CPSC 464
Algorithmic Curation of Inclusive Syllabi: Measuring and Increasing Diversity in Course Content

Winiboya Aboyure Mark Ayiah Adham Hussein Aly Moosa Lee Muita

Deborah Olorunisola Bohan Shakes Professor: Nisheeth K. Vishnoi

December 7, 2024

Abstract

Diversity in university syllabi fosters student engagement, enhances critical thinking, and ensures equitable representation of perspectives, preparing students for a globalized world. However, curating diverse syllabi presents significant challenges, including the time-intensive task of identifying inclusive materials and the lack of institutional support or incentives for regular syllabus revision. To address these issues, we introduce a framework that integrates quantitative metrics—such as Rao's Entropy, Jaccard's Distance, and relevance and overlap proportions—to evaluate and enhance syllabus diversity. By leveraging the Open Library API, this framework automates metadata retrieval and subject analysis, streamlining the process and reducing reliance on manual efforts. Empirical results suggest that the framework, when paired with an appropriate metric, effectively measures thematic richness and dissimilarity while providing actionable recommendations for diversifying syllabus content. Beyond higher education, the framework could support applications such as designing inclusive patient scenarios for medical training, curating diverse datasets for algorithmic fairness, and selecting representative case studies for business education. These applications highlight its potential to foster diversity across a variety of fields and contexts.

Contents

1 Introduction

1.1 High level description of the problem.

Curating a diverse university syllabus is difficult. While instructors may be encouraged to consider representation across various dimensions such as race, gender, sexuality, class, religion, and nationality, the process poses significant cognitive and logistical challenges inherent to curricular reform (12). These challenges include the effort required to identify reputable and inclusive materials, the difficulty of aligning these new materials with existing course content, and the need to redesign syllabi in ways that maintain pedagogical coherence. High switching costs and time pressures further hinder the process, often discouraging instructors from updating their syllabi. Changes to syllabi may also conflict with an instructor's professional identity, as adopting new teaching practices or reframing course content can temporarily diminish their sense of expertise or redefine their roles in ways they find uncomfortable. Consequently, syllabi frequently remain static from semester to semester, perpetuating the lack of diversity.

University syllabi often mirror the relative lack of diversity among faculty, perpetuating limited representation in academic content (8). Despite significant institutional reforms to diversify student bodies within the past 50 years, efforts to hire a diverse body of faculty have lagged behind. This discrepancy stems from systemic issues and structural barriers, including the under-representation of low-income students and students of color in graduate programs and reliance on internal referrals (10).

Our project aims to analyze book diversity and, while it does not explicitly incorporate fairness constraints in its algorithmic approach, it allows users to diversify their syllabi by focusing on specific areas of diversity that are often overlooked in traditional book classification systems. Ultimately, the user makes considerations regarding a book's utility, allowing them to choose from suggested books to integrate into their syllabi. By exploring various metrics for measuring diversity, we allow users to define their own concepts of diversity and fairness that align with their goals.

1.2 Motivation to study the problem.

Representation is crucial, particularly for individuals from historically underrepresented groups in academia, and cultivating intellectual diversity through diverse syllabi improves learning outcomes for all students. Studies demonstrate that incorporating teachers and content that reflect gender and racial diversity in classes leads to increased success in traditional classroom contexts. This approach fosters a deeper appreciation across entire classes for the value of diverse perspectives and role models (4). Moreover, when students see their backgrounds and identities represented in classroom materials, they tend to think more about their futures, ambitions, and engage with more achievement-related goals and activities (19). Ultimately, diverse and inclusive classrooms are better classrooms, and achieving greater diversity in classrooms begins with the adoption of more inclusive syllabi.

Yale College's mission is to "Educate talented students of diverse backgrounds to lead and serve in a complex and changing society," emphasizing the importance of a liberal arts education (18). To fulfill this and similar missions, universities must prepare students for the real world with coursework that reflects the diversity of modern society. This preparation requires not only interacting with diverse communities but also engaging with a wide range of perspectives in the authors and texts assigned in class. By introducing a mathematical framework to audit and diversify

syllabus content, we move toward creating a more equitable and enriching classroom experience for all students. We also improve the searchability of relevant, interdisciplinary books that are not initially accessible from a traditional library classification system. Thus, the framework's ability to highlight more relevant books potentially improves the utility of a syllabus.

1.3 Contributions

1.3.1 High Level Overview

Our project addresses the limitations of existing approaches to syllabus diversification by offering quantitative assessments of syllabus diversity alongside curated recommendations for diverse books through a user-friendly HTML interface. Users provide inputs by specifying a desired area of diversity, a list of books, and a metric to evaluate diversity. The tool also recommends books to add, giving instructors the flexibility to select between titles to incorporate into their syllabi.

This project focuses on diversifying syllabi for higher education contexts. However, the selection algorithm and diversity measurement framework presented have broader use cases. For instance, the proposed pipeline can be applied to libraries, where diverse content selection is an established problem (2). It may also be useful for students who wanted to engage with supplementary readings or academic centers who wanted to provide this tool as a resource to universities. Beyond books, this framework could be extended to suggest diverse case studies for business courses, inclusive patient scenarios for medical training, or culturally representative datasets for computer science projects. More broadly, we can employ our algorithmic approach to suggest diversified items in any context where sufficient differentiation between items exists.

1.3.2 Conceptual novelty

Our project advances syllabus diversification by addressing the key limitations of existing approaches and introducing innovative methods to measure and improve diversity. Previous efforts, such as the Augustana College Library audit (7) and Washington University's use of the Genderize API (8), relied heavily on manual data collection and binary gender classifications, providing limited insights into broader dimensions of diversity. Existing diversity measures often focus narrowly on simple percentages or binary definitions, failing to capture the full spectrum of identities and intersectionalities. Similarly, topic modeling approaches (3), while useful for some applications, are computationally intensive and unsuitable for analyzing full-length books due to restrictions on accessing data. Moreover, while recommender systems like those presented at the WSDM conference explore balancing accuracy and diversity (1), they lack explicit diversity metrics and actionable pathways for improvement.

In contrast, our project integrates a novel diversity auditing framework with a recommendation system that provides actionable suggestions for improving syllabus content. By leveraging the Open Library API, we automate subject and metadata retrieval, enabling a high-throughput solution that minimizes reliance on manual data entry. This not only expands the scope of measurable diversity beyond binary definitions but also provides tailored recommendations for users to diversify specific areas of underrepresentation. In doing so, we address both the limitations of existing tools and the practical challenges of improving syllabus diversity, offering a scalable, data-driven solution that bridges the gap between diversity assessment and actionable reform.

1.3.3 Technical novelty

Our algorithm presents the user with four metrics to diversify with. While several of these have direct applications to any set, using Rao's Entropy to audit diversity in a collection of books is novel. In addition, our demonstration offers a framework for users and policymakers to incorporate their preferred definitions of mathematical diversity by leveraging the integration of the Open Library Search API.

Although Rao's Entropy traditionally measures diversity across species, it can also be used to analyze the diversity of a syllabus. In this scenario, the "ecosystem" consists of two "species:" the overarching discipline of the syllabus and the area of desired diversity. In Rao's Entropy, a proportion of individual organisms within each species is calculated; similarly, the proportion of each tag (representing a topic in a book), is calculated. Thus, by finding Rao's Entropy for this "ecosystem," we can quantify the diversity of the syllabus while considering the "species" in the two disjoint sets. This provides a nuanced measurement of the balance and representation within the syllabus.

1.4 Diversity and Fairness Considerations

While fairness represents equitable treatment across groups, diversity focuses on representing varied identities and viewpoints to increase inclusivity. Thus, diversity does not imply fairness and vice versa. This project prioritizes diversity over strict fairness to ensure the inclusion of a wider range of perspectives. Our approach emphasizes diversity even if it disproportionately favors underrepresented areas, as long as it aligns with the instructor's goals for their syllabus. Unlike traditional diverse set selection algorithms that create maximally diverse subsets, our project focuses on recommending influential additions to existing syllabi rather than constructing an entirely new syllabus.

2 Related Prior Works

Diversity Measurement Tools

In order to fill in gaps in representation and expand the critical perspectives of their catalogs, researchers conduct diversity audits in collections of readings, including syllabi and libraries. However, these efforts involve manual data collection and/or percentages to attain. For example, a manual audit of the Augustana College library provided the library with accurate data on representational gaps based on the authors' identities (7). However, replicating this method in other libraries requires researchers to manually pull book and author information and provides no analysis beyond simple percentages about an author's protected attributes. For example, this method outputs the percentage of women authors on a syllabus.

A study at the Washington University in St. Louis collected a sample of syllabi from and developed the Genderize API to analyze instructor and author genders. (8). They then measure the percentage of female authors assigned in different disciplines and in courses taught by male or female instructors. While this method streamlines the process of auditing a bit, it also limits results to percentages. Further, the Genderize API uses U.S. census data to predict gender based on names and defines author gender based on a binary. This is a common approach, also seen in the Gender Balance Assessment Tool (17), which leveraged the `gender` and `predictrace` packages

in R to elucidate probable author gender and race. This method goes against feminist and queer scholarship, as it defines social identities in a mutually-exclusive manner (5).

Library Science for Book Classification

Our approach to understanding diversity focuses on bridging interdisciplinary studies and broader areas. The Library of Congress (LOC) Classification and the Dewey Decimal Classification (DDC) are two widely used systems for organizing books and other materials in libraries. These systems provide a standardized method for grouping books into distinct subject areas, enabling efficient retrieval and audits of library content. Both classification methods reflect the dominant academic disciplines of their time (both systems are from the 19th century), particularly emphasizing Western-centric and Eurocentric biases (6; 16). While these systems offer broad categorization capabilities, they map many interdisciplinary fields to a single classification number, limiting their flexibility and comprehensiveness. Library science researchers explore methods to make classification systems more inclusive and reflective of diverse perspectives. Efforts such as creating call guides for interdisciplinary studies, like LGBTQ+ Studies or Black Studies, help bridge some of these gaps (9). However, these approaches are manual and rely on human expertise to identify and classify relevant works and stand to benefit from an automated curation approach.

Diversity Improvement Tools

Importantly, no existing diversity measurement tool makes suggestions that improve upon the present diversity of a syllabus. In order to make recommendations which improve the diversity of a syllabus, we need to know the diversity of a piece of text. Research into measuring the diversity of a piece of text often focuses on access to the full text of the document. The paper "Text-Based Measures of Document Diversity" (3) uses topic modeling (Latent Dirichlet Allocation and Gibbs sampling) and a co-occurrence matrix to measure topic diversity using Rao's Quadratic entropy (see Section 3 for the definition of this metric). One success of this paper was that it suggests a ground truth diversity metric with high accuracy (AUC above 0.9) in practice. However, one key limitation in our ability to apply it to our case is that topic modeling entire books (instead of shorter articles, as done in the paper) is not feasible. For one, we lack access to the full-text versions of books, making a word co-occurrence approach unsuitable. Even with this access, topic modeling many books during our model's initial diversity measurement phase and item retrieval phase would be computationally expensive. Instead of topic modeling, we consider the Open Library API. Given a query (an ISBN number or an author and book title), it returns a list of subjects associated with that book. These subjects serve as a stand-in for this topic modeling approach by giving us an idea of the subjects in a book and allow us to use quadratic entropy to measure diversity.

Diversity Metrics

In order to measure the diversity of a syllabus with respect to a particular discipline area, we explore various metrics for quantifying diversity. We design our tool to be non-prescriptive in its ability to audit syllabus diversity and present multiple definitions of diversity to allow a user to gain a multi-faceted perspective on their syllabi. We employ four metrics, each one capturing a different amount of diversity. End users can choose to get recommendations through measuring (1) overlap, or the proportion of unique subjects in a syllabus, (2) relevance or the proportion of books

related to the area of desired diversity *at all*, (3) Jaccard Distance, which compares the *amount* of key words about an area of desired diversity to the topics of the syllabus or (4) Rao's Quadratic Entropy, which measures the amount of diversity of the topics in a syllabus by weighting them based on frequency and similarity.

Rao's Entropy (13) stands out among diversity measures because it incorporates similarity between categories, consideration of multivariate data, and computational efficiency for the project's scope and feasibility. Unlike other information theoretic approaches such as Shannon's Entropy (14) or Simpson's Index (15), Rao's Entropy explicitly accounts for the thematic similarity between categories in a collection. This is relevant because it weights the contribution to the overall syllabus diversity of two subjects that well-associated with one another differently than two disparate topics. Furthermore, while Shannon's Entropy measures uncertainty in a single dimension (richness), and Simpson's Index focuses on uniformity of distribution (fairness or evenness), Rao's Entropy accounts for both the number of distinct categories and the relative abundance of these categories.

3 Preliminaries / Problem Statement

3.1 Syllabus Structure

Traditionally, a syllabus is a collection of books, videos, articles, and other content. For our tool and framework, a syllabus S is defined as a collection of n books, where each book B_i is defined as:

$$B_i = (T_i, A_i, I_i)$$

For each book, T_i represents the title of the i-th book, A_i is the author, and I_i is the ISBN, a unique 13-digit identifier. As a result, a syllabus is formally defined as:

$$S = \{(T_1, A_1, I_1), (T_2, A_2, I_2), \ldots, (T_n, A_n, I_n)\}$$

In addition to the book identifiers, a syllabus S has a set of subjects $C = \{c_1, c_2, \ldots, c_f\}$, where f is the total number of subjects associated with topics, perspectives, or themes relevant to each book B_i in S. The subjects in C are further categorized as follows:

$C(\alpha) = \{c_1, c_2, \ldots, c_p\}$: Subjects obtained from books in the syllabus, which is related to the primary discipline α,
$C(\beta) = \{c_1, c_2, \ldots, c_q\}$: Subjects relevant to the area of desired diversity β.

Subjects are acquired from the Open Library API to evaluate thematic relevance.

3.2 Role of α and β

In this model, α represents the syllabus's primary area of focus (e.g., philosophy, sociology), while β denotes the area we aim to introduce for increased diversity (e.g., gender studies, Black studies). The algorithm balances diversity by augmenting S with additional relevant materials tagged with $C(\beta)$, while preserving the central relevance of $C(\alpha)$.

3.3 Diversity Measures

Four metrics are used to evaluate and compare the diversity of syllabi:

3.3.1 Rao's Entropy

Rao's Entropy quantifies diversity by considering the balance of categories in an ecosystem and their similarity. Unlike other metrics, it accounts for diverse topics and their dissimilarity, providing a nuanced measure of thematic spread. Rao's Entropy is adapted to measure the thematic diversity of a syllabus S. It captures the distinctiveness of subjects by weighting subject pairs based on their frequency and similarity:

$$\text{Rao}(S) = \sum_{i=1}^{|C(\alpha)|} \sum_{j=1}^{|C(\beta)|} p_i p_j \delta(i,j)$$

where p_i and p_j represent the probability of observing subjects i and j from $C(\alpha)$ and $C(\beta)$, respectively. The similarity $\delta(i,j)$ is calculated as $\delta(i,j) = \frac{x_i \cdot x_j}{\|x_i\|\|x_j\|}$, where x_i and x_j are subject embeddings normalized over the range $[0,1]$.

3.3.2 Relevance Proportion

The percentage of books in S with at least one subject in $C(\beta)$ quantifies the coverage of the area of desired diversity, allowing for a clear indication of representation within a syllabus. It differs from other metrics by focusing solely on coverage rather than evaluating the thematic or categorical spread of the content.

Let m be the number of books in S that contain at least one subject from $C(\beta)$ and let n be the total number of books in a syllabus S:

$$P_R = \frac{m}{n}$$

3.3.3 Overlap Proportion

This metric counts the unique subject topics represented in S, providing a measure of topic overlap between books. Unlike the metrics that account for balance or distribution, this metric only considers the variety of categories, making it a measure of topic breadth rather than thematic depth.

Let k be the number of unique subjects among all books in S, with an upper bound V, representing the representing the total number of subjects (non-unique) among all books:

$$P_B = \frac{k}{V}$$

3.3.4 Jaccard Distance

Jaccard Distance compares the non-overlap between two sets, such as the subjects associated with syllabus books ($c(\alpha)$) and those relevant to a desired diversity area ($c(\beta)$). It quantifies similarity by focusing on shared elements relative to the total elements in both sets. It differs from other metrics by concentrating particularly on the presence or absence of themes rather than their spread or balance

$$J(c(\alpha), c(\beta)) = 1 - \frac{|c(\alpha) \cap c(\beta)|}{|c(\alpha) \cup c(\beta)|}$$

3.4 Comparative Analysis of Diversity Metrics

Relevance Proportion and Overlap Proportion represent more simple, one-dimensional approaches to defining diversity. Overlap Proportion evaluates the extent to which the content (subjects) of the books within a syllabus coincide, offering a straightforward way to gauge the variety of topics represented, regardless of their relevance to the area of desired diversity. Relevance Proportion, on the other hand, assesses if the content of the syllabus is related to the area of desired diversity. However, this measure does not consider if the syllabus addresses the area of desired diversity in a comprehensive or balanced way. Jaccard Distance is more comprehensive, as it directly compares the syllabus content to the diversity area and quantifies the degree of overlap. This simultaneously captures how well the syllabus aligns with the desired diversity area and the distinctiveness of the ideas the syllabus contains. Lastly, Rao's Entropy expands on Jaccard Distance by incorporating the richness and relative abundance of the syllabus content, making it the most versatile and comprehensive metric for evaluating diversity in a wide range of contexts. Rao's Entropy also enables a deeper analysis by capturing subtle variations in content, making it particularly useful for identifying nuanced gaps or redundancies in the representation of diversity.

3.5 Recommender System for Syllabus Expansion

The recommender system combines the Library of Congress Classification and the Open Library Search API to identify books that meet both relevance and diversity goals. We create a query q to retrieve books that are in the same Library of Congress classification as the books on the syllabus $C(\alpha)$ and contain at least one subject from $C(\beta)$. The recommendation process screens potential books by measuring their impact on diversity, measured by Rao's Entropy, Jaccard Distance, Relevance Proportion, or Overlap Proportion. See Section 4 for the full algorithm.

3.6 Limitations

The proposed framework has the following limitations:

Library of Congress Classification Reliability. The LOC Classification assigns a primary classification to books based on dominant themes. As a result, a book may not have all its topic areas thoroughly represented by the classification despite the prominent subject in that book. To address this, we only test our model using the rich subject classification.

Subject Consistency. When using subjects from Open Library, differences in subject naming may affect the standardization of subject relevance across syllabi. To mitigate such an issue, the

recommender system tests with uniform probabilities for all words in a subject, and then another test with probabilities based on the number of mentions in the classification tree.

4 The Model/Framework

Syllabus input. The user will input a syllabus S which represents the readings for their course. In addition to the syllabus S, the user must enter the overarching discipline of the syllabus α, the desired diversity area within the syllabus β, and their preferred measure of diversity. For each book B_i in S, we query the Open Library Search API to retrieve the ISBN and update B_i accordingly.

Topic Extraction. For each book $B_i \in S$, we retrieve a set of subjects from the Open Library Search API and add them to a set of all subjects from the syllabus $C(\alpha)$. Subjects in $C(\beta)$ are defined according to keywords relating to the desired diversity area, as given in research on specific LOC classifications for diverse identities (9).

Diversity Calculation. The diversity of the syllabus S leverages one of four user-specified measures: Rao's Entropy, Jaccard Distance, Relevance Proportion, or Overlap Proportion.

Recommendation Process. We query the Open Library Search API using a query q to get a set of potential books R. The first book for the output set L is the book that independently maximizes or minimizes the diversity calculation. The remaining books are chosen in order of how well they diversify the new set of books L. These recommendations are output to the user. This is represented by the following algorithm.

Algorithm 1 Recommend Books

Input : number of recommendations k, syllabus S
Output: L: list of recommended books

$L \leftarrow [\,]$
R $\leftarrow$ `Run Query()`
best_book $\leftarrow$ `FindBestBook(`S, *books*`) // Select the first book based on syllabus` S
Append *best_book* to L, Remove *best_book* from *books*
while $length(L) < k$ **do**
    best_book $\leftarrow$ `FindBestBook(`L, *books*`) // Select the remaining books based on` L
    Append *best_book* to L, Remove *best_book* from *books*
return L
Function `FindBestBook(`*collection*, *books*`):`
    $max\_change \leftarrow 0$, $best\_book \leftarrow NULL$
    foreach *book* ***in*** *books* **do**
        diversity_change $\leftarrow$ `Diversity(`*collection* + *book*`)` $-$ `Diversity(`*collection*`)`
        if *diversity_change* $>$ *max_change* **then**
            $max\_change \leftarrow diversity\_change$
            $max\_change\_book \leftarrow book$
    return *best_book*

Outputs. As an output, the user will receive both a diversity score according to their original syllabus and preferred diversity measure and a list of books relevant to their course and the area of desired diversity to consider adding to their syllabus.

5 Theoretical Results

5.1 Definition of Diversity

Philosophically, diversity highlights the need for varied perspectives, experiences, and thoughts to challenge uniformity. Since the world is becoming more complex and heterogeneous every day, uniformity isn't a proper or accurate concept to represent it. Students attend classes to obtain knowledge about the world, and classes that only present or accept one perspective provide an incomplete understanding of the world. Thus, diversity is not merely about the presence of differences but about cultivating an understanding where those differences are valued and contribute meaningfully to collective outcomes.

Mathematically, diversity refers to the idea of a "spread" or distribution of vectors that minimizes redundancy and maximizes coverage in a given space. This concept can be visualized in terms of a set of vectors in a high-dimensional space, where the goal is to choose a new vectors, which represents a book, that is as distinct or dissimilar from the current set of vectors other as possible.

5.2 Rao's Entropy

Rao's Entropy quantifies diversity by considering not only the number of categories present but also how similar these categories are and how evenly distributed they are. It provides a comprehensive measure of diversity in a dataset by combining variety, balance, and relevance into a single metric. Though used in ecological situations, books, in this case, are similar to species. Given a list of subjects fiven by the syllabus $c(\alpha)$, a list of keywords relating to the area of desired diversity $c(\beta)$, the proportion of each keyword i in the area of desired diversity p_i, the proportion of each subject j in the overarching discipline area, and the cosine similarity $\delta(i,j)$ between each keyword i and each subject j normalized to $[0,1]$, we define Rao's Entropy for a syllabus S as follows:

5.2.1 Rao's Entropy Lower Bound

The lower bound for Rao's Entropy occurs when all items in the syllabus belong to the same category, meaning there is no similarity between the syllabus content and the area of desired diversity. In this case, $\delta(i,j) = 0$ for all $i \neq j$, and the entropy reduces to:

$$\text{Rao}(S) = \sum_{i=1}^{|C(\alpha)|} \sum_{j=1}^{|C(\beta)|} p_i p_j \cdot 0 = 0 \tag{2}$$

As a syllabus becomes more diverse, meaning the syllabus content becomes increasingly similar to the area of diversity by virtue of pairwise similarities $\delta(i,j)$ increasing for each $i \neq j$, Rao's Entropy approaches its upper bound.

5.2.2 Rao's Entropy Upper Bound

The upper bound for Rao's Entropy is derived by maximizing the similarity between any keyword i and subject j, thereby achieving the highest possible representation of the area of desired diversity β. In this case, $\delta(i,j) = 1$ for all $i \neq j$. Thus, we can now express Rao's Entropy in this case as:

$$\text{Rao}(S) = \sum_{i=1}^{|C(\alpha)|} p_i \sum_{j=1}^{|C(\beta)|} p_j \tag{1}$$

This simplifies under the guarantee that all probabilities p_i and p_j sum to 1. Hence, we obtain the upper bound:

$$\text{Rao}(S) = 1 \tag{1}$$

Thus, the bounds on Rao's Entropy are given by:

$$0 \leq \text{Rao}(S) \leq 1 \tag{2}$$

Usefulness: Rao's Entropy is most useful when trying to achieve a wholistic understanding of the diversity of a syllabus, while also taking into account the richness and relative abundance of its content. By virtue of the similarity measure $\delta(i,j)$, this measure is sensitive to the thematic content of the syllabus, capturing the presence of diverse elements in the direction of desired diversity as well as the degree to which they differ or overlap. Although Rao's Entropy is the most computationally intensive option, it provides a more complex, multidimensional assessment of the syllabus in comparison to the other three tools.

Limitations: The similarity measure $\delta(i,j)$ between keywords and subjects relies on subjective interpretations or computational approximations that inherently abstract the subjective interpretations of syllabus content. Further, Rao's Entropy assumes pairwise relationships between categories but does not account for hierarchical or multidimensional structures that may exist within the data. As a result, this measure may overlook complex interactions or contextual nuances that influence diversity in more sophisticated ways.

5.3 Relevance Proportion

This Relevance Proportion measures the proportion of books addressing the desired area of diversity. Given m as the number of books with at least one subject relevant to the area of desired diversity α and n as the total number of books in the syllabus, we define the percentage of books relevant to the area of desired diversity as follows:

$$P_R = \frac{m}{N}$$

5.3.1 Relevance Proportion Lower Bound

The lower bound for the percentage of books relevant to the area of desired diversity occurs when none of the books in the syllabus have subjects relevant to the desired diversity area. In this case $m = 0$, and substituting this into the equation for P_R returns:

$$P_R = \frac{0}{n} = 0 \tag{3}$$

As the diversity of the syllabus increases, meaning an increasing amount of books contain at least one subject in the area of desired diversity β, Relevance Proportion approaches its upper bound, described below.

5.3.2 Relevance Proportion Upper Bound

The upper bound for the percentage of books relevant to the area of desired diversity occurs when all books in the syllabus have at least one subject relevant to the desired diversity area. In this case $m = n$, and substituting this into the equation for P_R returns:

$$P_R = \frac{n}{n} = 1 \tag{3}$$

Thus, the bounds for P_R are given by:

$$0 \leq P_R \leq 1 \tag{4}$$

Usefulness: Relevance Proportion is most useful when seeking a straightforward and easily interpretable measure of whether or not a syllabus represents the area of desired diversity strongly. This metric is particularly well-suited for situations involving a large number of syllabi, where the goal is to perform an initial screening to quickly assess whether each syllabus reflects the desired diversity in significant proportions. From there, the user can flag syllabi with low proportions and prioritize assessing syllabi with higher proportions using more involved metrics such as Rao's Entropy.

Limitations: Relevance Proportion simplifies relevance to a binary classification, treating books as either strictly relevant or not to the desired area of diversity α. This does not allow for the consideration of varying degrees of relevance. It also measures only the presence of at least one relevant subject in a book, ignoring the depth or extent of coverage, which can inflate perceived diversity. Additionally, the metric assumes all books contribute equally to diversity, disregarding variations in their significance or emphasis within the syllabus.

5.4 Overlap Proportion

Overlap Proportion defines the proportion of unique subject topics represented in a syllabus. Given V as the total number of subject topics observed in the syllabus S and k as the total number of unique subclasses represented by the books in S, it measures the breadth of diversity by evaluating the variety of topics without considering balance or relevance. We define Overlap Proportion as follows:

$$\frac{k}{V} \tag{5}$$

5.4.1 Overlap Proportion Lower Bound

The lower bound for the number of unique categories, k, occurs when all books in the syllabus S belong to the same singular subject. In this case $k = 1$ and the Overlap Proportion is at its minimum value:

$$\frac{1}{V} \tag{6}$$

As the syllabus becomes more diverse—represented by an increase in the number of unique Open Library subjects——k increases toward its maximum possible value.

5.4.2 Overlap Proportion Upper Bound

The upper bound for the number of unique categories, k, occurs when every book in the syllabus contributes unique subclasses. In this scenario, the number of subclasses k are all unique, so $k = V$, and the Overlap Proportion reaches its upper limit:

$$k = V \tag{5}$$

Thus, the possible range for k is bounded by:

$$1 \leq k \leq V \tag{7}$$

As follows, the possible values for P_B are bounded by:

$$\frac{1}{V} \leq P_B \leq 1 \tag{8}$$

Usefulness: Overlap Proportion is most useful when trying to understand the range of topics covered in a syllabus, focusing less on the specific content and more on the overall breadth. This is well-suited for scenarios in which a user would like to compare diverse syllabi and target ones with more breadth, or when a user would like to identify syllabi that are highly specialized (i.e., those with high Overlap Proportions) and flag them as candidates for potential diversification efforts.

Limitations: Overlap Proportion does not take into account the area of desired diversity, and instead gives a measure of the spread of themes in the syllabus independent of β. Further, it overestimates diversity if categories are overly granular or poorly defined. Lastly, it ignores uneven distributions (e.g., one category dominating) and does not account for thematic or structural dissimilarities between categories.

5.5 Jaccard Distance

The Jaccard Distance compares the non-overlap between the list of subjects given by the syllabus $c(\alpha)$ and the list of keywords relating to the area of desired diversity $c(\beta)$. It quantifies dissimilarity by focusing on the elements that are unique to each set relative to the total elements in both sets. A lower distance represents a more diverse syllabus in the direction of desired diversity β.

5.5.1 Jaccard Distance Lower Bound

The lower bound for Jaccard Distance occurs when $c(\alpha)$ and $c(\beta)$ are identical, meaning there is complete overlap between the sets. In this case, $c(\alpha) \cap c(\beta) = c(\alpha) \cup c(\beta)$, which simplifies the Jaccard Distance expression to:

$$J(c(\alpha), c(\beta)) = 1 - \frac{|c(\alpha) \cap c(\beta)|}{|c(\alpha) \cap c(\beta)|} \tag{9}$$

Now, we solve for the lower bound:

$$1 - \frac{|c(\alpha) \cap c(\beta)|}{|c(\alpha) \cap c(\beta)|} = 1 - 1 = 0 \tag{10}$$

Jaccard Distance approaches its upper bound as the content of the syllabus becomes less aligned with the content of the area of desired diversity because of the elements in $c(\alpha)$ containing fewer elements in common with $c(\beta)$.

5.5.2 Jaccard Distance Upper Bound

The upper bound for Jaccard Distance occurs when there is no overlap between $c(\alpha)$ and $c(\beta)$, i.e., $c(\alpha) \cap c(\beta) = \emptyset$. In this case:

$$J(c(\alpha), c(\beta)) = 1 - \frac{0}{|c(\alpha) \cup c(\beta)|} = 1 \tag{8}$$

Thus, the bounds on Jaccard Distance are given by:

$$0 \leq J(c(\alpha), c(\beta)) \leq 1 \tag{11}$$

Usefulness: Jaccard Distance is most useful when you would like a measure of diversity and relevance to the area of desired diversity, but lack the computational power required for Rao's Entropy. Jaccard Distance provides a valuable proxy for Rao's Entropy and can serve as an initial pass through a syllabus to detect whether there is overlap between its content and the area of desired diversity. Furthermore, it balances considerations of both relevance and breadth in a way that Relevance Proportion and Overlap Proportion fail to assess as standalone metrics and in a more easily interpretable manner than Rao's Entropy.

Limitations: The Jaccard Score is sensitive to the size of the sets; larger sets may artificially inflate the score. Additionally, it does not account for the relevance or semantic relationships between elements, only their presence or absence.

6 Empirical Results

6.1 Demonstration

As part of our results, we constructed an online interface and syllabus parsing tool. This allows potential users, namely college course instructors, to input their syllabi as a Word document and receive a list of recommended titles via the proposed framework. The full code can be found here[1].

On the first page, the user is presented with a description of how to use our tool and its purpose. Then, we describe in simple language the purpose of each of the four measures of diversity and allow the user to choose between them. The code base offers flexibility to integrate new measures of diversity in the future. Subsequently, the user must define the parameters of the model, as in Section 4, which include a syllabus S, the overarching discipline α and the area of desired diversity β, and lastly the desired diversity metric that will be used to audit their syllabus.

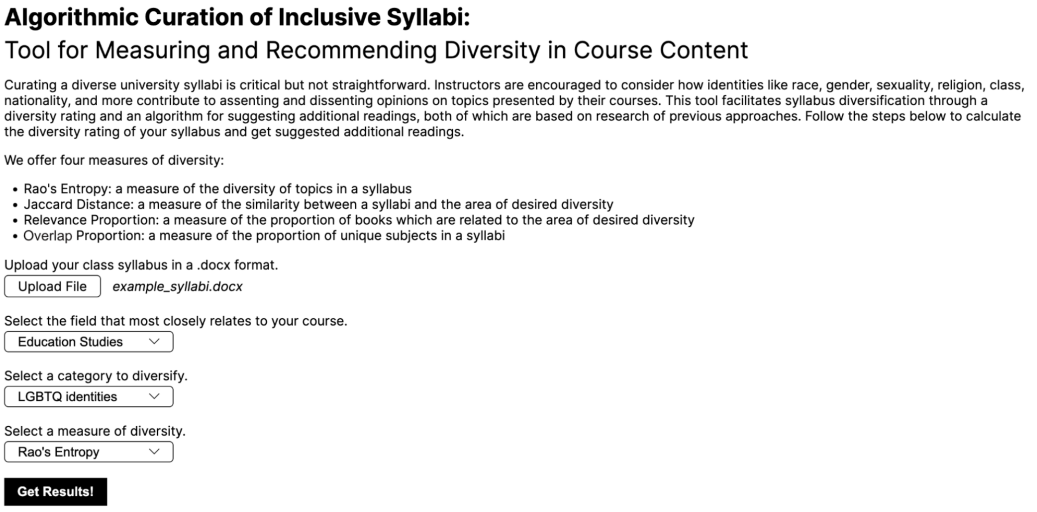


Figure 1: Landing Page for User Definition of Model Parameters.

On the second page, the tool offers an opportunity for the end user to validate the titles and ISBN numbers of the books within their syllabus to account for potential parsing errors.

[1] https://github.com/mark-ayiah/CPSC-464-Inclusive-Syllabi

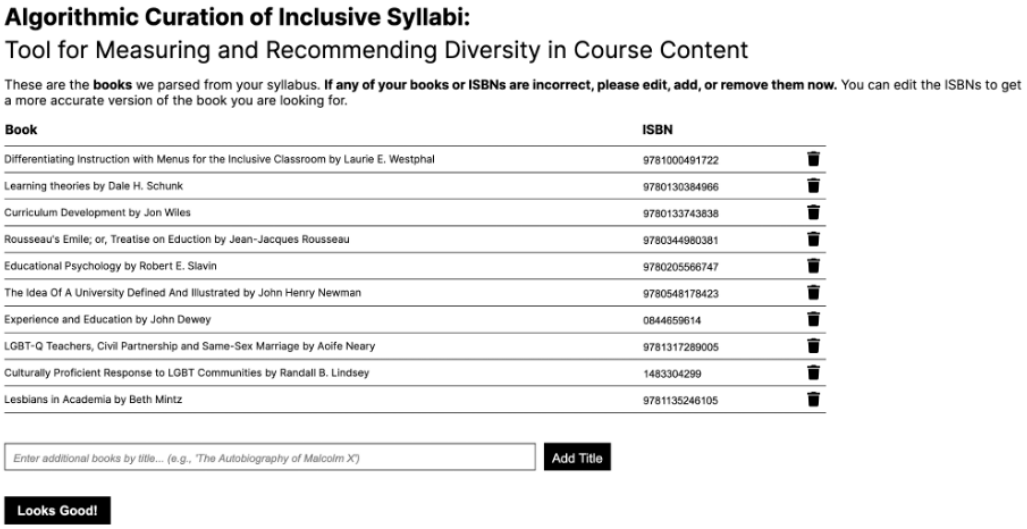


Figure 2: User Corrects Any Parsing Errors.

On the last page, the user is presented with a diversity rating calculated using the metric they previously selected. The user then receives recommended titles that can be used to diversify their syllabus in their selected area. The user can reload the tool and choose a different selected area of diversity to quickly furnish a long list of suggested books to add to their syllabus.

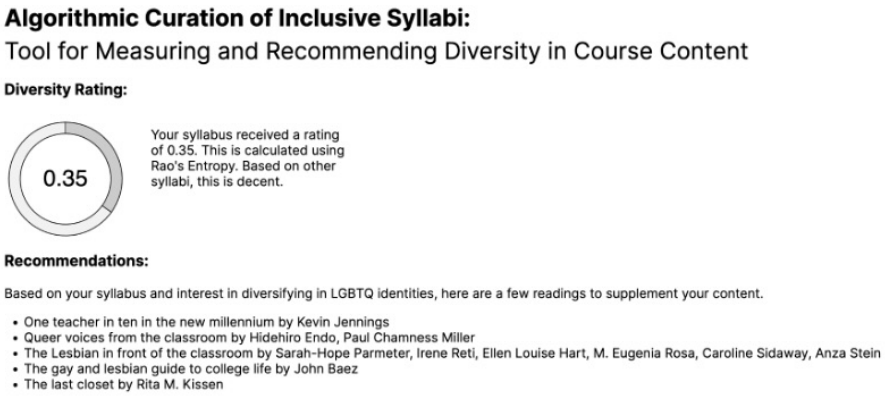


Figure 3: User Receives a Diversity Rating and a List of Potential Syllabus Additions.

6.2 Test Data Description

To test our diversity measures, we constructed a scenario in which a professor of an Education Studies (α) course wishes to diverse their syllabi to include more LGBTQ perspectives (β) (11). We designed four syllabi with varying levels of diversity in relation to the area of diversity β. Each syllabus has 10 books and and the proportion of books relevant to β is as follows: 0% of books in Syllabus 1 (Low) are relevant to β, 30% of books in Syllabus 2 (Low-Medium) are relevant to

β, 60% of books in Syllabus 3 (Medium-High) are relevant to β, and 100% of books in Syllabus 4 (High) are relevant to β.

When calculating Rao's Entropy and the Jaccard Distance, we subset for the top 10 key words of the syllabus and the top 5 key words of the diversity area. We settled on these values after a heuristic check of how they split the proportion of key word occurrences and the perceived quality of the key terms extracted.

6.3 Metrics.

Given that the outputs of the diversity measures are not directly comparable, we assess them independently on the four syllabi. The direction and rate of change of an effective diversity measure should correspond with the difference in diversity between syllabi.

6.4 Results

Rao's Entropy

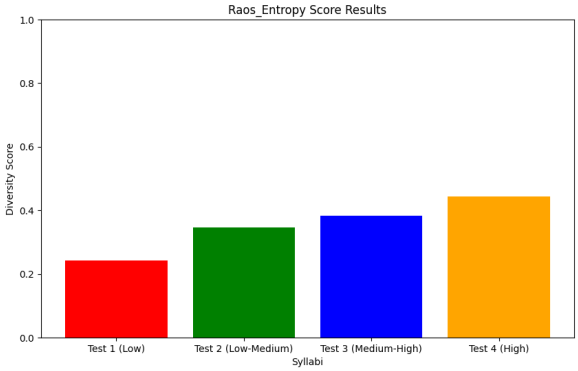


Figure 4: Rao's Entropy Performance on Different Syllabi.

As the number of books in the syllabus related to β increases, Rao's Entropy increases. This is in line with expected results. While the scores are concentrated between 0.2 and 0.5, this measure seems to capture increases in diversity, although it is not as strongly skewed towards its theoretical upper and lower bounds.

One key factor affecting the calculation is the embedding distance between key words in the syllabus and the discipline area. In the Low syllabus, the quadratic entropy is about 0.237. Looking at the embedding distance matrix, the lowest similarity is between the key words "lesbian" and "ability", with a cosine similarity score of 0.121. Generally, none of the similarity scores for the Low syllabus are greater than 0.43. These bounds on the similarity score explain why the Low syllabus is perceived as mildly diverse, despite having no subject term overlap with the diversity area. It is only when the syllabi contain subjects in the discipline area that any of the pairwise similarity scores, as expected, exceed 0.5.

Anther factor that affects the metric's sensitivity towards diversity is the size of the set of key words for the diversity area. The current version uses only the topic five key words ("gay", "lesbian",

"transgender", "culture", and "same-sex"), based on their frequency in LOC classification headers of other books within the same class ones on each syllabus. This produces Rao's Entropy scores which have greater distances from one another as diversity increases. In a separate version, we input up to 11 key words, including "bisexuality", "lgbt", and "queer". Because we selected these words ad hoc, the the distribution of the words was uniform. We, thus, observed a smaller difference between the 4 syllabi. Further, we found that the more keywords there were, the smaller the distance observed between the entropy of each syllabus.

Jaccard Distance

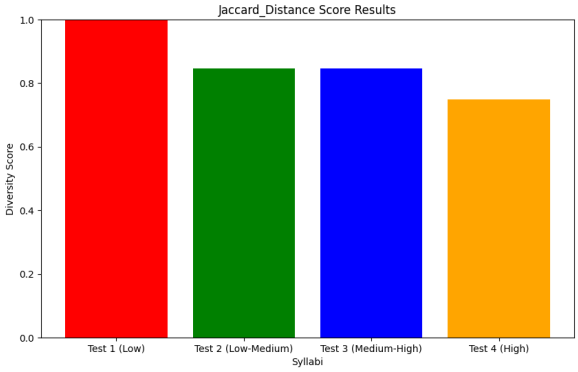


Figure 5: Jaccard Distance Performance on Different Syllabi.

As expected, when the syllabus diversity increases, the Jaccard Distance between the α and β decreases. Although the Low diversity syllabus had the greatest Jaccard Distance (reaching the upper bound of 1), the metric gave high distance scores to all syllabi, as they were all 0.75 or greater. This makes sense, as there are only 5 key words in the diversity area and 10 in the discipline area; a score of 0.75 means that there were 4 key words from the diversity area (distance = $\frac{11}{15}$) in the syllabus' books.

Overlap Proportion

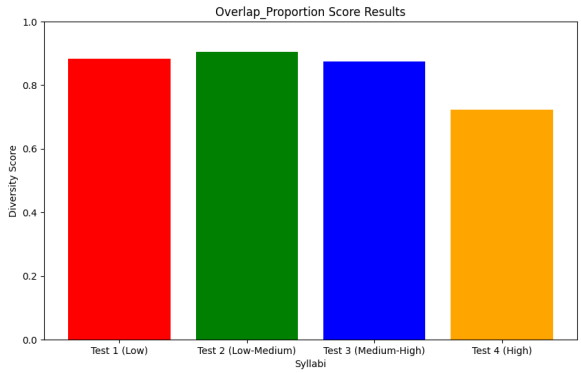


Figure 6: Overlap Proportion Performance on Different Syllabi.

Looking at raw subject tag counts also does not illuminate the diversity of a book collection. For the Low diversity syllabus, there are 28 distinct words that showed up in the subject tags for the book. The Low-Medium diversity syllabus has 35 distinct words, the Medium-High one has 36 distinct words, while the High diversity syllabus has 28 distinct words. Empirically, we see that there is little association between diversity and the raw number of distinct subjects. This is logically consistent with the idea that these subjects are moreso to benefit the searcher, rather than being indicative of a given book's main focus. This also further demonstrates the utility of Rao's Entropy, as it weights words more likely to be relevant higher in the overall calculation of diversity.

From this, it is clear that calculating the percentage of unique categories produces distinct results for each book collection. However, these outputs are not fully representative because this measure does not take into account the area of desired diversity β. While a higher percentage of unique LOC subclasses may indicate broader diversity, it does not necessarily reflect diversity within β. It is also only interpretable in the context of other syllabi. Therefore, the output can be difficult for the user to understand.

Relevance Proportion

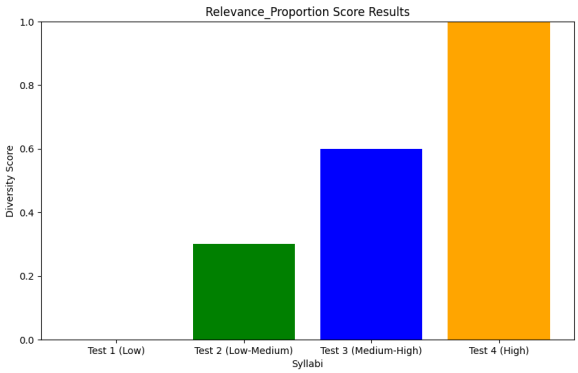


Figure 7: Relevance Proportion Performance on Different Syllabi.

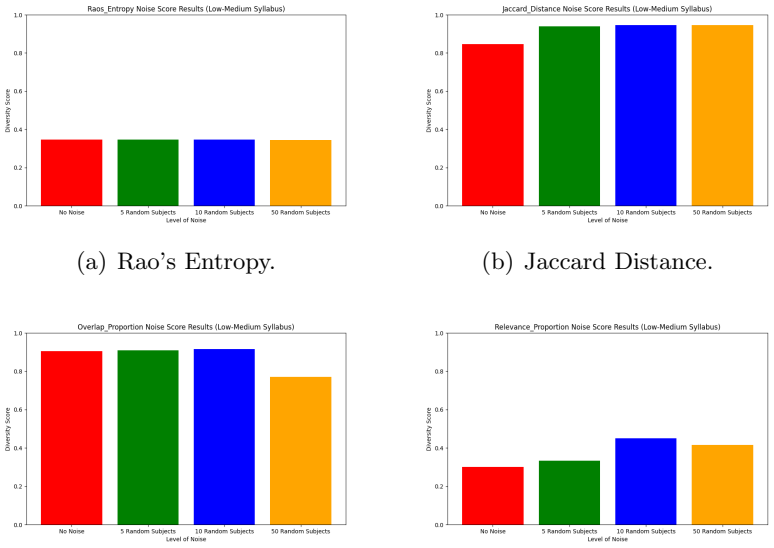
Although the percentage of books relevant to β can provide a general indication of how inclusive a syllabus is, it lacks nuance and does not account for the depth or strength of a book's relevance.

The proportions are explicitly defined to be 0%, 30%, 60%, and 100% in this test. This is helpful for understanding our test case, but is a difficult metric to maximize over in order to provide the best recommendations, beyond the order which Open Library returns by default.

7 Discussion of Empirical Results

Robustness

It is important to evaluate the robustness of these measures to random noise as this enables us to interpret the outputs of these metrics on syllabi with unknown relevance proportions. For these experiments, we evaluated their ability to withstand perturbation with irrelevant subjects on the Low-Medium diversity syllabi. We chose this syllabus because it represents the most likely case in which a syllabus has some books related to the area of desired diversity.



(a) Rao's Entropy.

(b) Jaccard Distance.

(c) Overlap Proportion.

(d) Relevance Proportion.

Figure 8: Performance of Metrics on Low-Medium Syllabus with 0, 5, 10 and 50 Random Subjects.

Some of the robustness tests help clarify the empirical results from the test syllabi. We see in Figure 9(a) that adding random subjects does not affect the entropy of the Low-Medium syllabus, as we look at a subset of the most common subjects. This presents Rao's Entropy as a metric that is resistant to noise, because it only takes the top N subjects. In contrast, we see in the other 3 figures that the perturbations affect the output. For Jaccard Distance, adding random words tends to increase the distance, which is expected, as these words may not be related to the diversity area. This makes it slightly less robust than Rao's Entropy, as the lack of consideration for relative abundance makes it less able to "throw out" or discount irrelevant words in the syllabus.

The Overlap Proportion appears robust to small additions (5 to 10 random subjects), but adding 50 random subjects leads to a decrease, when it *should* stay constant or increase. The Relevance Proportion is sensitive to additional subjects, but in a nonlinear manner. Given that it depends on the count of books with relevant subjects, randomly adding a subject in the diversity area to a book that has no relation to the diversity area may randomly increase the relevance proportion.

Overall, the metrics may need to be scaled by a nonlinearity or a constant value. The latter could be a subject sensitivity threshold that the end user adjusts in order to get more interpretable outputs.

Performance of Recommendations

For the recommendation step, we query the Open Library API as described in Section 3.5 and parse the results in order to gauge how well the metrics perform at selecting recommendations. We use the Low-Medium syllabus as the base case and apply the algorithm from Section 4 to return three books that might be acceptable for the end user.

			Objective Function			
Metric	Prior Score		Rao's	Jaccard	Relevance	Overlap
Rao's	0.347	Amount of Change	+0.029	+0.004	+0.014	-0.002
Jaccard	0.846		0	0	0	0
Relevance	0.3		+0.162	+0.117	+0.033	0
Overlap	0.906		-0.031	-0.009	-0.036	+0.002

Table 1: Change in Score Based on 3 Recommended Books. The columns mark the metric used to find the suggested books (objective function), while the rows show the prior score and how maximizing relative to each metric affects the score seen in the corresponding column of the table. The diagonal of the table shows how the prior score changed relative to the metric used to generate the recommended books.

All of the objective functions, with the exception of the Overlap Proportion, change the values of the other metrics in the desired directions.

Using Rao's Entropy as the objective function increases itself and the relevance proportion the most. While it is not ideal that it was unable to significantly (relative to random perturbation) improve its value, individual books seem to have a greater effect on the entropy. On its own, the first book recommendation increases Rao's Entropy by 0.05, which is relatively significant given that the example High diversity syllabus is 0.10 points above the Low-Medium prior score. It seems that considering all of the books may decrease the richness, leading to the decrease in overall added entropy. This makes it a useful measure if the end user only wants to add a few books for a small unit about the area of desired diversity, rather than making the whole course about the overlap between the diversity and the discipline area.

The relevance proportion was able to minimize the overlap proportion the most of the 4 variables. This potentially indicates a strong relationship between relevance and overlap that could be explored using a linear combination of the two variables. We could then use submodular maximization to determine which books are relevant to the area of desired diversity while minimizing the amount of new, irrelevant subjects that are introduced.

Regardless of objective function, the Jaccard Distance of the syllabus did not change in the presence of all three new recommendations. This is true regardless of the objective function, but surprising because the first book recommendation, when optimizing for Jaccard Distance, decreased the score by about 0.07. Adding any book which *could* decrease the distance *should* decrease it, regardless of the other book suggestions.

8 Conclusion

This study introduces a framework for auditing and enhancing syllabus diversity, leveraging Rao's Entropy, among other metrics, to capture both thematic richness and dissimilarity. The tool also leverages library science and an open source API, making its underpinnings more accessible to those who want to improve on the framework.

Through theoretical analysis and empirical evaluations, we demonstrate how our framework provides actionable recommendations for diversifying syllabi while preserving academic relevance.

The proposed tool addresses the limitations of existing approaches by automating diversity audits and offering tailored suggestions for areas of improvement, empowering educators to create more inclusive learning environments.

Limitations. The framework faces several challenges. It relies heavily on metadata from the Open Library API and Library of Congress classification system, which may not fully capture the complexity of a given book. These classification systems necessarily categorize books into a set of tags or a book code and are inherently simplistic measures of texts. In thinking about the themes of this course, the idea that data can never be objective applies to our classification system; books cannot be captured using quantitative metrics, and any representation of a book in the model contains assumptions and biases that are hidden to the end user. These biases might disproportionately affect interdisciplinary, non-Western books. Concretely, the LOC Classification provides more detailed information for English vs. non-English titles. The Open Library Search system provides more subject areas on popular books as compared to less-well known books. Bias mitigation within the recommendation system we propose is not within the scope of our project, yet it is important to recognize in our work.

Moreover, the current implementation focuses exclusively on books, excluding other critical materials such as journal articles and multimedia resources that many syllabi often employ. The tool we provide to users requires books to be entered using a parsing tool that is unable to accept information other than an ISBN code for a book. Furthermore, language and cultural barriers further constrain the applicability of our method, as non-English titles and tags are not accounted for in this version of our product.

8.1 Future Work

Future work will expand the framework's scope by incorporating multilingual support, additional data sources, and non-pairwise diversity metrics such as Determinantal Point Processes to address these limitations. Integrating advanced machine learning techniques to validate and enrich metadata will also enhance the reliability of diversity measures and recommendations.

Additions to the interface will allow users to specify publishing date ranges and expand the discipline and diversity areas available for end users to choose from by using librarian-approved LOC call guides. It is also possible to revisit the tree-like structure of the LOC classification system to improve key word search and generation. There is also an opportunity to improvement the model's interpretability by using a nonlinearity to scale the outputs towards more understandable values, or considering the random input base cases before displaying the final result.

We've conducted preliminary conversations with the PoorVu Center to explore how this tool could be integrated into teacher training or resources on Diversity, Equity, and Inclusion. By addressing these challenges and working with area experts, the framework has the potential to evolve into a versatile tool for fostering inclusivity in academia and beyond, with applications spanning library curation, educational platforms, and content recommendation systems.

Acknowledgements. ChatGPT was used to improve the grammar and clarity of sections throughout this report. For example, using prompts such as "How can I clarify this paragraph?" or"How can I improve this sentence?"

References

[1] Agrawal, R., Gollapudi, S., Halverson, A., and Ieong, S. Diversifying search results. In *Proceedings of the Second ACM International Conference on Web Search and Data Mining* (New York, NY, USA, 2009), WSDM '09, Association for Computing Machinery, p. 5–14.

[2] American Library Association. Diverse collections: An interpretation of the library bill of rights, 2006.

[3] Bache, K., Newman, D., and Smyth, P. Text-based measures of document diversity. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY, USA, 2013), KDD '13, Association for Computing Machinery, p. 23–31.

[4] Dasgupta, N., and Stout, J. G. Girls and women in science, technology, engineering, and mathematics: Steming the tide and broadening participation in stem careers. *Policy Insights from the Behavioral and Brain Sciences 1*, 1 (2014), 21–29.

[5] Devinney, H., Björklund, J., and Björklund, H. Theories of "gender" in nlp bias research. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (New York, NY, USA, 2022), FAccT '22, Association for Computing Machinery, p. 2083–2102.

[6] Drabinski, E. Queering the catalog: Queer theory and the politics of correction. *The Library Quarterly 83*, 2 (2013), 94–111.

[7] Emerson, M. E., and Lehman, L. G. Who are we missing? conducting a diversity audit in a liberal arts college library. *The Journal of Academic Librarianship 48*, 3 (2022), 102517.

[8] Harris, J. K., Croston, M. A., Hutti, E. T., and Eyler, A. A. Diversify the syllabi: Underrepresentation of female authors in college course readings. *PLOS ONE 15*, 10 (10 2020), 1–14.

[9] Howard, S. A., and Knowlton, S. A. Browsing through bias: The library of congress classification and subject headings for african american studies and lgbtqia studies. *Library Trends 67*, 1 (2018), 74–88. Volume 67, Number 1, Summer 2018.

[10] Office of Planning, Evaluation and Policy Development. Advancing diversity and inclusion in higher education. 1–75.

[11] Olorunisola, D., Ayiah, M., Moosa, A., Aboyure, W., Hussein, A., Shakes, B., and Muita, L. Cpsc 464 inclusive syallbi, 2024.

[12] Price, I., and Regehr, G. Barriers or costs? understanding faculty resistance to curricular change. *Canadian Medical Education Journal 13*, 3 (Apr. 2022), 113–115.

[13] Rao, C. Diversity and dissimilarity coefficients: A unified approach. *Theoretical Population Biology 21*, 1 (1982), 24–43.

[14] Shannon, C. E. A mathematical theory of communication. *The Bell System Technical Journal 27*, 3 (1948), 379–423.

[15] Simpson, E. H. Measurement of diversity. *Nature 163* (1948), 688–688.

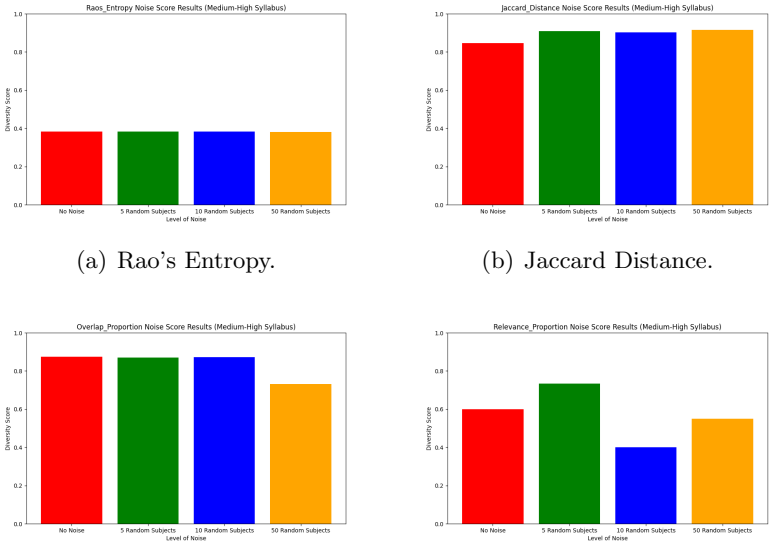
[16] Smith, T., and Appleton, L. Addressing classification system bias in higher education libraries in england. *portal: Libraries and the Academy 23*, 4 (2023), 819–841. Accessed: 2024-11-10.

[17] Sumner, J. L. The gender balance assessment tool (gbat): A web-based tool for estimating gender balance in syllabi and bibliographies. *PS: Political Science 38; Politics 51*, 2 (2018), 396–400.

[18] Yale University. Mission statement of Yale College, 2023.

[19] Zirkel, S. Is there a place for me? role models and academic identity among white students and students of color. *Teachers College Record 104*, 2 (2002), 357–376.

9 Appendix



(a) Rao's Entropy.

(b) Jaccard Distance.

(c) Overlap Proportion.

(d) Relevance Proportion.

Figure 9: Performance of Metrics on Medium-High Syllabus with 0, 5, 10 and 50 Random Subjects.