

# Comparative religion, topic models, and conceptualization: Towards the characterization of structural relationships between online religious discourses

Zachary K. Stine<sup>a</sup>, James E. Deitrick<sup>b</sup> and Nitin Agarwal<sup>a</sup>

<sup>a</sup>University of Arkansas at Little Rock, 2801 S. University Ave., Little Rock, AR 72204, United States

<sup>b</sup>University of Central Arkansas, 201 Donaghey Ave., Conway, AR 72035, United States

## Abstract

The similarity between the lexicons of different religious discourses does not necessarily reflect the similarity between the ways of understanding the world inherent in their discourses. Drawing on scholarship from comparative religion that distinguishes between surface-level, lexical distinctions and deeper grammatical and structural distinctions between two religious traditions, we present a computational approach to assessing the structural similarity between religious discourses irrespective of their lexical differences. We argue that unsupervised machine learning models trained on different discourses can be indirectly compared by how consistently they organize information as an operationalization of structural similarity. This consistency can be quantified as the mutual information between the models' clusterings of a designated set of comparison data. We present our approach through a case study comparing discussions from Reddit concerning Buddhism and Christianity.

## Keywords

comparative religion, topic modeling, information theory, digital religion

## 1. Introduction

Comparative analyses of culturally specific discourses are complicated by the possibility for the discourses being compared to reflect ways of understanding the world which are fundamentally similar yet expressed through distinct, culturally specific terms. This is specifically problematic for comparative religion in cases where it is possible for one to adopt the forms of a religious tradition without necessarily adopting the deeper structures beneath those forms (i.e., something like a worldview). For example, it has been argued that the religious life of Henry Steel Olcott—a notable convert to Buddhism—can be understood as comprising an American Protestant structure that informs Olcott's identity despite his adoption of a Buddhist and South Asian cultural lexicon [1]. A distinction in how religious identities are expressed is made here between the consciously-chosen forms that signal an identity—or cultural lexicon—and the deeper cultural structure—or cultural grammar—underlying those forms.

In this paper, we put forward an operationalization of how religious discourses might be empirically compared in such a way that reflects their similarity at the level of cultural grammar rather than cultural lexicon in order to measure what we are calling their structural similarity.

---

*CHR 2020: Workshop on Computational Humanities Research, November 18–20, 2020, Amsterdam, The Netherlands*

EMAIL: zkstine@ualr.edu (Z.K. Stine); deitrick@uca.edu (J.E. Deitrick); nxagarwal@ualr.edu (N. Agarwal)  
ORCID: 0000-0001-5211-0111 (Z.K. Stine); 0000-0002-5612-4753 (N. Agarwal)



© 2020 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



CEUR Workshop Proceedings (CEUR-WS.org)

We assume that a particular discourse reflects a way of understanding the world or some aspect of it [2]. In other words, a discourse reflects a cultural grammar or structure. However, as in the example of Olcott, a discourse may be expressed within a cultural lexicon that is incongruous with the underlying cultural grammar (see section 2.1 for a more detailed discussion of this phenomenon). Importantly, our use of the term “lexical” should be understood to refer to how culturally specific a particular term is, reflecting this notion of cultural lexicon.

Our approach is based on the assumption that a discourse divides the world, or some aspect of it, in a particular way. In other words, a categorization scheme is implicit within a discourse. Given this assumption, we argue that if discourses are structurally similar, they can be expected to produce categorization schemes that carve up information in a mutually consistent manner, despite differences in culturally specific lexicons used in each discourse. We operationalize this notion using unsupervised machine learning models that are trained on each discourse being compared. Each model learns a clustering scheme that is specific to the discourse used to train it, thereby acting as a plausible representation of that discourse’s categorization scheme. We then interrogate the relationship between categorization schemes (represented by the learned models) by forcing each model to apply its discourse-specific scheme to the unseen discourse with which it is being compared. We then measure the mutual consistency with which each discourse-specific model classifies both its own discourse and the comparison discourse using the mutual information between the resulting clusterings.

In order to better clarify what we are attempting to do in this approach, we draw on and extend a particular usage of the term “conceptualization.” A clustering of a data set can be understood as implying a particular conceptualization of that data, and multiple clusterings may imply various ways that a researcher might conceptualize the data, with potential differences or similarities between them [3]. In this sense, clustering data leads a researcher to interpret the resulting clusters as salient concepts for understanding the data. In that case, a single data set is explored through various clusterings in order to find useful conceptualization schemes. Here, we use “conceptualization” to mean how one discourse—as represented within a model trained on it—organizes another discourse in terms of its own semantic elements. In other words, the representation of a different, unseen discourse by a model trained on a different discourse can be understood as how the training discourse “conceptualizes” this unseen discourse.

This usage of “conceptualization” is especially useful given the type of unsupervised model we use: latent Dirichlet allocation (or LDA). LDA learns two things from a corpus: a set of word-usage patterns (or topics), which can be understood as corpus-specific concepts, and a representation of each document in the corpus as a mixture of these word-usage patterns [4]. Importantly, these word-usage patterns may be characterized by the corpus-specific lexicon alongside less corpus-specific terms. When we force such a model to represent the documents of a different corpus as mixtures of its own word-usage patterns, we get a representation of the different corpus through the lens of the corpus which was used to train the model. In other words, we get a conceptualization of this different corpus in terms of the training corpus. We argue that the mutual consistency with which both corpus-specific models “conceptualize” each other reflects their structural similarity—the degree to which each discourse-as-model categorizes the training corpus of each. In this operationalization, the structural similarity is reflected by the mutual information between how two models organize input, regardless of how different the actual word-usage patterns are between the two models. In this way, we are not comparing the features of each model directly, but instead are comparing only how consistent these corpus-specific features are applied by each model. From this, we get a mapping from the

“true” word-usage patterns (from the model trained on the corpus) to those used in another model’s conceptualization of the corpus. This mapping can be usefully thought of as the interpretation of one model’s topics by another model.

Motivated by prior work in comparative religion concerning encounters between Buddhism and American Protestantism [5, 6], we explore the empirical implications of this operationalization in a narrow case study between two English-language discourses from the popular discussion platform, Reddit: *r/Buddhism* and *r/Christianity*. Importantly, there is no reason to assume that either discourse we examine constitutes a general representation of global Buddhism or Christianity (assuming such general forms, untethered from particular social systems, are even valid to begin with). Instead, these discourses should be understood to reflect only the particular versions of Buddhism and Christianity which emerge from these online communities. In other words, rather than focus our comparisons on abstract representations of Buddhism and Christianity, we focus our comparisons on specific communities engaged in discussing Buddhism and Christianity. Therefore, our findings should not be construed as reflections of Buddhism and Christianity as transcendent forms, but as contingent upon these online communities. We include two additional communities to help contextualize our results.

Far from being trivial or unserious objects of scholarly inquiry, such online discourses offer valuable insights into how religious traditions are understood and engaged with in popular culture. In recent years, a body of literature has emerged specifically around the study of religion in digital contexts under the name of “digital religion” [7]. Given the popularity of Reddit, it is reasonable to think that an understanding of its religious communities does have salience for understanding popular conceptions of religious traditions in the English-speaking world. Additionally, the quantity of data that is available from these communities is sufficiently large to be an obstacle to researchers analyzing these data without the aid of computational tools. Quantitative methods are underused within the study of digital religion [8, 9], and so another goal of this work is to demonstrate how such methods may be imported and customized as useful complements to qualitative methods.

We find evidence that our proposed operationalization of structural similarity accords with our expectations about the relationship between the two subreddits’ discourses and with the discourses of two secondary subreddits. Additionally, we investigate which features from models of *r/Buddhism* and *r/Christianity* are most responsible for their structural similarities by calculating the pointwise mutual information between each possible pair of features. We find the context in which the two corpora are compared is highly influential on which feature pairs emerge as most strongly related between models. We also find that, while these feature pairs may have stark differences between the lexical items that characterize them, their mappings between models often appear surprisingly reasonable as if analogies for each other within their different lexical contexts.

In the following sections of this article, we provide background for understanding the theoretical framework we present, describe the data and methods used to illustrate this framework within the case study of the *r/Buddhism* and *r/Christianity* discourses from Reddit, present our findings from the case study, and briefly discuss what these findings suggest about our operationalization and directions for further investigation.

## 2. Background

In the following subsections, we provide background information necessary for constructing our argument that the mutual information between topic models trained on lexically distinct religious discourses can be understood as a reflection of their structural relationship.

### 2.1. Comparative religion and religious creolization

While this comparative problem may be faced in a variety of cultural contexts, we explore it from within the context of comparative religion, and so a brief consideration of the problems faced in comparative religion will provide important context for understanding the challenges faced by this work. Paden identifies three primary criticisms that have been levied against traditional comparative approaches [10]. First, comparativism may mislead by suppressing differences between cultures, engaging in colonialist reductiveness. Second, comparativists have sometimes been guilty of introducing theological or ontological assumptions into their work in an unscientific manner. Finally, charges have been made that comparativism is untheoretical in that it lacks the ability to explain religious differences and similarities.

The use of empirical methods in comparative religion has been suggested as a possible antidote to this last criticism [11], and while computational methods are certainly not objective, they at least reduce the ways in which researchers may introduce their own faulty assumptions into analysis or make those assumptions explicit. However, the potential for reductionism in computational approaches is worth consideration. Computational methods, specifically those from machine learning, are effective in identifying large-scale patterns within data too numerous for individuals to comb through. Such large-scale analyses require a trade-off between the particular and the general. In other words, machine learning methods excel at illuminating trends and generalities, but potentially at the expense of finer-grained variation. While reductionism is certainly a concern, it has been argued by some that a preoccupation with reductionism has substantially hindered comparative religion [12, 13].

With these challenges in mind, we now turn to the comparative work undertaken by Deitrick concerning the relationship between the social ethics of engaged Buddhism and mainstream American religion, which serves as the inspiration for the present study. In [5], Deitrick invokes a theory of religious creolization put forward to describe the religious life of Henry Steel Olcott [1]. This theory posits a distinction between a religion’s grammatical structures and the particular lexical forms through which these structures are expressed. In the case of American engaged Buddhism, Deitrick argues that, in terms of its social ethics, it can be understood as the adoption of a Buddhist lexicon to describe cultural structures that ultimately reflect mainstream American religion. Deitrick refers to this as an “inverse creole faith” in that it reverses the power dynamics of what is typically referred to as “creole”—a dominant group adopts the lexicon of a minority group [5].

In the present study, we are interested in whether the Buddhist discourse from Reddit is only lexically distinct from the Christian discourse, or if it is both lexically and structurally distinct.

### 2.2. Religion on Reddit

Reddit consists of a large number of communities, called subreddits, which facilitate discussions around a defined theme or topic. Users can author submissions to a subreddit and author

comments within discussion threads that accompany each submission. Data from Reddit have been usefully analyzed in work ranging from the effectiveness of hate speech bans [14], violations of community norms [15], persuasion [16], birth narratives [17], and discourses around China [18].

Reddit is a useful source of popular discourses for several reasons. Most importantly, each community constitutes a discourse that is endogenously defined. Constructing a corpus that represents a particular religious tradition is complicated by the decisions that must be made about which documents to include and exclude from the corpus. In the case of Reddit, such consequential decisions are avoided: The community of users and their discussions presents an unambiguously delineated discourse. Additionally, comparative analyses of subreddits have the benefit that all subreddits being analyzed are subject to the same effects that stem from simply being on Reddit, whether in the form of demographic trends of its users or the affordances of the platform. Each subreddit we analyze is predominantly English-language.

While a number of subreddits exist which focus on Buddhist and Christian traditions, we limit ourselves to *r/Buddhism* and *r/Christianity* for two reasons. First, our primary goal in this paper is to explain our proposed approach for making structural comparisons between religious discourses; therefore, we analyze these two subreddits to serve as a focused case study. Second, *r/Buddhism* and *r/Christianity* appear to be the most general subreddits dedicated to their respective religious traditions as well as having the largest discussion histories. We are more interested in popular conceptions of Buddhism generally rather than engagements with more specific traditions within, for example, Theravada, Mahayana, or Vajrayana Buddhism. Similarly, we are interested in general conceptions of Christianity rather than in specific denominations. This is not to suggest that communities with a narrower focus on more specific traditions and denominations are irrelevant to our questions, but simply that, within the current study, we are interested in the two most popular subreddits that involve discussions of Buddhism or Christianity.

Various sects and denominations are surely represented to some extent in these communities, but there is no reason to think they are represented in a balanced way—certain perspectives may loom larger than others. However, to reiterate a previous point, we are not studying *r/Buddhism* because we mistakenly believe it to be an accurate representation of global Buddhist perspectives. Instead, we study it because it is a wildly popular Buddhist discussion community on a wildly popular social media platform and its discourse is therefore salient for understanding Buddhism within popular English-language online culture. The same applies to *r/Christianity*. In future work, we intend to extend our approach to other communities including several smaller sect-specific subreddits alongside those analyzed here. However, *r/Buddhism* and *r/Christianity* remain reasonable and interesting starting places for our case study for the reasons just given.

To provide context for our results comparing *r/Buddhism* and *r/Christianity*, we also report results comparing them with two other subreddits: *r/religion* and *r/math*. Our rationale for including *r/religion* is that we expect it to reflect a tendency in Western culture to associate the notion of religion with Abrahamic traditions and especially with Christianity. We include *r/math* because we expect that, while *r/Buddhism* and *r/Christianity* may present two distinct discourses, they are more likely to reflect similar conceptualization schemes with each other than with discussions about mathematics. Additionally, the inclusion of *r/math* serves as a check to make sure that our approach is still capable of showing dissimilarity and not simply forcing all corpora being compared to appear mostly similar.

### 2.3. Latent Dirichlet allocation

We use latent Dirichlet allocation (LDA) to represent each discourse as a topic model. LDA views a corpus as the result of a generative statistical process in which each document in the corpus is generated by drawing a probability distribution over a set of “topics”—probability distributions over the vocabulary of the corpus—from which each word in the document is then drawn [4]. In training, LDA attempts to infer the distribution over topics for each document in the corpus as well as the distributions over the vocabulary (or “topics”). The learned topics correspond to latent features underlying the corpus. While these features may sometimes correspond to colloquial usages of “topic,” they are better understood as patterns of word-usage, or as [19] suggests, contexts. The topics of LDA can also be understood to reflect several concepts from the sociology of culture [20]. LDA not only provides a representation of each document in the corpus as a mixture of these features but can also provide representations of unseen documents not included in the training corpus as mixtures of these features.

An unsupervised algorithm, LDA learns the topics and document-topic distributions without any specifications of what the content of its features ought to look like. However, LDA does require the selection of the number of topics,  $k$ . Different choices of  $k$  may influence the specificity of the learned features, with smaller values of  $k$  yielding more general topics and larger values yielding more specific topics [21, 19]. Quantitative evaluation of LDA models is a complex problem, and qualitative evaluation is typically necessary to ensure that a model is understandable and therefore helpful to a researcher [22]. Ultimately, it may not make sense to think of one model as more correct than another, even if one appears optimal according to one or more evaluation metrics, but to simply see each as plausible representations of the training corpus.

LDA has been previously used within the context of religious studies including a comparative analysis of three Confucian texts [23] and an investigation into mind-body holism in medieval Chinese thought [24]. LDA has also been used in comparative contexts outside of religious studies to compare the proceedings of natural language processing conferences over time [25] and to compare two discourses about China from Reddit [18]. In each of these cases, LDA is used to train a common topic model that is shared by each of the collections being compared. The relevant documents, terms, or collections of documents are then compared within this shared topic space. This approach makes sense when the objects being compared are not characterized by distinct lexicons or if such lexical distinctions are of interest. What differentiates the approach we describe here is that we are not comparing objects within a shared topic space but are instead comparing how topic models try to fit unseen, lexically distinct discourses into their own topic spaces that are specific to their training discourses. We are not looking at which topics are associated with which discourse but are instead comparing how much consistency exists between how models place documents from different discourses within their own discourse-specific features. In other words, we are looking at how different models “conceptualize” other discourses and measuring the consistency between those conceptualizations rather than measuring the similarity between the concepts themselves.

The LDA models trained on the discussions of *r/Buddhism* and *r/Christianity* can be thought of as representations of their corresponding discourses, where we understand a discourse as a way of understanding the world or some aspects of it [2]. While useful, these representations are not perfect, functioning more like metonyms of the corresponding discourses [26]. We propose thinking of LDA models as not only representations of a discourse, but also as operationalizations of a discourse in that we can deploy the organizational scheme of the



model in novel contexts to see how the model organizes new information, i.e., how it conceptualizes. In addition to learning features and a representation of the training corpus as mixtures of those features, a trained LDA model also has the ability to infer the topic mixtures of new documents using the posterior parameter for document-topic distributions (typically notated as  $\alpha$ ) that becomes the prior in the inference process for new documents' topic distributions. When inferring the topic distributions of unseen documents, this prior acts as the conceptual disposition of the model, which is taken in along with the observed text of the new document to determine its topic distribution. If we were to ask the model to infer the topic mixture of a blank document, it would simply assign this prior topic mixture.

Contrary to the usual goals of machine learning, we do not want these discourse models to generalize beyond their training data. Instead, we want them to reflect only the conceptual schemes latent in their training corpus. Rather than examine the similarity between the features of the models, which reflect differences in lexical content, we are interested in the mutual consistency between how the models conceptualize—do certain features tend to be co-applied to documents regardless of the lexical differences that constitute those features?

## 2.4. Information theory

Information theory provides a useful means for quantifying the kinds of relationships we are trying to uncover between discourses. To quantify the consistency with which two LDA models conceptualize a discourse, we use the mutual information between each model's topic assignments. Introduced in the context of communication channels by [27], the mutual information of two random variables,  $I(X; Y)$ , quantifies the reduction in uncertainty about  $X$  (or  $Y$ ) that is provided by knowing  $Y$  (or  $X$ ) given in bits [28]. A common usage of mutual information is to measure how similarly two clustering schemes partition a set of observations (e.g., [29]). Typically, this is done for hard clusterings in which each observation is assigned to a single class, as distinct from LDA, which assigns observations (documents) to a mixture of multiple classes (topics). A method for “hardening” topic mixtures from LDA is proposed by [18]. However, we calculate the mutual information between the probabilistic clusters of documents, following [30], which does have some complications. Other information theoretic quantities exist for comparing two clusterings, including variations based on mutual information (e.g., [31]) and the metric, variation of information [32], which we plan to compare with the standard mutual information in further work.

Additionally, measures of information divergence provide a useful means for quantifying how lexically distinct two discourses are and how distinguishing each term is individually. One such quantity, the Kullback-Leibler divergence, provides an asymmetric measure of how much one probability distribution differs from an expectation based on another distribution [33]. The Kullback-Leibler divergence (or KLD) has been previously used alongside LDA to characterize the reading behavior of Charles Darwin [34], innovation within parliamentary speeches [35], and legislative change [36]. The Jensen-Shannon divergence (or JSD) is a symmetrical divergence derived from the KLD [37]. The JSD has been previously used to measure the distinguishability between distributions of features from violent and non-violent court trials [38]. It has also been used to measure the difference between LDA topics (e.g., [39]).

The contribution of each feature to the total JSD between distributions can also be calculated. For example, this is done in [38] to identify which trial features most distinguish violent from non-violent trials and vice versa. We use the JSD between the relative frequencies of each word between subreddits to quantify how lexically distinct the discourses of the subreddits are

from each other. Additionally, we can characterize the extent to which each word functions as part of a discourse’s lexicon by calculating each word’s individual contribution to the total JSD between discourses. In this context, a word’s contribution to the JSD between discourses represents how strongly the word implies one discourse over another.

### 3. Methods and Data

In this section, we describe our data collection, preprocessing steps, and put together our framework built from the topics introduced in the previous section, explaining it in parallel to the methods we use to compare the discourses of *r/Buddhism* and *r/Christianity*.<sup>1</sup>

#### 3.1. Data collection and preprocessing

We collect data from the two subreddits of primary interest, *r/Buddhism* and *r/Christianity*, as well as for *r/religion* and *r/math*. For the subreddits of interest, we first collected all available submission IDs from the creation date of the subreddit through the end of 2019. These submission IDs were collected from the service PushShift.io (using the Python wrapper PSAW), which maintains historical data from Reddit. We then used Reddit’s own Application Programming Interface (API) (using the Python wrapper, PRAW) to collect the submission title, body text, and all comments for each submission ID, which were written to CSV files along with relevant metadata such as user ID and timestamps.

After collecting the submissions from each subreddit, we performed basic preprocessing on the text. Tokens are lowercase strings with a minimum length of three characters. URLs are tokenized so that they are reduced to their hostname with hyphens replacing any punctuation (e.g., “en.wikipedia.org” becomes “en-wikipedia-org”). References to users and subreddits are preceded by “u/” and “r/” respectively. We preserve these indicators when tokenizing so that a distinction is made in cases where a user name or subreddit name overlaps with another word type. For example, if a comment references the subreddit, *r/Buddhism*, that reference will be assigned to the word type, “r-buddhism” in order to distinguish it from the word type, “buddhism.” Tokens other than URLs, user names, and subreddit names do not include punctuation or numeric characters.

We created a custom set of 42 stopwords from the most frequent words in each subreddit which were removed from all documents. Additionally, words which occurred in fewer than five documents within each subreddit were removed from all documents. After word removal, the final vocabulary was limited to words that were within the 30,000 most frequent words of a subreddit. Using this final vocabulary, only documents with 20 or more tokens were included in each subreddit’s corpus. An overview of the data collected can be seen in Table 1.

#### 3.2. Quantifying the lexical distinctness of discourses

While it might be reasonable to take it for granted that the discourses of *r/Buddhism* and *r/Christianity* use cultural lexicons that distinguish each from the other, we use the JSD between the relative word frequencies from each subreddit to quantify the degree to which they are lexically distinct from each other. For each word type in the combined vocabulary of the subreddits, we calculate the probability of each word within a subreddit as the number

---

<sup>1</sup>All code used for this analysis is available at [https://github.com/zacharykstine/chr2020\\_comp\\_relg\\_lda](https://github.com/zacharykstine/chr2020_comp_relg_lda).



**Table 1**

Overview of Collected Data

Subreddit	Date Created	Subscribers as of 2020-06-22	Accessible Submissions	Submissions in Corpus	Raw Vocab Size
<i>r/Buddhism</i>	2008-03-25	254,693	87,792	66,108	223,356
<i>r/Christianity</i>	2008-01-25	241,539	412,930	298,502	618,370
<i>r/math</i>	2008-01-24	1,198,611	155,873	103,471	237,742
<i>r/religion</i>	2008-01-25	53,167	88,390	31,283	160,562

of times that word occurs divided by the total number of tokens present in all documents from that subreddit. We then calculate the JSD between the two distributions for each pair of subreddits under consideration.

Additionally, we calculate the individual contributions of each word to the JSD between *r/Buddhism* and *r/Christianity* to see if the words which contribute the most to the total JSD reasonably correspond to what we would expect to see in the cultural lexicons of the subreddits. The way in which we calculate the JSD contribution of each term differs slightly from the method used by [38]. There, the authors calculate the partial KLD of each feature from one distribution to the mean of the two distributions, which quantifies how much each feature signals one particular distribution over the other. Here, we simply calculate the per-feature JSD contributions by calculating the partial KLD of each feature for both distributions. This results in two partial KLD values for each feature from which we take the mean to get the partial JSD of the feature. Done this way, we can see which terms are most distinguishing between the two subreddits from both directions, rather than which terms distinguish one subreddit over the other.

### 3.3. Structural comparisons between discourses

We now propose and explain our implementation of the structural comparisons between the discourses of *r/Buddhism* and *r/Christianity*. We separately train LDA models with 30 topics on the *r/Buddhism* corpus and the *r/Christianity* corpus using the Gensim package for Python [40]. For brevity, we will refer to the model trained on *r/Buddhism* as model B, and the 30-topic model trained on *r/Christianity* as model C, and refer to the  $i^{\text{th}}$  topic of a model as B. $i$  or C. $i$ . After training each model, we get three primary results: a set of “topics” (or features) as probability distributions over the vocabulary, a representation of all documents in the training corpus as distributions of topics, and a way to infer topic distributions for unseen documents. We qualitatively choose labels for each topic based on the highest probability words in the topic as well as close readings of exemplar documents of the topic.

A more common way to compare these two models would be to calculate the similarity or distance between the topics from one model and the topics from the other model (e.g., as in [39]). However, we are less interested in how similar the models’ topics are, and more interested in how similarly the models apply their topics. This is a substantial distinction. We are acknowledging that the two different models, trained on two different corpora, may have completely different topics. However, as long as the models apply those topics to documents in a mutually consistent fashion, the models functionally conceptualize the documents similarly. Our assumption is that, if two models that organize input in a mutually consistent fashion, then they are similar at a structural level regardless of how different their particular features

are from each other.

It is common to think of LDA models as primarily being the set of inferred topics, but this is only half of the full picture. In addition to their topics, LDA models instantiate a particular organizational scheme that takes input text and categorizes it as a mixture of those topics, weighing the observed text being input with a model’s learned disposition for applying its topics. In other words, LDA models can be thought of as both a representation and an operationalization of a discourse, with topics being the former and the way in which models apply those topics to particular documents being the latter. Drawing on and extending the use of the term in [41] and [3], we frame this activity of assigning topics to new information things as conceptualization—the activity of the model representing the novel information in terms of its own discursive features (topics) and dispositions (the trained model’s posterior document-topic distribution parameters, which act as a prior when doing inference on new documents). By comparing how two models apply their topics to a set of documents (rather than comparing their topics directly), we are comparing how each model conceptualizes that particular set of documents. If two models conceptualize information in a mutually consistent way, then they share a kind of similarity that is deeper than the particular forms their concepts (or topics) take. This is what we are referring to as structural similarity, distinct from lexical similarity.

To quantify this shared consistency between two ways of conceptualizing input, we calculate the mutual information between their conceptualizations of a set of documents. Given two clusterings of the same set of objects, the mutual information between the two clusterings represents how much information knowing one cluster assignment provides for knowing the assignment made by the other clustering. As previously noted, mutual information is typically used to quantify the similarity between two “hard” clusterings—those in which each object is assigned to a single cluster. However, we calculate the mutual information between document-topic distributions from two models following the proposed method in [30] by multiplying the transposed document-topic probability matrix from one model with the document-topic matrix of the other model to create a kind of contingency table from which the joint and marginal probabilities of the topics from the two models can be calculated.

Calculating the mutual information between two LDA models in this way requires us to choose the set of documents across which the two models will be compared, and there is no reason to suppose that the mutual information between models will be the same when different document sets are used when calculating it. If we assume that the topic assignments made on the same documents which were used to train the model are the “true” topic assignments of that corpus, we can think of the mutual information between the models based on that corpus as representing how well the other model is able to interpret the first.

For example, when comparing models B and C on the *r/Buddhism* corpus used to train model B, we consider the topic assignments made by model B to be the “true” assignments, since B was trained from this corpus. The topic assignments made by model C, on the other hand, represent something very different. Model C, acting as an extension of its training corpus, is forced to apply its own set of topics (or contexts) from *r/Christianity* to *r/Buddhism*. In other words, model C conceptualizes *r/Buddhism* based on the broad discourse underlying *r/Christianity*. So if model B assigns a document from *r/Buddhism* to have high probability of topic B.*i*, the topic assignment made by model C can be understood as model C’s interpretation of B.*i*. If model C is highly certain about how to assign a topic mixture, perhaps assigning the document to have high probability for topic C.*j*, then model C can be understood as interpreting B.*i* as C.*j* within the context of this single document. If, on the other hand, model C is highly

uncertain about how to assign a topic mixture to the document, the resulting distribution of topics may be highly spread out, lacking a clear mapping from model B to C.

As this is repeated over all of the documents from *r/Buddhism*, if *B.i* and *C.j* continue to occur with high probability in the same documents, then the association between them (in the context of *r/Buddhism*) continues to strengthen. If, however, model C applies a variety of topics to documents with topic *B.i*, whether by topic distributions that are continually spread out over the topics or by applying high probability topics which vary from document to document, then the interpretation of *B.i* by model C becomes less clear. This relationship between the topics is quantified by the mutual information between the models.

Specific relationships between a single topic from one model with a single topic from the other model are quantified by the pointwise mutual information between them. The mutual information is simply the expected pointwise mutual information between all topic pairs across models. Importantly, the mutual information between two models is contingent on the set of documents over which they are compared. As we will show, the mutual information between B and C will depend on the comparison corpus, and more notably, the strongest mappings between topic pairs will also depend on the comparison corpus.

The argument we are exploring here is that if two models representing two lexically distinct discourses are functionally similar (in that they organize information similarly), then the discourses represented by the two models are structurally similar—the two discourses divide aspects of the world up into similar categories, despite using different lexical items to describe the categories. The degree of structural similarity between models is reflected in the mutual information between them on a particular discourse.

In the present article, we empirically explore this argument by comparing how models trained on the discourses of *r/Buddhism* and *r/Christianity* interpret each other by calculating the mutual information between their topic assignments twice: once for each corpus to act as the comparison corpus. To contextualize these results, we compare them to the self-mutual information of each corpus and corresponding model. We also compare how each model interprets models trained on the discussions of *r/math* and *r/religion*. For each comparison, we refer to the model trained on the comparison corpus as the source model and refer to the model trained on a corpus other than the comparison corpus as the interpreting model.

To better understand what the mutual information between models represents, we look at which topic pairs between models B and C have the largest pointwise mutual information. To assess how different our proposed method for comparing models is from a direct comparison of topics between models, we also calculate the distance between all topic pairs using the Jensen-Shannon divergence.

To get a sense of how dependent these results are on using models with 30 topics, we also train models on each subreddit with 60 topics and calculate the mutual information between them. To differentiate models with different numbers of topics, we subscript *k* with the model name (e.g.,  $B_{30}$  or  $C_{60}$ ).

## 4. Results

In this section, we report the results of our methodology within the narrow case study of the subreddits previously described. Our goal in reporting these results is to illustrate the empirical implications of the way we have defined and operationalized the notion of structural similarity.

**Table 2**  
Jensen-Shannon Divergence Between Vocabulary Distributions

Subreddit	Subreddit	JSD (bits)
<i>r/Christianity</i>	<i>r/religion</i>	0.092
<i>r/Buddhism</i>	<i>r/religion</i>	0.116
<i>r/Buddhism</i>	<i>r/Christianity</i>	0.131
<i>r/Buddhism</i>	<i>r/math</i>	0.224
<i>r/Christianity</i>	<i>r/math</i>	0.238

#### 4.1. Lexical comparisons

When we calculate the JSD between the vocabulary distributions of each subreddit, we find that *r/Christianity* and *r/religion* have the least divergence between them. In other words, they are the least lexically distinct pair. We also find that *r/Buddhism* is slightly less lexically distinct from *r/religion* than from *r/Christianity*. The JSD values between the vocabulary distributions of the subreddits are given in Table 2. The relationships between subreddits that emerge from their lexical distinctness provide a good baseline against which we can compare their structural similarity. As we will show in the sections below, there is some disagreement between the ordering of lexical similarity between subreddits in Table 2 with the orderings we obtain from their structural similarity reported in the subsections that follow. This disagreement, though slight, is an encouraging sign that our approach to calculating structural similarity is not simply a more complicated, but functionally equivalent, calculation of the lexical similarity—it measures something different.

A sample of the twenty-two words with the largest contributions to the JSD between the vocabulary distributions of *r/Buddhism* and *r/Christianity* is provided in Table 3. Most of these highly distinguishing terms are reasonable candidates for the cultural lexicons of either subreddit. Some terms such as “practice” or “suffering” may not be unique to a single religious lexicon. However, their relatively large JSD contributions indicate that they are highly distinguishing terms between the subreddits—they are strong signals of one discourse over the other. Importantly, this way of quantifying the extent to which a word functions as part of a discourse’s lexicon is dependent on the discourse it is being compared with.

#### 4.2. Mutual information between models

Given that we are calculating mutual information between probabilistic clusterings of documents, we first calculate the mutual information between each model and itself. This self-mutual information for each model gives us a rough sense of the maximum mutual information possible for the corpus on which the model was trained. For that reason, when we report mutual information between models trained on different corpora, we also report what percentage of the self-mutual information that value is, according to the self-mutual information of the source corpus. The self-mutual information values for each subreddit can be found in Table 4.

When we calculate the mutual information between models  $B_{30}$  and  $C_{30}$  with *r/Buddhism* as the comparison corpus, we get 0.182 bits or 59% of the mutual information model  $B_{30}$  has with itself. Similarly, we find that  $B_{60}$  and  $C_{60}$  have mutual information of 0.133 bits (57% of the self-mutual information of  $B_{60}$ ). When we calculate the mutual information between  $B_{30}$  and  $C_{30}$  within the context of the *r/Christianity* corpus, we get 0.168 bits or 42% of the self-mutual

**Table 3**Word Types with the Largest JSD Contributions Between  $r/Buddhism$  and  $r/Christianity$ 

Word Type	Contribution to JSD (bits)	Word Type	Contribution to JSD (bits)
god	$5.66 \times 10^{-3}$	sin	$9.67 \times 10^{-4}$
buddhism	$2.61 \times 10^{-3}$	christians	$9.34 \times 10^{-4}$
buddha	$2.42 \times 10^{-3}$	mind	$8.10 \times 10^{-4}$
jesus	$2.02 \times 10^{-3}$	self	$7.04 \times 10^{-4}$
church	$1.79 \times 10^{-3}$	suffering	$6.78 \times 10^{-4}$
buddhist	$1.78 \times 10^{-3}$	dharma	$6.66 \times 10^{-4}$
bible	$1.53 \times 10^{-3}$	path	$6.30 \times 10^{-4}$
christ	$1.26 \times 10^{-3}$	zen	$5.55 \times 10^{-4}$
meditation	$1.22 \times 10^{-3}$	karma	$5.38 \times 10^{-4}$
practice	$1.13 \times 10^{-3}$	faith	$5.20 \times 10^{-4}$
christian	$1.04 \times 10^{-3}$	enlightenment	$5.16 \times 10^{-4}$

**Table 4**

Self-Mutual Information of Models

Training Corpus	Self-MI (bits) $k = 30$	Self-MI (bits) $k = 60$
$r/math$	0.629	0.547
$r/Christianity$	0.401	0.309
$r/Buddhism$	0.306	0.236
$r/religion$	0.250	0.265

information of model  $C_{30}$ . We likewise find 0.125 bits of mutual information between  $B_{60}$  and  $C_{60}$  (40% of the self-mutual information of  $C_{60}$ ).

These results can be understood to reflect how well the two models—and as imperfect representations, the two discourses—interpret each other. In the form of their corresponding models, the discourse of  $r/Christianity$  is capable of interpreting the discourse of  $r/Buddhism$  better than  $r/Buddhism$  can interpret  $r/Christianity$ . This is true in the case of the models with 30 topics as well as the 60-topic models (see Tables 5 and 6).

Simply knowing the mutual information values between does not provide strong intuitions about their structural similarity, so we contextualize these values with comparisons to  $r/religion$  and  $r/math$ . Given the number of topics in the model trained on  $r/religion$  that reflect generally Abrahamic and monotheistic religious concerns, we expect  $r/Christianity$  and  $r/religion$  to have higher structural similarity with each other than any other subreddit pairing. We find that the largest mutual information between any two subreddits occurs between  $r/Christianity$  and  $r/religion$  when the comparison corpus is  $r/Christianity$ . This is the case for both the 30-topic and 60-topic models. In the case of the 30-topic models,  $r/religion$  interprets  $r/Christianity$  with 54% of the self-mutual information of  $r/Christianity$ , the third-highest. In the 60-topic models,  $r/religion$  interprets  $r/Christianity$  with 63% of the self-mutual information of  $r/Christianity$ , rising to the second-highest.

In the case of  $r/math$ , we expect both  $r/Buddhism$  and  $r/Christianity$  to be highly distinct, both lexically and structurally. Accordingly, the four comparisons done between the subreddits of interest and  $r/math$  generate the lowest four mutual information values (as percentages of

**Table 5**

Mutual Information Between Models with 30 Topics

Interpreting Model Corpus	Source Model Corpus	MI (bits)	Percent of Source Self-MI
<i>r/Christianity</i>	<i>r/religion</i>	0.198	79%
<i>r/Christianity</i>	<i>r/Buddhism</i>	0.182	59%
<i>r/religion</i>	<i>r/Christianity</i>	0.218	54%
<i>r/religion</i>	<i>r/Buddhism</i>	0.137	45%
<i>r/Buddhism</i>	<i>r/religion</i>	0.108	43%
<i>r/Buddhism</i>	<i>r/Christianity</i>	0.168	42%
<i>r/math</i>	<i>r/Buddhism</i>	0.095	31%
<i>r/Christianity</i>	<i>r/math</i>	0.167	27%
<i>r/Buddhism</i>	<i>r/math</i>	0.139	22%
<i>r/math</i>	<i>r/Christianity</i>	0.078	20%

**Table 6**

Mutual Information Between Models with 60 Topics

Interpreting Model Corpus	Source Model Corpus	MI (bits)	Percent of Source Self-MI
<i>r/Christianity</i>	<i>r/religion</i>	0.183	69%
<i>r/religion</i>	<i>r/Christianity</i>	0.196	63%
<i>r/Christianity</i>	<i>r/Buddhism</i>	0.133	57%
<i>r/religion</i>	<i>r/Buddhism</i>	0.118	50%
<i>r/Buddhism</i>	<i>r/Christianity</i>	0.125	40%
<i>r/Buddhism</i>	<i>r/religion</i>	0.098	37%
<i>r/math</i>	<i>r/Buddhism</i>	0.081	34%
<i>r/Christianity</i>	<i>r/math</i>	0.143	26%
<i>r/math</i>	<i>r/Christianity</i>	0.075	24%
<i>r/Buddhism</i>	<i>r/math</i>	0.117	21%

the appropriate self-mutual information). This is true for the 30-topic models and for the 60-topic models. Mutual information values for models with 30 topics can be seen in Table 5, and values for models with 60 topics can be seen in Table 6.

### 4.3. Pointwise mutual information between topics

While the mutual information between models on a comparison corpus provides a high-level picture of the relationship between models, it is also possible to dig into which features of the discourses are mapped together by looking at which topic pairs between models have the highest pointwise mutual information. For brevity, we only focus on the 30-topic models, B<sub>30</sub> and C<sub>30</sub>, as a case study for which we obtain all 900 pointwise mutual information values for each combination of topics for both *r/Buddhism* and *r/Christianity* with each as the source corpus. As examples, we report the ten topic pairs with the highest pointwise mutual information in Tables 7 and 8, annotated with our qualitative topic labels based on high-probability topic words and close readings of exemplar documents for each topic. Notably, these examples reveal that, despite their lexical differences, these mappings appear surprisingly reasonable in many cases.



**Table 7**

Ten Topic Pairs with Highest PMI Between 30-topic Models Trained on *r/Buddhism* and *r/Christianity* and Compared on the Documents of *r/Buddhism*

<i>r/Buddhism</i> Source Topics	<i>r/Christianity</i> Interpreted Topics	Pointwise Mutual Information
B.16 Relationships	C.15 Relationships	3.095
B.24 Dietary Ethics & Meat	C.18 Abortion	2.797
B.05 Repeated Text	C.27 Repeated Text: Moderators	2.761
B.05 Repeated Text	C.10 Repeated Text: Verse Bot	2.743
B.21 Intl. Politics & Conflict	C.08 American Politics & Race	2.670
B.12 Text Quotations	C.23 Bible Verses	2.665
B.25 Precepts	C.25 Sex & Morality	2.617
B.03 Monastic Practice & Monks	C.03 Churches & Fellowship	2.583
B.25 Precepts	C.29 Sexual Preferences	2.295
B.26 Source Text Discussion	C.22 Historical Jesus & Accuracy	2.156

The topic pairs with high pointwise mutual information suggest interesting analogies. For example, the association that emerges between topics B.24 and C.18 suggests that discussions about dietary ethics are to *r/Buddhism* what discussions about abortion are to *r/Christianity*. The content of these discussions is considerably distinct lexically. Yet, these divisive ethical and moral debates occur in both subreddits with the particular focus of the debates marking the discourse as that of *r/Christianity* (in the case of abortion) or of *r/Buddhism* (in the case of eating meat).

This example provides important clues as to how this method of comparison works. When model B<sub>30</sub> encounters discussions about abortion in *r/Christianity*, it is confronted with terms that are not prominent in its training corpus from *r/Buddhism*. None of the topics in model B<sub>30</sub> include the term “abortion” as a high-probability term and so the term does not play much of a role in model B<sub>30</sub> choosing an appropriate topic mixture. Instead, model B<sub>30</sub> is forced to ignore lexically distinct terms like “abortion” in favor of terms that are less distinguishing between the two discourses. Thus a common structural property between discourses emerges that we might label as something that is non-discourse specific such as “contentious ethical issues.”

Additionally, we find that the relative strength of the associations between topics is dependent on the comparison corpus used. The interpretation by model C<sub>30</sub> of B.24 as C.18 has the second-highest pointwise mutual information (see Table 7), whereas the interpretation by model B<sub>30</sub> of C.18 as B.24 ranks tenth (see Table 8).

In order to assess how different these topic mappings are from those we might get using a more standard method of comparing topics directly, we calculate the Jensen-Shannon divergence between each pair of topics between model B<sub>30</sub> and model C<sub>30</sub>. The ten most similar topic pairs (i.e., those with the lowest Jensen-Shannon divergence) can be seen in Table 9. We find that, while overlap certainly exists, the ten most similar pairs of topics between models are not necessarily those that appear most salient when making indirect comparisons within the context of a comparison corpus.

Topics B.16 and C.15 appear as the most similar when compared directly in this way. This is also true when compared indirectly through the interpretation of *r/Buddhism* by *r/Christianity* (in the form of model B<sub>30</sub> and C<sub>30</sub>) as shown in Table 7). However, this topic pair is ranked

**Table 8**

Ten Topic Pairs with Highest PMI Between 30-topic Models Trained on *r/Buddhism* and *r/Christianity* and Compared on the Documents of *r/Christianity*

<i>r/Christianity</i> Source Topics	<i>r/Buddhism</i> Interpreted Topics	Pointwise Mutual Information
C.04 Prayer	B.07 Schools & Sects	3.113
C.27 Repeated Text: Moderators	B.05 Repeated Text	3.103
C.16 Health	B.19 Mental Health	2.844
C.14 Science & Evolution	B.17 Mind & Reality	2.821
C.25 Sex & Morality	B.25 Precepts	2.711
C.01 Resources & Bible Versions	B.06 Resources	2.705
C.03 Churches & Fellowship	B.03 Monastic Practice & Monks	2.538
C.01 Resources & Bible Versions	B.26 Source Text Discussion	2.478
C.29 Sexual Preferences	B.25 Precepts	2.382
C.18 Abortion	B.24 Dietary Ethics & Meat	2.274

**Table 9**

Ten Topic Pairs from 30-topic Models with Lowest Jensen-Shannon Divergence

<i>r/Buddhism</i> Topics	<i>r/Christianity</i> Topics	Jensen-Shannon Divergence (bits)
B.16 Relationships	C.15 Relationships	0.153
B.09 Debate, Opinions, Questions	C.28 Debate, Non-Christians, Criticisms	0.159
B.00 Advice	C.17 Advice	0.185
B.09 Debate, Opinions, Questions	C.09 Debate, Theology, Apologetics	0.222
B.18 Dealing with People	C.17 Advice	0.258
B.09 Debate, Opinions, Questions	C.07 Bible & Interpretation	0.270
B.17 Mind & Reality	C.09 Debate, Theology, Apologetics	0.271
B.18 Dealing with People	C.28 Debate, Non-Christians, Criticisms	0.282
B.21 Intl. Politics & Conflict	C.21 Money & Society	0.293
B.00 Advice	C.11 References, Stories, Humor	0.294

twelfth when indirectly compared through the interpretation of *r/Christianity* by *r/Buddhism*. Evidently, the choice of comparison corpus is consequential for how salient the same topic pair is within the comparison. The extent of how consequential the differences are between direct and indirect comparisons can be severe. When *r/Buddhism* interprets *r/Christianity*, the relationship between C.04 and B.07 is strongest. When *r/Christianity* interprets *r/Buddhism*, this pairing is ranked 32nd. When compared directly using the Jensen-Shannon divergence, the pair is ranked 672nd. Clearly, indirect comparisons between topics within the context specified by a comparison corpus are capable of painting substantially different pictures of how the features between two models are mapped.

## 5. Discussion

Our goal in reporting the above results is not to prove the validity of our operationalization of structural similarity but to provide a glimpse of what this operationalization looks like within a narrow case study, and to see how closely the results of this case study conform to our intuitions.

Further work is therefore necessary to continue exploring the method we have proposed here. While we have suggested one possible operationalization of structural similarity, there are likely to be many different possible operationalizations which may overcome limitations present in ours.

An important limitation of the analysis we present here is that we have only considered two sets of LDA models for representing the discourses. LDA models trained on the same corpus and with the same parameters may still exhibit differences due to the randomness in the training process. For this reason, it is possible that particularities within these models may produce mutual information that is highly dependent on those particularities. In future work, we will examine the relationships between corpora where each is represented by a variety of LDA models in order to get a more robust reading of the mutual information that tends to occur between models trained on different corpora.

We believe that an important strength of the approach we outline here is that it does not require any significant modifications to each corpus beyond standard preprocessing. However, our next steps will include an approach in which each corpus is modified in such a way that it is forced to be less lexically distinct from the corpora with which it is compared. Possibilities for reducing the lexical distinctness between two corpora might include the removal of certain terms based on their contribution to the JSD between the vocabulary distributions of the corpora being compared. Additionally, the methods put forward by [42] to reduce the correlation between the topics of an LDA model and metadata may be appropriate for this context as well.

If our attempt at quantifying structural relationships between discourses has some validity, we can begin to explore comparative religion (and perhaps comparative culture more broadly conceived) as a meta-clustering problem in which relationships between various clustering schemes learned from different discourses suggest similarities and differences that go far deeper than lexical distinctions. This is similar to the meta-clustering problem described in [3], except in that case, the different clusterings being compared are all learned from the same set of observations. Our case, wherein each clustering is learned from a different set of observations, brings up additional complications. Most importantly, it is not clear whether or not the structural similarity, as we have defined it here, between two discourses is stable across various contexts in which the discourses are compared (i.e., the comparison corpus). As our results show, the structural similarity is contingent on the context in which the discourses are compared. However, it is possible that, as two discourses are compared within a greater variety of comparison corpora, that their structural similarity becomes stable. Even if a stable trend of structural similarity does not emerge between discourses, then examining the contexts in which their structural similarity differs should still offer useful insights.

## 6. Conclusion

Drawing from the comparative religion research in [5, 6] and the framing of unsupervised machine learning models as conceptualization schemes found in [3] and [41], we have proposed a computational theory of the structural similarity between lexically distinct religious discourses—discourses that are characterized by distinct lexicons. We have argued that, if two unsupervised machine learning models organize information with a high degree of mutual consistency as quantified by the mutual information between them, then they share a high degree of structural similarity, regardless of the lexical distinctions between the models' representa-

tions. Using latent Dirichlet allocation as our model of choice, we developed our theory and explored its empirical implications for a case study comparing the discourses of two discussion communities from Reddit: *r/Buddhism* and *r/Christianity*. The results from this case study suggest that our method for quantifying structural similarity has merit and warrants further exploration.

## Acknowledgments

This research is funded in part by grants from the U.S. National Science Foundation (OIA-1946391, OIA-1920920, IIS-1636933, ACI-1429160, and IIS-1110868), U.S. Office of Naval Research (N00014-10-1-0091, N00014-14-1-0489, N00014-15-P-1187, N00014-16-1-2016, N00014-16-1-2412, N00014-17-1-2675, N00014-17-1-2605, N68335-19-C-0359, N00014-19-1-2336, N68335-20-C-0540), U.S. Air Force Research Lab, U.S. Army Research Office (W911NF-17-S-0002, W911NF-16-1-0189), U.S. Defense Advanced Research Projects Agency (W31P4Q-17-C-0059), Arkansas Research Alliance, the Jerry L. Maulden/Entergy Endowment at the University of Arkansas at Little Rock, and the Australian Department of Defense Strategic Policy Grants Program (SPGP) (award number: 2020-106-094) to the third co-author, Nitin Agarwal. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding organizations. The researcher gratefully acknowledges the support.

## References

- [1] S. Prothero, *The White Buddhist: The Asian Odyssey of Henry Steel Olcott*, 1st. ed., Indiana University Press, Indianapolis, IN, 1996.
- [2] M. Jørgensen, L. J. Phillips, *Discourse Analysis as Theory and Method*, 1st ed., Sage, London, 2002.
- [3] J. Grimmer, G. King, General purpose computer-assisted clustering and conceptualization, *Proceedings of the National Academy of Sciences* 108 (2011) 2643–2650. doi:10.1073/pnas.1018067108.
- [4] D. M. Blei, A. Y. Ng, M. I. Jordan, Latent Dirichlet allocation, *Journal of Machine Learning Research* 3 (2003) 993–1022.
- [5] J. E. Deitrick, *Mistaking the Boat for the Shore?: A Critical Analysis of Socially Engaged Buddhism in the United States*, Ph.D. thesis, University of Southern California, Los Angeles, CA, 2000. UMI Number: 3041445.
- [6] J. E. Deitrick, Engaged Buddhist ethics: Mistaking the boat for the shore, in: C. Queen, C. Prebish, D. Keown (Eds.), *Action Dharma: New Studies in Engaged Buddhism*, RoutledgeCurzon Critical Studies in Buddhism, 1st. ed., RoutledgeCurzon, New York, NY, 2003, pp. 252–269.
- [7] H. Campbell, Making space for religion in internet studies, *The Information Society* 21 (2005) 309–315. doi:10.1080/01972240591007625.
- [8] M. Lövheim, H. A. Campbell, Considering critical methods and theoretical lenses in digital religion studies, *New Media & Society* 19 (2017) 5–14. doi:10.1177/1461444816649911.
- [9] T. Hutchings, Digital humanities and the study of religion, in: P. Svensson, D. T. Goldberg (Eds.), *Between Humanities and the Digital*, 1st. ed., The MIT Press, Cambridge, MA, 2015, pp. 283–294.

- [10] W. E. Paden, Comparative religion, in: J. R. Hinnells (Ed.), *The Routledge Companion to the Study of Religion*, 1st. ed., Routledge, New York, NY, 2005, pp. 208–225.
- [11] L. H. Martin, Comparison, in: W. Braun, R. T. McCutcheon (Eds.), *Guide to the Study of Religion*, 1st. ed., Cassell, London, 2005, pp. 45–56.
- [12] E. Slingerland, Who’s Afraid of Reductionism? The Study of Religion in the Age of Cognitive Science, *Journal of the American Academy of Religion* 76 (2008) 375–411. doi:10.1093/jaarel/1fn004.
- [13] F. Cho, R. K. Squiers, Religion as a Complex and Dynamic System, *Journal of the American Academy of Religion* 81 (2013) 357–398. doi:10.1093/jaarel/1ft016.
- [14] E. Chandrasekharan, U. Pavalanathan, A. Srinivasan, A. Glynn, J. Eisenstein, E. Gilbert, You can’t stay here: The efficacy of reddit’s 2015 ban examined through hate speech, *Proc. ACM Hum.-Comput. Interact.* 1 (2017). doi:10.1145/3134666.
- [15] E. Chandrasekharan, M. Samory, S. Jhaver, H. Charvat, A. Bruckman, C. Lampe, J. Eisenstein, E. Gilbert, The internet’s hidden rules: An empirical study of reddit norm violations at micro, meso, and macro scales, *Proc. ACM Hum.-Comput. Interact.* 2 (2018). doi:10.1145/3274301.
- [16] C. Tan, V. Niculae, C. Danescu-Niculescu-Mizil, L. Lee, Winning arguments: Interaction dynamics and persuasion strategies in good-faith online discussions, in: *Proceedings of the 25th International Conference on World Wide Web, WWW ’16*, International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 2016, p. 613–624. doi:10.1145/2872427.2883081.
- [17] M. Antoniak, D. Mimno, K. Levy, Narrative paths and negotiation of power in birth stories, *Proc. ACM Hum.-Comput. Interact.* 3 (2019). doi:10.1145/3359190.
- [18] Z. K. Stine, N. Agarwal, Comparative discourse analysis using topic models: Contrasting perspectives on china from reddit, in: *International Conference on Social Media and Society, SMSociety’20*, Association for Computing Machinery, New York, NY, USA, 2020, p. 73–84. doi:10.1145/3400806.3400816.
- [19] C. Allen, J. Murdock, Lda topic modeling: Contexts for the history & philosophy of science, 2020. Preprint of a chapter forthcoming in Ramsey, G., De Block, A.(Eds.) *The Dynamics of Science: Computational Frontiers in History and Philosophy of Science*. Pittsburgh University Press; Pittsburgh.
- [20] P. DiMaggio, M. Nag, D. Blei, Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of u.s. government arts funding, *Poetics* 41 (2013) 570–606. URL: <http://www.sciencedirect.com/science/article/pii/S0304422X13000661>. doi:<https://doi.org/10.1016/j.poetic.2013.08.004>, topic Models and the Cultural Sciences.
- [21] D. Nguyen, M. Liakata, S. DeDeo, J. Eisenstein, D. Mimno, R. Tromble, J. Winters, How we do things with words: Analyzing text as social and cultural data, *CoRR abs/1907.01468* (2019). [arXiv:1907.01468](https://arxiv.org/abs/1907.01468).
- [22] M. E. Roberts, B. M. Stewart, D. Tingley, Navigating the local modes of big data: The case of topic models, in: R. M. Alvarez (Ed.), *Computational Social Science: Discovery and Prediction*, 1st. ed., Cambridge University Press, New York, NY, 2016, pp. 51–97.
- [23] R. Nichols, E. Slingerland, K. Nielbo, U. Bergeton, C. Logan, S. Kleinman, Modeling the contested relationship between analects, mencius, and xunzi: Preliminary evidence from a machine-learning approach, *The Journal of Asian Studies* 77 (2018) 19–57. doi:10.1017/S0021911817000973.
- [24] E. Slingerland, R. Nichols, K. Neilbo, C. Logan, *The Distant Reading of Religious Texts*:

- A “Big Data” Approach to Mind-Body Concepts in Early China, *Journal of the American Academy of Religion* 85 (2017) 985–1016. URL: <https://doi.org/10.1093/jaarel/lfw090>. doi:10.1093/jaarel/lfw090.
- [25] D. Hall, D. Jurafsky, C. D. Manning, Studying the history of ideas using topic models, in: *Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP '08*, Association for Computational Linguistics, USA, 2008, p. 363–371.
  - [26] A. Piper, There will be numbers, *Journal of Cultural Analytics* (2016). doi:10.22148/16.006.
  - [27] C. E. Shannon, A mathematical theory of communication, *The Bell system technical journal* 27 (1948) 379–423.
  - [28] T. M. Cover, J. A. Thomas, *Elements of Information Theory*, 2nd. ed., John Wiley Sons, Inc., Hoboken, NJ, 2006.
  - [29] I. S. Dhillon, S. Mallela, D. S. Modha, Information-theoretic co-clustering, in: *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '03*, Association for Computing Machinery, New York, NY, USA, 2003, p. 89–98. doi:10.1145/956750.956764.
  - [30] Y. Lei, J. C. Bezdek, J. Chan, N. Xuan Vinh, S. Romano, J. Bailey, Generalized information theoretic cluster validity indices for soft clusterings, in: *2014 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*, 2014, pp. 24–31.
  - [31] M. E. J. Newman, G. T. Cantwell, J.-G. Young, Improved mutual information measure for clustering, classification, and community detection, *Phys. Rev. E* 101 (2020) 042304. URL: <https://link.aps.org/doi/10.1103/PhysRevE.101.042304>. doi:10.1103/PhysRevE.101.042304.
  - [32] M. Meilä, Comparing clusterings—an information based distance, *Journal of Multivariate Analysis* 98 (2007) 873 – 895. URL: <http://www.sciencedirect.com/science/article/pii/S0047259X06002016>. doi:<https://doi.org/10.1016/j.jmva.2006.11.013>.
  - [33] S. Kullback, R. A. Leibler, On information and sufficiency, *Ann. Math. Statist.* 22 (1951) 79–86. URL: <https://doi.org/10.1214/aoms/1177729694>. doi:10.1214/aoms/1177729694.
  - [34] J. Murdock, C. Allen, S. DeDeo, Exploration and exploitation of victorian science in darwin’s reading notebooks, *Cognition* 159 (2017) 117 – 126. URL: <http://www.sciencedirect.com/science/article/pii/S0010027716302840>. doi:<https://doi.org/10.1016/j.cognition.2016.11.012>.
  - [35] A. T. J. Barron, J. Huang, R. L. Spang, S. DeDeo, Individuals, institutions, and innovation in the debates of the french revolution, *Proceedings of the National Academy of Sciences* 115 (2018) 4607–4612. URL: <https://www.pnas.org/content/115/18/4607>. doi:10.1073/pnas.1717729115. arXiv:<https://www.pnas.org/content/115/18/4607.full.pdf>.
  - [36] Z. K. Stine, N. Agarwal, A quantitative portrait of legislative change in ukraine, in: R. Thomson, H. Bisgin, C. Dancy, A. Hyder (Eds.), *Social, Cultural, and Behavioral Modeling*, Springer International Publishing, Cham, 2019, pp. 50–59.
  - [37] J. Lin, Divergence measures based on the shannon entropy, *IEEE Transactions on Information Theory* 37 (1991) 145–151. doi:10.1109/18.61115.
  - [38] S. Klingenstein, T. Hitchcock, S. DeDeo, The civilizing process in london’s old bailey, *Proceedings of the National Academy of Sciences* 111 (2014) 9419–9424. URL: <https://www.pnas.org/content/111/26/9419>. doi:10.1073/pnas.1405984111. arXiv:<https://www.pnas.org/content/111/26/9419.full.pdf>.
  - [39] F. Morstatter, J. Pfeffer, H. Liu, K. Carley, Is the sample good enough? Comparing data from Twitter’s streaming API with Twitter’s firehose, in: *International AAAI Conference*



on Web and Social Media, 2013.

- [40] R. Řehůřek, P. Sojka, Software Framework for Topic Modelling with Large Corpora, in: Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks, ELRA, Valletta, Malta, 2010, pp. 45–50. <http://is.muni.cz/publication/884893/en>.
- [41] K. D. Bailey, Typologies and Taxonomies: An Introduction to Classification Techniques, Quantitative Applications in the Social Sciences, 1st. ed., Sage, Beverly Hills, CA, 1994.
- [42] L. Thompson, D. Mimno, Authorless topic models: Biasing models away from known structure, in: Proceedings of the 27th International Conference on Computational Linguistics, Association for Computational Linguistics, Santa Fe, New Mexico, USA, 2018, pp. 3903–3914. URL: <https://www.aclweb.org/anthology/C18-1329>.