# Modeling Latent Topic Interactions using Quantum Interference for Information Retrieval

Alessandro Sordoni
sordonia@iro.umontreal.ca

Jing He
hejing@iro.umontreal.ca

Jian-Yun Nie
nie@iro.umontreal.ca

DIRO, Université de Montréal
Montréal, H3C 3J7, Québec

## ABSTRACT

Recently, increasing attention has been given to a possible reinterpretation of information retrieval issues in the more general probabilistic framework offered by Quantum Theory. In this paper, we investigate the use of the well-known wave-like phenomenon of Quantum Interference for topic models such as Latent Dirichlet Allocation (LDA). We use interference effects in order to model interactions between latent topics. Our aim is to elaborate a way to build more precise document models starting from original LDA estimations. Experiments in ad-hoc retrieval show statistically significant improvements on several TREC collections.

**Categories and Subject Descriptors:** H.3.3 [Information Search and Retrieval]: Retrieval Models

**Keywords:** Quantum Interference; Topic Models.

## 1. INTRODUCTION

Latent Dirichlet Allocation (LDA) [3] is a well-known probabilistic topic model based upon the vision that documents are mixtures of topics and a topic is a probability distribution over terms. To generate a document, LDA first samples a per-document multinomial distribution over topics from a Dirichlet. Then, it repeatedly samples a topic from this multinomial and samples a term from the topic distribution. Therefore, each topic has a given probability of appearing in a document and a term has a probability of being generated by a topic. In Language Modeling (LM) framework for information retrieval, the application of LDA has been shown to be very effective [9, 10]. Indeed, LDA produces a "semantic" document model which can be used for matching queries to documents beyond the term level, thus addressing the vocabulary mismatch between queries and documents.

Under a Dirichlet, the topic proportions are nearly independent. This hampers the ability of LDA to capture topic correlations, which are often present in natural language documents. A number of hierarchical probabilistic models tried to model correlations between topics [3, 8]. However, the application of such models for IR tasks has not shown to be as effective as one would expect [10].

In LDA, each term is assumed to be generated by one latent single topic. This entails that the probability of seeing a given term in a document is computed by marginalizing over its unknown topic assignment, i.e. by applying the Law of Total Probability (LTP)[1]. This calculation does not take into account the inherent interactions between topics in generating document terms. The present work is motivated by the following observation: if two topics *war* and *oil* are well represented in a document, one would expect the term *Iraq* to have a high probability in the document model. By positing that each term is generated by one topic, i.e. by applying the LTP, the term *Iraq* has a high chance to be less represented than topic-specific terms such as *army*, *fighter*, *extraction* or *pipeline*. We propose to address this problem by modeling interactions between topics. To this end, we take inspiration from the well-known phenomenon of Quantum Interference (QI).

QI is considered one of the largest mysteries in Quantum Theory (QT) [5]. However, it does not constitute in itself a probabilistic conundrum. It can be interpreted as the result of a special probabilistic parametrization which allows for relaxing disjointness constraints in a handy way. In this work, we take inspiration from QI and modify the document model estimations obtained through the LTP by adding an *interference term* which accounts for the interactions between topics. We elaborate an analogy with one of the most known interference experiments, the double-slit experiment [5]: topics are associated to waves and a document is represented as a superposition of such waves, allowing for the appearance of interference. We do not intend to capture global topics correlations by modifying the training phase, as attempted in the works cited above, which have been shown to be ineffective in retrieval tasks [10]. Our aim is to elaborate a computationally affordable way to build more precise document models starting from original LDA estimations. In our model, the term *Iraq* is naturally boosted in the final document model because there is an interference between several topics on such term that must be taken into account.

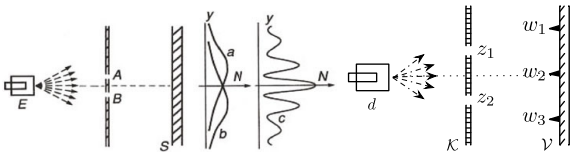
## 2. RELATED WORKS

In order to capture global correlations between topics, in [3] the Dirichlet is substituted with a logistic normal dis-

[1]The LTP states that for a collection of mutually disjoint and exhaustive events $\{e_1, \ldots, e_n\}$, $e_i \in \Omega$, the probability for an event $u \in \Omega$ is calculated as $p(u) = \sum_i p(u|e_i)p(e_i)$.


*CIKM'13,* Oct. 27–Nov. 1, 2013, San Francisco, CA, USA.

http://dx.doi.org/10.1145/2505515.2507854.




(a) (b)

**Figure 1: (a) The double-slit experiment [5]. $\mathsf{a}$ ($\mathsf{b}$) is the distribution over positions observed by closing $\mathsf{B}$ ($\mathsf{A}$). If the photon really passed through either $\mathsf{A}$ or $\mathsf{B}$, then we would expect a mixture $\hat{\mathsf{c}} = p(\mathsf{A})\mathsf{a}+(1-p(\mathsf{A}))\mathsf{b}$. Instead, the curve $\mathsf{c}$ is obtained, whose peaks and valleys correspond respectively to constructive and destructive interference. (b) Our analogy, where $N = 3$, $K = 2$.**

tribution, whose covariance matrix specifies the correlation between pair of topics. In order to model beyond-pairwise correlations, Li and McCallum [8] elaborate the Pachinko Allocation Model (PAM) and choose to represent and learn arbitrary-arity topic correlations by using a directed acyclic graph. In this model, each term is generated by a topic, sampled from a hierarchy of super-topics. These models generally allow to obtain lower perplexity results on held-out data and to discover fine-grained topics resulting in more human-readable topic distributions. However, they suffer a larger computational burden in the training phase compared to LDA. From an IR point of view, the thorough study conducted by Yi and Allen [10] shows that the increased semantic resolution of correlated topic models does not seem to increase retrieval performance over the original LDA model.

In IR, interference effects have inspired and motivated several studies. Zuccon and Azzopardi [11] proposed a novel ranking principle in which interference-like effects are used to revise the probability of relevance of a document based on the documents that have already been retrieved. By adopting a more theoretical standpoint, the work by Melucci [1] questions interference effects from a query expansion perspective and shows that taking into account interference effects is important in order to increase the retrieval effectiveness. The work by Gonzáles and Caicedo [2] is perhaps the most related to our investigation. The authors propose Quantum Latent Semantic Analysis (QLSA), a modification of the traditional Latent Semantic Analysis [6] (LSA) obtained by changing the document representation used. QLSA does not aim to explicitly model topic interactions. The authors argue that interference effects between latent dimensions could naturally arise by adopting the new document representations. It is difficult to understand how topic interactions could be modeled by merely changing document representations and without modifying the decomposition strategy. Moreover, it is not clear how interference directly affects retrieval performance. Differently from that work, we work out an explicit interference formula upon the LDA model and we apply it successfully in the retrieval phase.

## 3. AN ANALOGY

In LDA, a document model is obtained by marginalizing over topic assignments, i.e. by applying the LTP. Formally:

$$p(w|\theta_d^{LDA}) = \sum_k p(w|z=k,\Phi)p(z=k|\theta_d) = \sum_k \theta_{dk}\phi_{kw} \quad (1)$$

where $z \in \{1,\dots,K\}$ is a topic index, $w \in \{1,\dots,N\}$ a term of a vocabulary, $\theta_d = (\theta_{d1},\dots,\theta_{dK})$ the topic proportions for the document $d$ and $\Phi$ is a $N \times K$ matrix $\Phi = (\phi_1,\dots,\phi_K)$ containing the distributions over terms defined by each topic $\phi_k = (\phi_{k1},\dots,\phi_{kN})$. By Eq. 1, the probability of a term $w$ in the final document model will be within the convex hull defined by the extremal points $\phi_{kw}$, i.e. it can never be higher than $\max_k \phi_{kw}$, its maximum probability assigned by a topic distribution. This could harm terms such as *Iraq* which are not likely to be topic-specific, but important if topics such as *war* and *oil* are highly represented in $d$.

A strict analogy can be drawn with the double-slit experiment [1, 5, 11]. In short, if a photon is shot towards a barrier with two slits, the probability of detection at position $s$ on a screen behind the barrier is not the average of the probabilities of detection at the same position if it had passed through either one of the slits, i.e. as calculated with the LTP (Fig. 1a). The photon somewhat passes through both slits and interferes with itself, much like a wave. It is said that the photon propagates in a *superposition* of two waves, one scattering from each slit and traveling towards the detection screen. The amount of interference at a given position depends on the phase difference between the two waves hitting that position. Generally, the interference distribution can be written as $\mathsf{c} = \hat{\mathsf{c}} + I(\hat{\mathsf{c}})$, where $\hat{\mathsf{c}}$ is the expected distribution and $I(\hat{\mathsf{c}})$ is the interference term.

If the topics are considered as slits, the document as the photon and a term as a position behind the barrier, Eq. 1 naturally translates the classical account of the double-slit experiment, in which we compute the expected position behind the barrier $p(w|\theta_d^{LDA})$ by positing that the photon passes through only one of the slits $k$. QI effects break this straightforward prediction by putting a document in a superposition of topics such that any topic can contribute in generating a term $w$ (Figure 1b).

In order to take into account interference effects, we should (1) represent the wave scattering from each topic slit and (2) represent a document as a superposition of such waves. In the next section, these notions are translated into the quantum formalism.

## 4. TOPIC INTERFERENCE

In QT, the probabilistic picture of random experiments can be elegantly expressed within a complex Hilbert space $\mathcal{H}^n$ of dimensionality $n$ [5]. Given a unit vector $u \in \mathcal{H}^n$, $\|u\|_2 = 1$, the projector on such vector $uu^\dagger$ is called a *dyad* and is an elementary event of the quantum probabilistic space. The symbol $^\dagger$ denotes the complex conjugate transpose. In this setting, orthogonal dyads correspond to disjoint events. If $\{u_1,\dots,u_n\}$ is an orthonormal basis of $\mathcal{H}^n$, then the collection of dyads $\{u_1u_1^\dagger,\dots,u_nu_n^\dagger\}$ forms a classical sample space. A convenient orthonormal basis is the standard basis, noted by $\mathcal{E} = \{e_1,\dots,e_n\}$ where $e_i = (\delta_{i1},\dots,\delta_{in})$, $\delta_{ij} = 1$ iff $i = j$, else $\delta_{ij} = 0$.

Quantum particles such as photons are represented as *states* $v = (v_1,\dots,v_n)$, $v \in \mathcal{H}^n$, $\|v\|_2 = 1$. Each entry of such vector is a complex number that can be represented in its polar form, $v_j = |v_j|e^{i\psi_j}$, where $|v_j|$ is called *amplitude*, $e^{i\psi_j} = \cos\psi_j + i\sin\psi_j$ is the complex exponential, $i = \sqrt{-1}$, and $\psi_j$ is called *phase*. These two quantities are necessary in order to take into account the wave-behavior of quantum particles. Each state defines a probability distribution on the Hilbert space by means of the Born's rule [5]: the probability of a dyad $uu^\dagger$ given a state $v$ is the squared cosine of the angle between $v$ and $u$, i.e. $p(uu^\dagger|v) = |u^\dagger v|^2$. Note

that if the event is a member of the standard basis $\mathcal{E}$, its probability is simply the square of the amplitude of the corresponding entry in the state vector, i.e. $p(e_i e_i^\dagger|v) = |v_i|^2$.

In the case of LDA, we associate to the vocabulary sample space the standard basis in $\mathcal{H}^N$, where $N$ is the size of the vocabulary. Therefore, each term event $e_w e_w^\dagger$ corresponds to an orthogonal dimension of the vector space. In our analogy of the double-slit experiment, each topic $z \in \{1, \ldots, K\}$ corresponds to a wave defining a probability distribution over terms. This situation can be modeled by defining $K$ states $\{z_1, \ldots, z_K\}$ in $\mathcal{H}^N$. In this work, we define the states $z_k$ in order to reproduce the LDA statistics over the vocabulary sample space. Hence, we set $z_k = (z_{k1}, \ldots, z_{kN})$, where:

$$z_{kw} = \sqrt{\phi_{kw}} e^{i\psi_{kw}}, \tag{2}$$

where $\psi_{kw}$ is a free variable representing the phase of the wave corresponding to topic $k$ for term $w$. The role of such quantity will be clearer shortly when the interference formula will be introduced. Following the parameterization given in Eq. 2, one can show that the probability of a term given a topic corresponds to the LDA statistics, i.e. $p(e_w e_w^\dagger|z_k) = (\sqrt{\phi_{kw}})^2 = p(w|z = k, \Phi)$.

Until now, we assigned a clear mathematical status to each topic wave. In order to allow interference between topics, a document must be represented as a superposition of topics. Superposition is not a classical probabilistic concept. From a mathematical point of view, superposition is obtained by linear combination. Therefore, one can represent a document as a superposition of the topic states $\{z_1, \ldots, z_K\}$ by $d = \frac{1}{Z_d} \sum_k \zeta_k z_k$, where the coefficients $\zeta_k$, $\sum_k |\zeta_k|^2 = 1$, quantify how well each topic is represented in the document and the normalization factor $Z_d$ ensures that $d$ is a state defining a proper probability distribution on the underlying vector space. In this work, we choose to set $\zeta_k = \sqrt{p(z = k|\theta_d)} = \sqrt{\theta_{dk}}$, i.e. the topic proportions estimated by the traditional LDA model, and therefore $d = \frac{1}{Z_d} \sum_k \sqrt{\theta_{dk}} z_k$. The document state can be explicitly written as $d = (d_1, \ldots, d_N)$ where:

$$d_w \propto \sum_k \sqrt{\theta_{dk}\phi_{kw}} e^{i\psi_{kw}}, \tag{3}$$

up to a normalization factor. Superposition enables the topics to interact by interference effects. Indeed, the probability of a term in a given document is calculated by:

$$\begin{aligned} p(e_w e_w^\dagger|d) = |d_w|^2 &\propto |\sum_k \sqrt{\theta_{dk}\phi_{kw}} e^{i\psi_{kw}}|^2 \\ &= \sum_k \theta_{dk}\phi_{kw} + 2\sum_{i<j} \sqrt{\theta_{di}\theta_{dj}} \sqrt{\phi_{iw}\phi_{jw}} \cos(\psi_{iw} - \psi_{jw}). \end{aligned} \tag{4}$$

The equation above defines an interference document model: the first component corresponds to the classical document model given by the LTP and corresponding to Eq. 1; the second part corresponds to the interference term which boosts or penalizes the probability for term $w$ in the final document model depending on the phase differences $\psi_{iw} - \psi_{jw}$. If a pair of topics is *in phase* for a given term $w$, i.e. $\psi_{iw} - \psi_{jw} = 0$, then $\cos(\psi_{iw} - \psi_{jw}) = 1$ and the interference term will be positive: the probability of seeing $w$ in the final document model increases. Note that if $\forall w, i, j,\ \psi_{iw} - \psi_{jw} = \pi/2$, then $\cos(\psi_{iw} - \psi_{jw}) = 0$: the interference term disappears, i.e the original LDA document model is recovered.

Modeling topic interactions by setting phase differences is certainly flexible but it is challenging to determine how actually one should set or estimate such parameters. This would certainly be an interesting research direction for future works. In this study, we choose to simplify this task by assuming that (1) the interference does not depend on the particular term, i.e. $\forall w$, $\cos(\psi_{iw} - \psi_{jw}) = \cos(\psi_i - \psi_j)$ and (2) the interference is proportional to a similarity measure between topic distributions, i.e. $\cos(\psi_i - \psi_j) \propto \Delta(\phi_i, \phi_j)$. We choose to define $\Delta$ as the cosine similarity between topic distributions, i.e. $\Delta(\phi_i, \phi_j) = \frac{\phi_i^\top \phi_j}{\|\phi_i\| \|\phi_j\|}$. The assumption here is that if two topics share common terms, then they interfere positively. On the contrary, if two topics are orthogonal to each other, the interference term will vanish. Note that we are discarding the possibility of negative interference between topics, i.e. $0 \leq \Delta(\phi_i, \phi_j) \leq 1$. Introducing negative interference terms certainly increases the modeling power of the proposed model. What we often observe in practice is that two topics can be unrelated or positively related somehow. This is at least the case for the topics that we extract in LDA, in which we only consider how topics can jointly generate terms in documents, but not how a topic precludes another. The goal of this paper is not to create a complex model to account for all kinds of interference between topics. Instead, we show that when the possible interference is considered in some (albeit simplified) way, the modified model performs better than the original LDA model. For this purpose, discarding negative interference between topics appears as a reasonable simplification. The final interference document model has the form:

$$p(e_w e_w^\dagger|d) = \frac{1}{Z_d}(\sum_k \theta_{dk}\phi_{kw} + 2\sum_{i<j} \Delta(\phi_i, \phi_j) \sqrt{\theta_{di}\theta_{dj}\phi_{iw}\phi_{jw}}), \tag{5}$$

where $Z_d = 1 + 2\sum_{i<j} \sqrt{\theta_{di}\theta_{dj}} \Delta(\phi_i, \phi_j)^2$ is a normalization factor that can be computed offline. If $w$ has a high probability of being generated by two related topics, i.e. if both $\Delta(\phi_i, \phi_j)$ and $\phi_{iw}\phi_{jw}$ are large, and these topics have high probability of being present in the document, i.e. if $\theta_{di}\theta_{dj}$ is also large, then $w$ gets boosted in the final document model. Therefore, the model will penalize topic-specific terms and favor terms that are less probable but shared by important topics. With respect to our previous example, this could favor *Iraq* because it is likely to be shared by related topics such as *war* and *oil*.

# 5. EXPERIMENTAL SETUP

In our experiments, we choose three commonly used TREC newswire corpora: AP and SJMN collections containing respectively 242,918 and 90,257 documents with topics 51-150; the WSJ collection containing 173,252 documents with topics 51-100 and 151-200. These small-sized collections are often used in the literature for keeping the training process of LDA computationally affordable. The performance is measured using Mean Average Precision (MAP) and is evaluated on the top 1000 retrieved documents. Statistical significance for MAP is determined using a two-sided Fisher's randomization test with 25,000 permutations and $\alpha < 0.05$.

As done by previous work [9, 10], we rely on the query likelihood scoring function for the ad-hoc task evaluated here. We use as a first baseline the classical formula of the language model based on Dirichlet smoothing (denoted `LM`) [10]:

$$p(w|\theta_d^{LM}) = (1 - \alpha_d) p(w|\theta_d^{ML}) + \alpha_d\, p(w|\theta_C^{ML}), \tag{6}$$

| | AP | | SJMN | | WSJ | |
|---|---|---|---|---|---|---|
| $K$ | LBDM | QLBDM | LBDM | QLBDM | LBDM | QLBDM |
| 50 | .2422 | **.2521**$^{\alpha}$ | .2100 | **.2155**$^{\alpha}$ | .3073 | **.3119**$^{\alpha}$ |
| 100 | .2480 | **.2580**$^{\alpha}$ | .2240 | **.2296**$^{\alpha}$ | .3050 | **.3105**$^{\alpha}$ |
| 200 | .2578 | **.2653**$^{\alpha}$ | .2232 | **.2282**$^{\alpha}$ | .3159 | **.3205** |
| 400 | .2581 | **.2682**$^{\alpha}$ | .2237 | **.2275** | .3116 | **.3145** |
| 600 | .2624 | **.2707**$^{\alpha}$ | .2290 | **.2338**$^{\alpha}$ | .3170 | **.3210**$^{\alpha}$ |

**Table 1: Comparison of results (MAP). The symbol $^{\alpha}$ denotes statistical significance over LBDM.**

where $\theta_d^{ML}, \theta_C^{ML}$ are the maximum likelihood estimators of the document LM and collection LM and $\alpha_d = (\frac{\mu}{\mu+|d|})$ controls the amount of smoothing. As a second baseline, we report LDA-Based document models [9] (denoted LBDM), which integrate semantic information mined by LDA, i.e. $p(w|\theta_d^{LDA})$, by adding a level of smoothing over LM:

$$p(w|\theta_d^{LBDM}) = \lambda\, p(w|\theta_d^{LM}) + (1-\lambda)\, p(w|\theta_d^{LDA}), \qquad (7)$$

where $\lambda \in [0,1]$ controls the amount of semantic matching. Finally, our proposed interference model (denoted QLBDM) is obtained by substituting $p(w|\theta_d^{LDA})$ in the above equation with the interference model $p(e_w e_w^{\dagger}|d)$ calculated using Eq. 5. The free parameters are set following [10]: $\mu = 1000$ and the semantic smoothing parameter $\lambda$ is optimized by linear search over $\{0, 0.1, ..., 1\}$ for each collection, method and number of topics tested. LDA is trained by running 50 iterations of Collapsed Gibbs Sampling. The number of topics $K$ ranges in $\{50, 100, 200, 400, 600\}$.

## 6. RESULTS

The MAP obtained by the baseline LM was .2217 for AP, .2014 for SJMN and .2842 for WSJ. The obtained results for LBDM and QLBDM are reported in Table 1. As found in previous work [10, 9], our results show that LDA is indeed beneficial for IR and its performance increases with the number of topics used. Moreover, QLBDM statistically outperforms LBDM across different collections and number of topics. Interestingly, QLBDM reaches the best performances obtained by LBDM at a fraction of the number of topics $K$. For example, on SJMN, QLBDM at $K = 100$ reaches comparable performance of LBDM at $K = 600$. A similar behavior can be observed on WSJ and AP. It seems that interference is especially effective for small $K$. In this case, the topic distributions are likely to be coarse-grained thus showing a large overlap. The interference term can be helpful in highlighting and boosting shared patterns in the topic distributions thus producing a more precise document model. In addition, the diminishing returns can be explained by analyzing Fig. 2. We plot the empirical cumulative distribution of the cosines $F_{\Delta}(x) = \frac{2}{K(K-1)} \sum_{i<j} I(\Delta(\phi_i, \phi_j) \leq x)$, where $I$ is the indicator function. By increasing $K$, $\Delta(\phi_i, \phi_j)$ becomes smaller in average, i.e. topics share less common terms. Therefore, the interference factor will play a smaller role.

In Fig. 3, we analyze how LBDM and QLBDM behave with respect to the smoothing parameter $\lambda$ for different values of $K$. Even if not reported explicitly, the pattern on AP was found to be similar. Generally, for any $\lambda > 0.1$, QLBDM stays significantly above the baseline LM. This behavior seems to be more pronounced for high values of $K$. As a result, QLBDM is less sensitive to the choice of smoothing parameter $\lambda$ than LBDM. For LBDM, the optimal $\lambda \in [0.6, 0.8]$, while for QLBDM $\lambda \in [0.3, 0.5]$. The smaller values for QLBDM suggest that interference generates more smoothed document models by

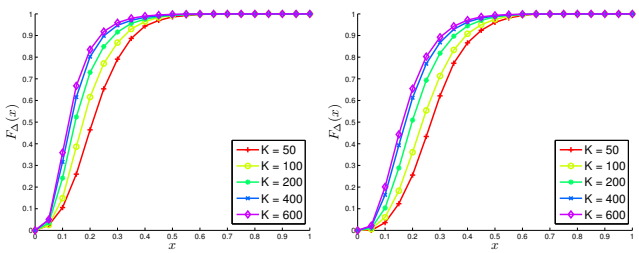



**Figure 2: The cumulative distribution $F_{\Delta}(x)$ for SJMN (left) and WSJ (right) for different values of $K$.**

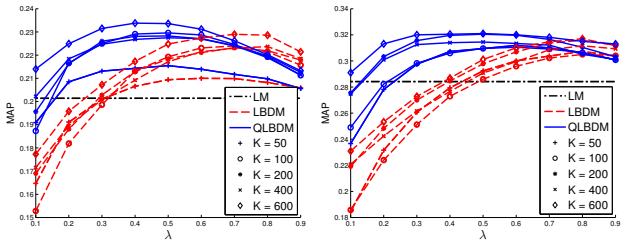



**Figure 3: $\lambda$ variation on SJMN (left) and WSJ (right).**

reducing the relative differences between terms probabilities. In order to highlight such differences, the smoothing parameter $\lambda$ should be set lower.

## 7. CONCLUSION

We presented an application of QI for modeling topic interactions in LDA. Our main focus was to take into account the inherent interactions between topics in generating document terms. The interference term is related to a measure of similarity between pair of topics. The model penalizes topic-specific terms favoring terms that are less probable but shared by similar important topics. Experimental evaluation showed the usefulness of our approach in that significant improvements are obtained over two strong baselines across different number of topics and collections. However, much work remains. Solutions involving more reasonable phase factors have the potential to make yet more significant improvements over LBDM. This will be part of our future investigation.

## 8. REFERENCES


[1] M. Melucci. An investigation of quantum interference in information retrieval. In *Proc of IRFC*, pages 136–151, 2010.

[2] F. A. González and J. C. Caceido. Quantum Latent Semantic Analysis. In *Proc. of ICTIR*, pages 52–63, 2011.

[3] D. M. Blei and J. D. Lafferty. Correlated topic models. In *Proc. of NIPS*, pages 147–155, 2006.

[4] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022, 2003.

[5] R. Feynman. *The Feynman Lectures on Physics*, Vol. 3, 1963.

[6] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman. Indexing by latent semantic analysis. *JASIST*, 41:391–407, 1990.

[7] T. L. Griffiths. Finding scientific topics. *Proc. Nat. Acad. of Sci.*, 101:5228–5235, 2004.

[8] W. Li and A. McCallum. Pachinko allocation: DAG-structured mixture models of topic correlations. In *Proc. of ICML*, pages 577–584, 2006.

[9] X. Wei and W. B. Croft. LDA-based document models for ad-hoc retrieval. In *Proc. of SIGIR*, pages 178–185, 2006.

[10] X. Yi and J. Allan. A comparative study of utilizing topic models for information retrieval. In *Proc. of ECIR*, pages 29–41, 2009.

[11] G. Zuccon and L. Azzopardi. Using the quantum probability ranking principle to rank interdependent documents. In *Proc. of ECIR*, pages 357–369, 2010.