability of the model to recover the topical influence weights. As one might expect, LDA is already capable of inferring topic distributions which are good enough to perform the regression on, without fully exploiting the additional feedback from the regression. This algorithm has a similar running time to the standard collapsed Gibbs sampler for LDA, as the regression step is not a bottleneck.

In all experiments, we set the hyper-parameters to $\alpha = 0.1, \beta = 0.1$ and the $\sigma$ parameter for the truncated Gaussian in TIRE to be 1. We interleaved regression steps every 10 Gibbs iterations. For exploratory data analysis experiments the models were trained for 500 burn-in iterations, and the samples from the final iterations were used for the analysis.

[2] http://clair.eecs.umich.edu/aan/

[3] http://www.arbylon.net/resources.html, published by Gregor Heinrich and based on an earlier collection due to Sam Roweis.

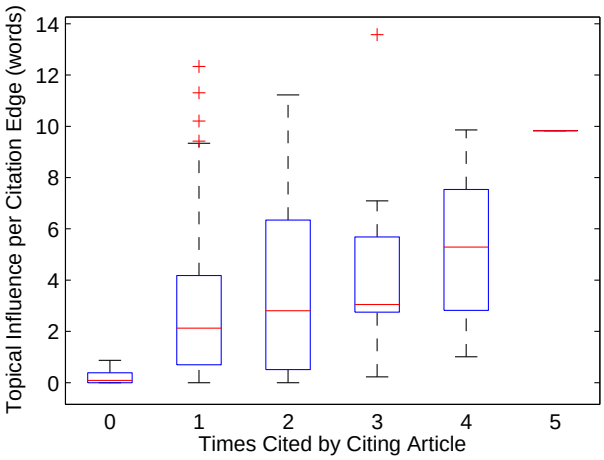


Figure 3: Topical influence per edge versus number of times cited by the citing article (NIPS). Several articles had zero in-text citations due to author or dataset errors.

### 5.1 Model Validation using Metadata

It is not immediately obvious how to best validate an unsupervised model of citation influence. Ground truth is not well-defined and human evaluation requires extensive knowledge of the individual papers in the corpora. With this in mind, we explore how topical influence scores relate to document metadata, which serves as a proxy for ground truth.

In many cases, if article $c$ is repeatedly cited in the text of article $d$ it may indicate that $d$ builds heavily on $c$. We would therefore expect to see an association between repeated citations and edge-wise topical influence $l^{(c,d)}$. For each of the 106 papers in the NIPS corpus with at least three distinct references, we counted the number of repeated citations for the most influential and least influential references according to the TIRE model (Figure 3). Overall, the "most influential" references were cited 171 times in the text of their citing articles, while the "least influential" references were cited 128 times. Of the 45 articles where the counts were not tied, the most influential references had the higher citation counts 33 times. A sign test rejects the null hypothesis that the median difference in citation counts between least and most influential references is zero at $\alpha = 0.05$, with $p$-value $\approx 5 \times 10^{-4}$.

Self-citations, where at least one author is in common between cited and citing articles, are also informative (Figure 4). Authors often build upon their own work, so we would expect self-citations to have higher edge-wise topical influence on average. For ACL the mean topical influence for a self citation edge is 2.80 and for a non-self citation is 1.40. For NIPS the means are 5.05 (self) and 3.15 (non-self). A two-sample t-test finds these differences are both significant at $\alpha = 0.05$.

### 5.2 Prediction Experiments

We also used a document prediction task to explore whether the posited latent structure is predictively useful. We selected roughly 10% of the articles in each corpus (170 and 330 documents for NIPS and ACL, respectively) for testing, chosen among the articles that made at least one citation. We held out a randomly selected set of 50% of their words and evaluated the log probability of the held out partial documents under each model. This is equivalent to evaluating on a set of new documents with the same set of references as the held out set. Evaluation was performed using annealed importance sampling (Neal, 2001), as in Wallach et al. (2009) except we used multiple samples per likelihood computation.

The TIR models were compared to LDA and an "additive" version of DMR with link function $\alpha_k^{(d)} = x^{(d)\intercal}\lambda_k + \alpha$, where the $\lambda$s were constrained to be positive and given an exponential prior with mean one. For DMR, binary feature vectors encoded the presence or absence of each possible citation. For each algorithm, we burned in for 250 iterations, then executed 1000 iterations, optimizing topical influence weights/DMR parameters every 10th iteration. Held-out log probability scores were computed by performing AIS with every 100th sample, and averaging the results to estimate the posterior predictive probability $Pr(\text{held out article}|\text{training set, citations, model})$.

It was found that all of the regression methods had superior predictive performance to LDA on these corpora, demonstrating that topical influence has predictive value (Table 1). Although DMR performed slightly better than TIR predictively, TIR was competitive despite the fact that it has a factor of $K$ less regression parameters. Note that DMR does not provide an interpretable notion of influence.

### 5.3 Exploring Topical Influence

In this section we explore the inferred topical influence scores $l^{(d)}$, total topical influence scores $T^{(d)}$ and edgewise topical influence scores $l^{(c,d)}$ (recall their definitions in Equations 1, 2 and 3, respectively). Table 2 shows the most influential articles in the ACL corpus, according to citation counts, topical influence and total topical influence (the latter two inferred with the TIR model). The most frequently cited paper within the ACL corpus, written by Papineni et al., introduces BLEU, a technique for evaluating machine translation (MT) systems.[4] This paper is of great importance to the computational linguistics community because the method that it introduces is widely used to validate MT systems. However, the BLEU article has a relatively low *topical* influence value of 0.58, consistent with the fact that most of the papers that cite it use the technique as part of their *methodology* but do not *build upon its ideas*. We emphasize that topical influence measures a specific dimension of scientific importance, namely the tendency of an article to influence the ideas (as mediated by the topics) of citing articles; papers with low topical influence such as the BLEU article may be important for other reasons.

Ranking papers by their influence weights $l^{(d)}$ (Table 2, middle) has the opposite difficulty to ranking by citation counts — the papers with the highest topical influence were typically cited only once, by the same authors. This makes sense, given what the model is designed to do. The lone citing papers were certainly topically influenced by these articles.

A more useful metric, however, is the total topical influence $T^{(d)}$ (the bottom sub-table in Table 2). This is the total number of words of prior concentration, summed over all of its citers, that the article has contributed, and is a measure of the total corpus-wide topical influence of the paper. This metric ranks the BLEU paper at 5th place, down from 1st place by citation count. The ACL paper with the highest total topical influence, by David Chiang, won the ACL best paper award in 2005.

The behavior of the different metrics is echoed in the NIPS corpus (Table 3). The most cited paper, "Handwritten Digit Recognition," by

[4] Citations within the corpora are of course only a small fraction of the total set of citations for many of these papers.

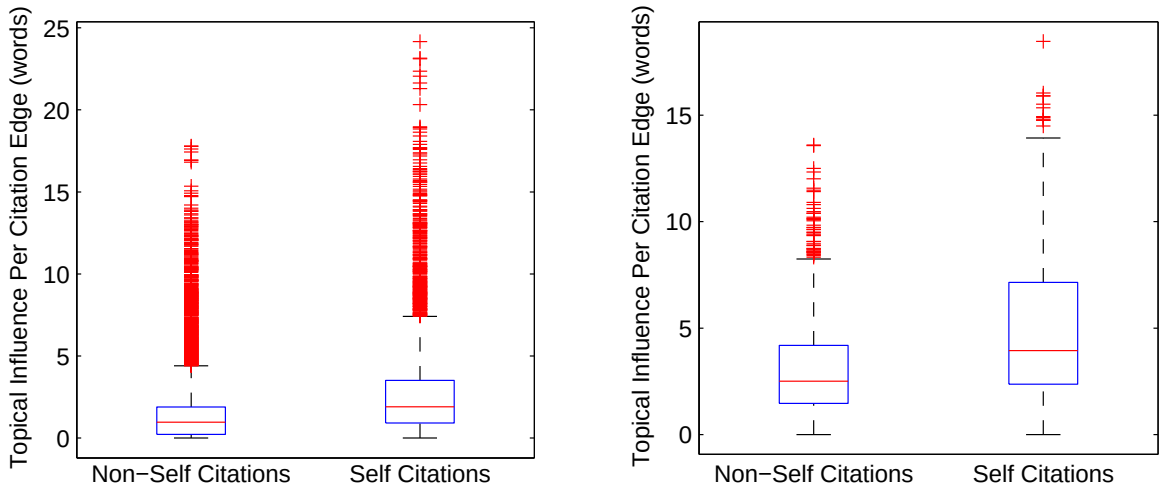


Figure 4: Topical influence for self and non-self citation edges. Left: ACL. Right: NIPS.

| | ACL | | | NIPS | | |
|---|---|---|---|---|---|---|
| | Wins | Losses | Average Improvement | Wins | Losses | Average Improvement |
| TIR | 297 | 33 | 65.7 | 150 | 20 | 38.2 |
| TIRE | 276 | 54 | 63.0 | 148 | 22 | 38.7 |
| DMR | 302 | 28 | 79.1 | 157 | 13 | 48.4 |

Table 1: Wins, losses and average improvement for log probabilities of held-out articles, versus LDA. Each "Win" corresponds to the model assigning a higher log probability score for the test portion of a held-out document than LDA assigned to that document.

Le Cun et al. (1990), is an early successful application of neural networks. The paper does not introduce novel models or algorithms, but rather, in the authors' words, "show[s] that large back propagation (BP) networks can be applied to real image recognition problems." Thus, although it is has an important role as a landmark neural network success story, it does not score highly in terms of *topical* influence. This paper is ranked 13th according to total topical influence, with a score of 1.6. The top two-ranked papers according to total topical influence, on Gaussian Process Regression and POMDPs respectively, were both seminal papers that spawned large bodies of related work. An interesting case is the third-ranked paper in the NIPS corpus, by Wang et al., on the theory of early stopping. It is only referenced three times, but has a very high topical influence of 19.3 words. All three citing papers are also on the theory of early stopping, and one of the papers, by Wang and Venkatesh, directly extends a theoretical result of this paper. Although it is easy to see why this paper scores highly on topical influence, in this case the metric has perhaps overstated its importance. A limitation of topical influence is that it can potentially give more credit than is due when an article is cited by a small number of topically similar papers, due to overfitting. This is likely to be an issue for any topic-based approach for modeling scientific influence. However, topics help to absorb lexical ambiguity and author-specific idiosyncracies, mitigating the problem relative to word-based approaches.

Using the TIRE model, we can also look at influence relationships between pairs of articles. Tables 4 and 5 show the most and least topically influential references, and the most and least influenced citing papers, for three example articles from ACL and NIPS, respectively. The model correctly assigns higher influence scores along the edges to and from relevant documents. For the ACL papers, the BLEU algorithm's article is inferred to have zero topical influence on Chiang's paper, consistent with its role

| | |
|---|---|
| **Top 5 Articles by Citation Count** | |
| 140 | BLEU: a Method for Automatic Evaluation of Machine Translation. K. Papineni, S. Roukos, T. Ward, W. Zhu. |
| 105 | Minimum Error Rate Training in Statistical Machine Translation. F. Och. |
| 64 | A Hierarchical Phrase-Based Model for Statistical Machine Translation. D. Chiang. |
| 64 | Accurate Unlexicalized Parsing. D. Klein, C. Manning. |
| 59 | Unsupervised Word Sense Disambiguation Rivaling Supervised Methods. D. Yarowsky. |
| **Top 5 articles by Topical Influence** | |
| 11.38 | Refining Event Extraction through Cross-document Inference. H. Ji, R. Grishman. |
| 11.37 | Bayesian Learning of Non-compositional Phrases with Synchronous Parsing. H. Zhang, C. Quirk, R. Moore, D. Gildea. |
| 10.48 | A Plan Recognition Model for Clarification Subdialogues. D. Litman, J. Allen. |
| 10.38 | PCFGs with Syntactic and Prosodic Indicators of Speech Repairs. J. Hale et al. |
| 10.30 | Referring as Requesting, P. Cohen |
| **Top 5 Articles by Total Topical Influence** | |
| 111.46 (1.74 $\times$ 64) | A Hierarchical Phrase-Based Model for Statistical Machine Translation. D. Chiang. |
| 101.12 (6.74 $\times$ 15) | Maximum Entropy Based Phrase Reordering Model for Statistical Machine Translation. D. Xiong, Q. Liu, S. Lin. |
| 98.56 (5.80 $\times$ 17) | A Logical Semantics for Feature Structures. R. Kasper, W. Rounds. |
| 85.15 (2.18 $\times$ 39) | Discriminative Training and Maximum Entropy Models for Statistical Machine Translation. F. Och, H. Ney |
| 81.82 (0.58 $\times$ 140) | BLEU: a Method for Automatic Evaluation of Machine Translation, K. Papineni, S. Roukos, T. Ward, and W. Zhu. |

Table 2: Most influential articles in the ACL Conference corpus, according to citation counts (top), topical influence $l^{(d)}$ inferred by TIR (middle), and total topical influence $T^{(d)}$ inferred by TIR (bottom). For total topical influence, the breakdown of $T^{(d)} = l^{(d)} \times$ citation count is shown in parentheses.

| | |
|---|---|
| **Top 5 Articles by Citation Count** | |
| 26 | Handwritten Digit Recognition with a Back-Propagation Network. Y. Le Cun, et al. |
| 19 | Optimal Brain Damage. Y. Le Cun, J. Denker, S. Solla. |
| 17 | A New Learning Algorithm for Blind Signal Separation. S. Amari, A. Cichocki, H. Yang. |
| 17 | Efficient Pattern Recognition Using a New Transformation Distance. P. Simard, Y. Le Cun, J. Denker. |
| 14 | The Cascade-Correlation Learning Architecture. S. Fahlman, C. Lebiere. |
| **Top 5 articles by Topical Influence** | |
| 29.7 | Synchronization and Grammatical Inference in an Oscillating Elman Net. B. Baird, T. Troyer, F. Eeckman. |
| 26.3 | Learning the Solution to the Aperture Problem for Pattern Motion with a Hebb Rule. M. Sereno. |
| 25.9 | ALVINN: An Autonomous Land Vehicle in a Neural Network. D. Pomerleau. |
| 25.1 | Some Estimates of Necessary Number of Connections and Hidden Units for Feed-Forward Networks. A. Kowalczyk. |
| 24.7 | Complex- Cell Responses Derived from Center-Surround Inputs: The Surprising Power of Intradendritic Computation. B. Mel, D. Ruderman, K. Archie. |
| **Top 5 Articles by Total Topical Influence** | |
| 84.7 (10.6 $\times$ 8) | Gaussian Processes for Regression. C. Williams, C. Rasmussen. |
| 63.9 (7.1 $\times$ 9) | Reinforcement Learning Algorithm for Partially Observable Markov Decision Problems. T. Jaakkola, S. Singh, M. Jordan. |
| 57.9 (19.3 $\times$ 3) | Optimal Stopping and Effective Machine Complexity in Learning. C. Wang, S. Venkatesh, J. Judd. |
| 54.7 (10.9 $\times$ 5) | Links Between Markov Models and Multilayer Perceptrons. H. Bourlard, C. Wellekens. |
| 51.2 (3.7 $\times$ 14) | The Cascade-Correlation Learning Architecture. S. Fahlman, C. Lebiere. |

Table 3: Most influential articles in the NIPS corpus, according to citation counts (top), topical influence $l^{(d)}$ inferred by TIR (middle), and total topical influence $T^{(d)}$ inferred by TIR (bottom).

| | | |
|---|---|---|
| **A Hierarchical Phrase-Based Model for Statistical Machine Translation. D. Chiang.** | | |
| Most influential reference | 1.48 | Discriminative Training and Maximum Entropy Models for Statistical Machine Translation. F. Och and H. Ney. |
| Least influential reference | 0.00 | BLEU: a Method for Automatic Evaluation of Machine Translation. K. Papineni, S. Roukos, T. Ward, W. Zhu. |
| Most influenced citer | 2.54 | Toward Smaller, Faster, and Better Hierarchical Phrase-based SMT. M. Yang, J. Zheng. |
| Least influenced citer | 0.60 | An Optimal-time Binarization Algorithm for Linear Context-Free Rewriting Systems with Fan-out Two. C. Gmez-Rodrguez, G. Satta. |
| **Unsupervised Word Sense Disambiguation Rivaling Supervised Methods. D. Yarowsky.** | | |
| Most influential reference | 2.52 | Subject-dependent Co-occurrence and Word Sense Disambiguation. J. Guthrie, L. Guthrie, Y. Wilks, H. Aidinejad. |
| Least influential reference | 0.53 | Word-sense Disambiguation using Statistical Methods. P. Brown, S. Della Pietra, V. Della Pietra, R. Mercer. |
| Most influenced citer | 1.81 | Discriminating Image Senses by Clustering with Multimodal Features. N. Loeff, C. Alm, D. Forsyth. |
| Least influenced citer | 0.00 | Semi-supervised Convex Training for Dependency Parsing. Q. Wang, D. Schuurmans, D. Lin. |
| **Accurate Unlexicalized Parsing. D. Klein, C. Manning.** | | |
| Most influential reference | 3.87 | Parsing with Treebank Grammars: Empirical Bounds, Theoretical Models, and the Structure of the Penn Treebank. D. Klein and C. Manning. |
| Least influential reference | 0.81 | Efficient Parsing for Bilexical Context-Free Grammars and Head Automaton Grammars. J. Eisner, G. Satta. |
| Most influenced citer | 1.67 | Evaluating the Accuracy of an Unlexicalized Statistical Parser on the PARC DepBank. T. Briscoe, J. Carroll. |
| Least influenced citer | 0.00 | Finding Contradictions in Text. M. de Marneffe, A. Rafferty, C. Manning. |

Table 4: Least and most influential references and citers, and the influence weights along these edges, inferred by the TIRE model for three example ACL articles.

| | | |
|---|---|---|
| **Feudal Reinforcement Learning. P. Dayan, G. Hinton** | | |
| Most influential reference | 5.47 | Memory-based Reinforcement Learning: Efficient Computation with Prioritized Sweeping. A. Moore, C. Atkeson. |
| Least influential reference | 0.00 | A Delay-Line Based Motion Detection Chip. T. Horiuchi, J. Lazzaro, A. Moore, C. Koch. |
| Most influenced citer | 3.36 | The Parti-Game Algorithm for Variable Resolution Reinforcement Learning in Multidimensional State-Spaces. A. Moore. |
| Least influenced citer | 1.71 | Multi-time Models for Temporally Abstract Planning. D. Precup, R. Sutton. |
| **Optimal Brain Damage. Y. Le Cun, J. Denker , S. Solla.** | | |
| Most influential reference | 2.82 | Comparing Biases for Minimal Network Construction with Back-Propagation. S. Hanson, L. Pratt. |
| Least influential reference | 0.15 | Skeletonization: A Technique for Trimming the Fat from a Network via Relevance Assessment. M. Mozer, P. Smolensky. |
| Most influenced citer | 3.08 | Structural Risk Minimization for Character Recognition. I. Guyon, V. Vapnik, B. Boser, L. Bottou, S. Solla. |
| Least influenced citer | 0.64 | Structural and Behavioral Evolution of Recurrent Networks. G. Saunders, P. Angeline, J. Pollack. |
| **An Input Output HMM Architecture. Y. Bengio, P. Frasconi.** | | |
| Most influential reference | 5.29 | Credit Assignment through Time: Alternatives to Backpropagation. Y. Bengio, P. Frasconi. |
| Least influential reference | 0.00 | Induction of Multiscale Temporal Structure. M. Mozer |
| Most influenced citer | 2.66 | Learning Fine Motion by Markov Mixtures of Experts. M. Meila, M. Jordan. |
| Least influenced citer | 1.47 | Recursive Estimation of Dynamic Modular RBF Networks. V. Kadirkamanathan, M. Kadirkamanathan. |

Table 5: Least and most influential references and citers, and the influence weights along these edges, inferred by the TIRE model for three example NIPS articles.

in the paper as an evaluation technique. The paper most topically influenced by Chiang's paper, written by Yang and Zheng, aims to improve upon the ideas in that paper. In the NIPS corpus, the article by Bengio and Frasconi, on recurrent neural network architectures, extends previous work by the same authors, which is correctly assigned the highest topical influence. A particularly interesting case is the paper by Dayan and Hinton, which is heavily influenced by a paper by Moore, and in turn strongly influences a later paper by Moore, thus illustrating the interplay of scientific influence between authors along the citation graph. These three papers were on reinforcement learning, while the lowest scoring reference and citer were on other subjects.

## 6 Conclusions / Discussion

This paper introduced the notion of topical influence, a quantitative measure of scientific impact which arises from a latent variable model called topical influence regression. The model builds upon the ideas of Dirichlet-multinomial regression to encode influence relationships between articles along the citation graph. By training TIR, we can recover topical influence scores that give us insight into the impact of scientific articles. The model was applied to two scientific corpora, demonstrating the utility of the method both quantitatively and qualitatively.

In future work, the proposed framework could readily be extended to model other aspects of scientific influence, such as the effects of authors and journals on topical influence, and to exploit the context in which citations occur. From an exploratory analysis perspective, it would be instructive to compare topical influence trajectories over time for different papers. This could be further facilitated by explicitly modeling the dynamics of each article's topical influence score. The TIR framework could potentially also be applicable to other application domains such as modeling how interpersonal influence affects the spread of memes via social media.

To complement TIR, it would be useful to also have systems for identifying articles which are important for alternative reasons, such as providing methodological tools and/or demonstrating important facts. Ultimately a suite of such tools could feed into a system such as Google Scholar or Citeseer. We envision that this line of work will also be useful for building visualization tools to help researchers explore scientific corpora.

## Acknowledgments

Supported by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior National Business Center contract number D11PC20155. The U.S. government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoI/NBC, or the U.S. Government.

## References


[A'Hearn2004] B. A'Hearn. 2004. A restricted maximum likelihood estimator for truncated height samples. *Economics & Human Biology*, 2(1):5–19.

[Blei et al.2003] D.M. Blei, A.Y. Ng, and M.I. Jordan. 2003. Latent Dirichlet allocation. *The Journal of Machine Learning Research*, 3:993–1022.

[Brin and Page1998] S. Brin and L. Page. 1998. The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, 30(1-7):107–117.

[Chang and Blei2009] J. Chang and D. Blei. 2009. Relational topic models for document networks. In *Artificial Intelligence and Statistics*, pages 81–88.

[Cohn and Hofmann2001] D. Cohn and T. Hofmann. 2001. The missing link-a probabilistic model of document content and hypertext connectivity. In *Advances in Neural Information Processing Systems*, pages 430–436.

[Dietz et al.2007] L. Dietz, S. Bickel, and T. Scheffer. 2007. Unsupervised prediction of citation influences. In *Proceedings of the 24th International Conference on Machine Learning*, pages 233–240.

[Gerrish and Blei2010] S. Gerrish and D.M. Blei. 2010. A language-based approach to measuring scholarly impact. In *Proceedings of the 26th International Conference on Machine Learning*, pages 375–382.

[Griffiths and Steyvers2004] T.L. Griffiths and M. Steyvers. 2004. Finding scientific topics. *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl 1):5228.

[He et al.2009] Q. He, B. Chen, J. Pei, B. Qiu, P. Mitra, and L. Giles. 2009. Detecting topic evolution in scientific literature: how can citations help? In *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, pages 957–966. ACM.

[Le Cun et al.1990] B.B. Le Cun, JS Denker, D. Henderson, RE Howard, W. Hubbard, and LD Jackel. 1990. Handwritten digit recognition with a back-propagation network. In *Advances in Neural Information Processing Systems*, pages 396–404.

[Lin2008] J. Lin. 2008. Pagerank without hyperlinks: Reranking with pubmed related article networks for biomedical text retrieval. *BMC bioinformatics*, 9(1):270.

[Mimno and McCallum2008] D. Mimno and A. McCallum. 2008. Topic models conditioned on arbitrary features with Dirichlet-multinomial regression. In *Uncertainty in Artificial Intelligence*, pages 411–418.

[Nallapati et al.2011] R. Nallapati, D. McFarland, and C. Manning. 2011. Topicflow model: Unsupervised learning of topic-specific influences of hyperlinked documents. In *International Conference on Artificial Intelligence and Statistics*, pages 543–551.

[Neal2001] R.M. Neal. 2001. Annealed importance sampling. *Statistics and Computing*, 11(2):125–139.

[Radev et al.2009] D. R. Radev, P. Muthukrishnan, and V. Qazvinian. 2009. The ACL anthology network corpus. In *Proceedings, ACL Workshop on Natural Language Processing and Information Retrieval for Digital Libraries*, pages 54–61, Singapore.

[Shaparenko and Joachims2009] B. Shaparenko and T. Joachims. 2009. Identifying the original contribution of a document via language modeling. In *Machine Learning and Knowledge Discovery in Databases*, pages 350–365. Springer.

[Teufel et al.2006] S. Teufel, A. Siddharthan, and D. Tidhar. 2006. Automatic classification of citation function. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 103–110. Association for Computational Linguistics.

[Wallach et al.2009] H.M. Wallach, I. Murray, R. Salakhutdinov, and D. Mimno. 2009. Evaluation methods for topic models. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 1105–1112. ACM.

[Wallach2006] H.M. Wallach. 2006. Topic modeling: beyond bag-of-words. In *Proceedings of the 23rd International Conference on Machine Learning*, pages 977–984. ACM.

[Ziman1968] J.M. Ziman. 1968. *Public knowledge: an essay concerning the social dimension of science*. Cambridge University Press.