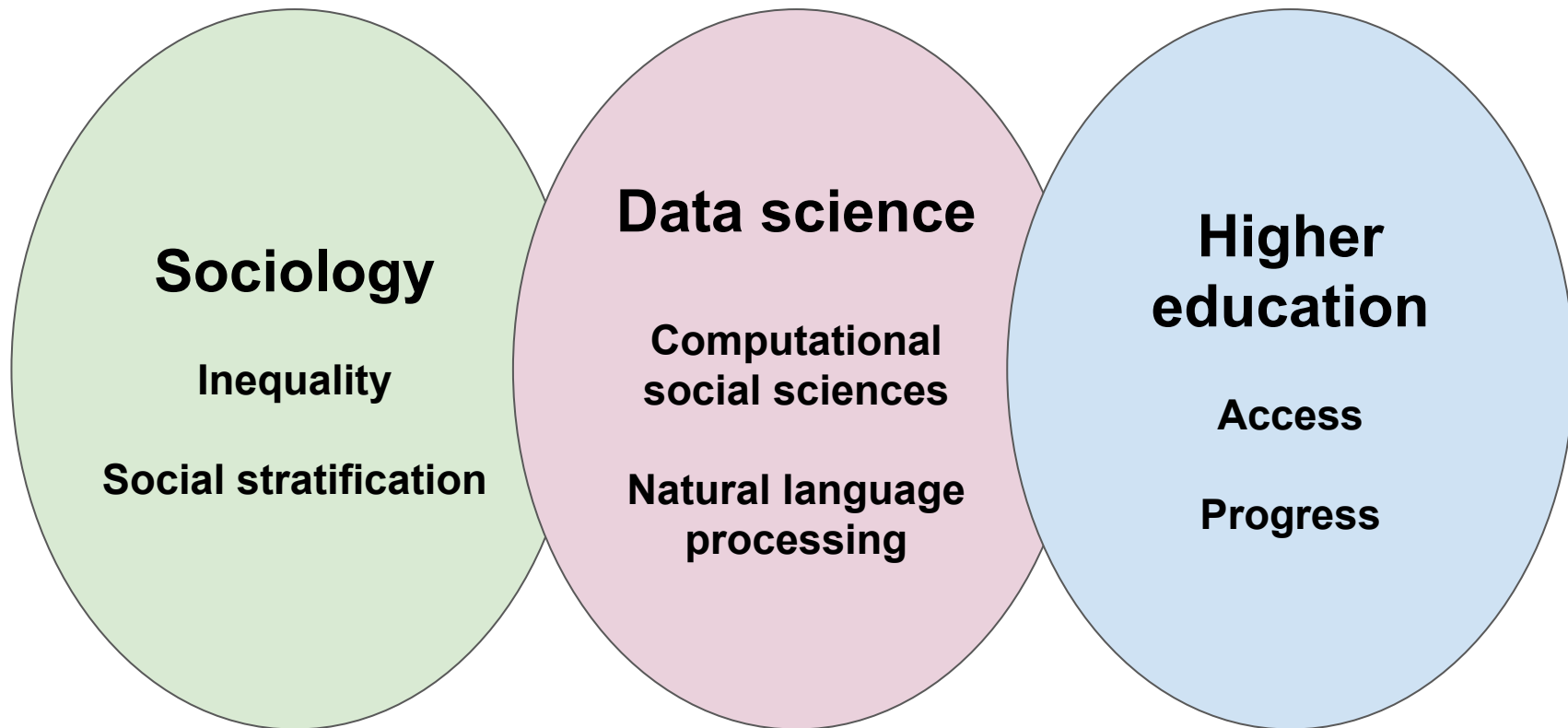# Defining Diversity Discourse in Global Higher Education: Computationally Comparing the USA, India, the UK, and Germany

Léon Marbach
Stanford University

Academic interests

Sociology

Inequality

Social stratification

Data science

Computational social sciences

Natural language processing

Higher education

Access

Progress

Academic background

BS in Applied Maths & CS Sorbonne University	BA in Social Sciences Sciences Po
MA in Public Policy Sciences Po	PhD in Education Data Science MS in Statistics Stanford University

Research experiences

Influence of peers, parents and professionals in higher education choices in France

Access to and progress in pre-medical studies in France

Comparative analysis of university discourses on diversity

TODAY!

Algorithms and college admissions (dissertation project)

This project

- International collaboration with **Sonia Giebel** (WZB), **Yasemin Soysal** (WZB), and **Saurabh Khanna** (University of Amsterdam)

- Motivation:

  - Universities play an important role in espousing and enacting diversity and inclusion worldwide
  - Universities as both global and local institutions
  - Like all cultural ideas, diversity is subject to contestation and redefinitions

- **RQ: What does diversity mean and how does it manifest in contemporary global higher education?**

- Study of university websites in 4 countries: USA, UK, Germany, and India

- Data collection fall 2023-winter 2024, now: analysis begins…

What is topic modeling?

- Useful when you have a large quantity of text and it is impossible to human read them all

- Goal is to identify topics from a set of documents known as a corpus

- Documents as collection of topics and topics as collection of words

- Completely inductive method, unsupervised because no predefined tags or training data

- Several different possible algorithms (e.g. LDA)

- Increasingly used in the social sciences to study newspaper articles, tweets, academic abstracts, etc.

Topics

gene 0.04
dna 0.02
genetic 0.01
, , ,

life 0.02
evolve 0.01
organism 0.01
, , ,

brain 0.04
neuron 0.02
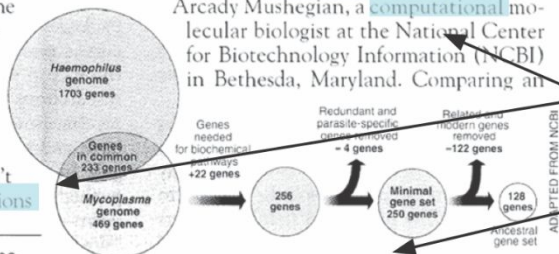nerve 0.01
...

data 0.02
number 0.02
computer 0.01
, , ,

Documents

Seeking Life's Bare (Genetic) Necessities

COLD SPRING HARBOR, NEW YORK— How many genes does an organism need to survive? Last week at the genome meeting here,* two genome researchers with radically different approaches presented complementary views of the basic genes needed for life. One research team, using computer analyses to compare known genomes, concluded that today's organisms can be sustained with just 250 genes, and that the earliest life forms required a mere 128 genes. The other researcher mapped genes in a simple parasite and estimated that for this organism, 800 genes are plenty to do the job—but that anything short of 100 wouldn't be enough.

Although the numbers don't match precisely, those predictions

"are not all that far apart," especially in comparison to the 75,000 genes in the human genome, notes Siv Andersson of Uppsala University in Sweden, who arrived at the 800 number. But coming up with a consensus answer may be more than just a genetic numbers game, particularly as more and more genomes are completely mapped and sequenced. "It may be a way of organizing any newly sequenced genome," explains Arcady Mushegian, a computational molecular biologist at the National Center for Biotechnology Information (NCBI) in Bethesda, Maryland. Comparing an
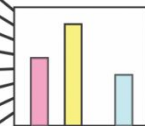
Haemophilus genome 1703 genes

Genes in common 233 genes

Mycoplasma genome 469 genes

Genes needed for biochemical pathways +22 genes

Redundant and parasite-specific genes removed -4 genes

256 genes

Minimal gene set 250 genes

Related and modern genes removed -122 genes

128 genes

Ancestral gene set

ADAPTED FROM NCBI

Stripping down. Computer analysis yields an estimate of the minimum modern and ancient genomes.

* Genome Mapping and Sequencing, Cold Spring Harbor, New York, May 8 to 12.

SCIENCE • VOL. 272 • 24 MAY 1996

Topic proportions and assignments

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77-84.

Step 1: Webscraping

- Sample comprises:

    - All universities in the UK **(N = 164)**

    - All universities in Germany except vocational schools **(N = 328)**

    - A representative sample of universities in India **(N = 250)**

    - A representative sample of universities in the USA **(N = 260)**

- Homepages and <u>one-click-away</u> from homepages

    - Is it even legal? IRB approval

    - Takes a lot of computational power!

    - Some websites blocked us

Step 2: Embeddings (BERTopic, Grootendorst 2022)

Algorithms do not understand text! Transforming our input documents into numerical representations using sentence-transformers.
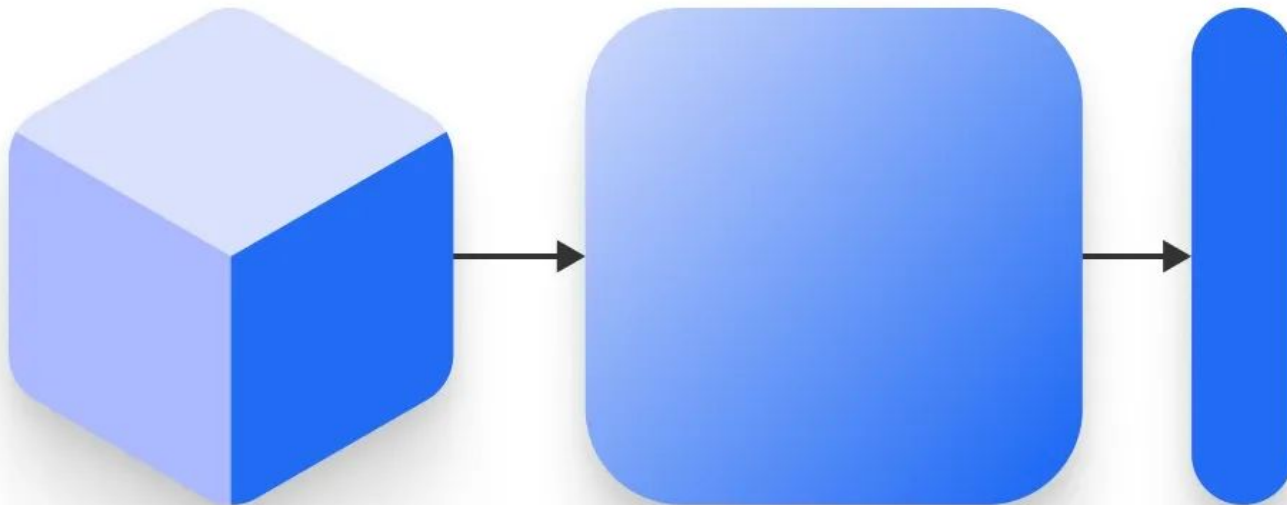
"Lion is the king of the jungle."

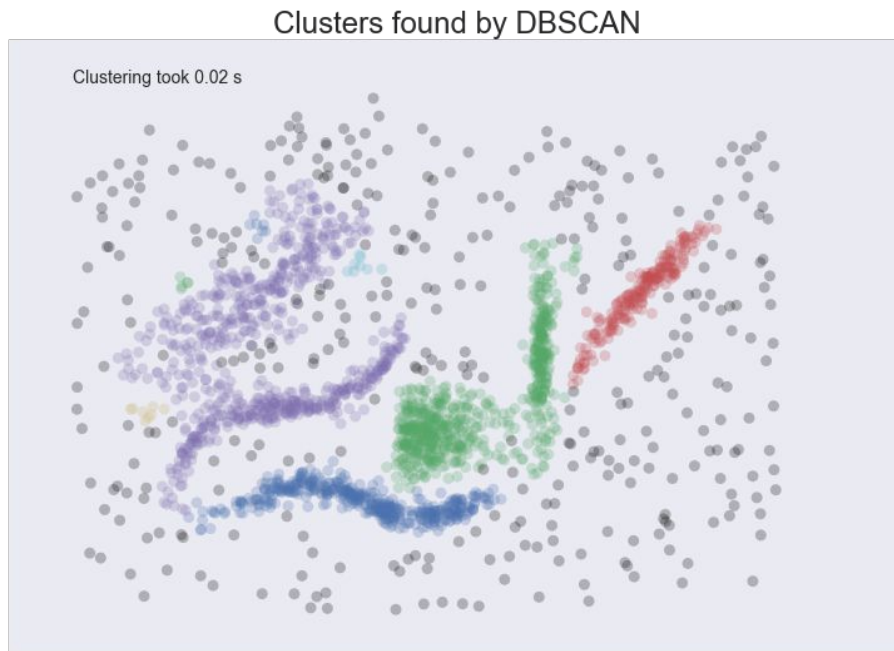"The tiger hunts in this forest."

"Everybody loves New York."

Step 3: Dimensionality Reduction

Due to curse of dimensionality, we reduce the dimensions of embeddings.

Step 4: Clustering

After reducing the dimensionality of our input embeddings, we need to cluster them into groups of similar embeddings to extract our topics.
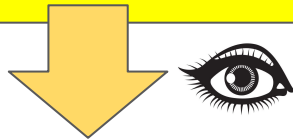
Clusters found by DBSCAN

Clustering took 0.02 s

Step 5: Topic representation

TF-IDF: what words are most distinctive of each topic → we get a list of words or expressions and we need then to make sense of them (read and interpret)

Example:

['abroad students', 'international students', 'studying abroad', 'international admissions', 'abroad experience', 'international student', 'abroad advisor', 'global experience', 'abroad abroad', 'global engagement']

International students and travel abroad

Final output

The probability that Doc 1 is about Topic 1 is 0.1
= 10% of Doc 1 talks about Topic 1

Total number of topics

	Topic 1	Topic 2	Topic 3	…	Topic X	Total
Doc 1	0.1	0.0	0.5	…	0.0	**1.0**
Doc 2	0.0	0.0	0.0	…	1.0	**1.0**
Doc 3	0.16	0.03	0.06	…	0.04	**1.0**
…	…	…	…	…	…	**…**
Doc N	0.67	0.04	0.03	…	0.07	**1.0**

Total number of documents

Probabilities in each row sum up to 1.0

USA: topics most closely related to "diversity"

Correlation	Topic name	Associated phrases
0.56	Diversity and inclusion	['inclusion diversity', 'diversity inclusion', 'diversity equity', 'equity inclusion', 'equity diversity', 'inclusive community', 'inclusive excellence', 'diversity', 'diverse', 'office diversity']
0.54	Multicultural	['multicultural student', 'multicultural affairs', 'cultural committee', 'office multicultural', 'multicultural', 'diversity', 'multicultural mcc', 'hall students', 'students faculty', 'empowerment club']
0.50	Indigenous and native communities	['humanities social', 'humanities', 'social sports', 'native indigenous', 'greenfield intercultural', 'powwow natives', 'activist', 'natives', 'activist following', 'natives event']
0.47	Black students	['celebrates black', 'mlk week', 'black month', 'black student', 'black movement', 'mlk', 'black writers', 'african american', 'civil rights', 'luther king']
0.47	Scholarships	['diversity graduation', 'donors faculty', 'future students', 'students need', 'students donors', 'big scholarship', 'centennial student', 'color institutional', 'organizations 52', '15 graduates']
0.46	Graduate students	['uncommon colleges', 'classesacademic', 'grads', 'grads life', 'grads quick', 'making difference', 'ready grads', 'athletics culture', 'culture research', 'scholarships aid']
0.46	?	['cultural social', 'honorees', 'provost research', 'vice provost', 'volunteer academic', 'provost', 'essay', 'athletic career', 'spiritual cultural', 'academic performances']
0.45	Campus life and advising	['student leaders', 'archives academic', 'student live', 'provide advisors', 'advising topics', 'student trainigs', 'grades research', 'student availabe', 'schools sawyer', 'interfaith weekly']

Germany: topics most closely related to "diversity"

Correlation	Topic name	Associated phrases (translated to English)
0.59	Diversity strategies and certificates	['diversity strategy', 'equal opportunities diversity', 'diversity management', 'diversity equality', 'diversity price', 'diversity', 'diversity audit', '2023 diversity', 'diversity dimensions', 'diversity']
0.52	Gender equality and engagement	['gender diversity', 'equality diversity', 'gender equality', 'prospective students', 'equal opportunities', 'diversity measures', 'regional student secretariat', 'economic social sciences', 'have a say in campus life', 'diversity']
0.47	Family-friendliness	['diversity activities', 'family-friendly diversity', 'diversity cosmopolitan', 'diversity', 'partnerships research associations', 'data processing central personnel representatives', 'strategy sustainability', 'understanding of quality family-friendly', 'promoting research reports', 'career fair organizational team']
0.42	Intercultural and interreligious	['intercultural theology', 'interreligious', 'religious', 'religious research', 'religious cultural', 'intercultural', 'theology', 'religious', 'university diversity', 'orientation week']
0.42	Global and cultural civic engagement	['intercultural responsible', 'diversity sustainability', 'engagement globalization', 'specializations intercultural', 'personal development thesis', 'gender diversity', 'transformation politics', 'internationality study advice', 'strategic diversity', 'civil society commitment']
0.41	Internationalization and Praxis	['internationality practice-integrated', 'humanities', 'dahlem humanities', 'practice-integrated memberships', 'internationalization internationality', 'engineering science department', 'scholars risk', 'scholars', 'engineering science department', 'practice-integrated']
0.41	Initiatives	['initiatives fostering', 'initiatives', 'transformation initiatives', 'faculty research', 'organizations', 'governance accreditations', 'positions research', 'endowment entrepreneurial', 'research faculty', 'sustainable transformation']

UK: topics most closely related to "diversity"

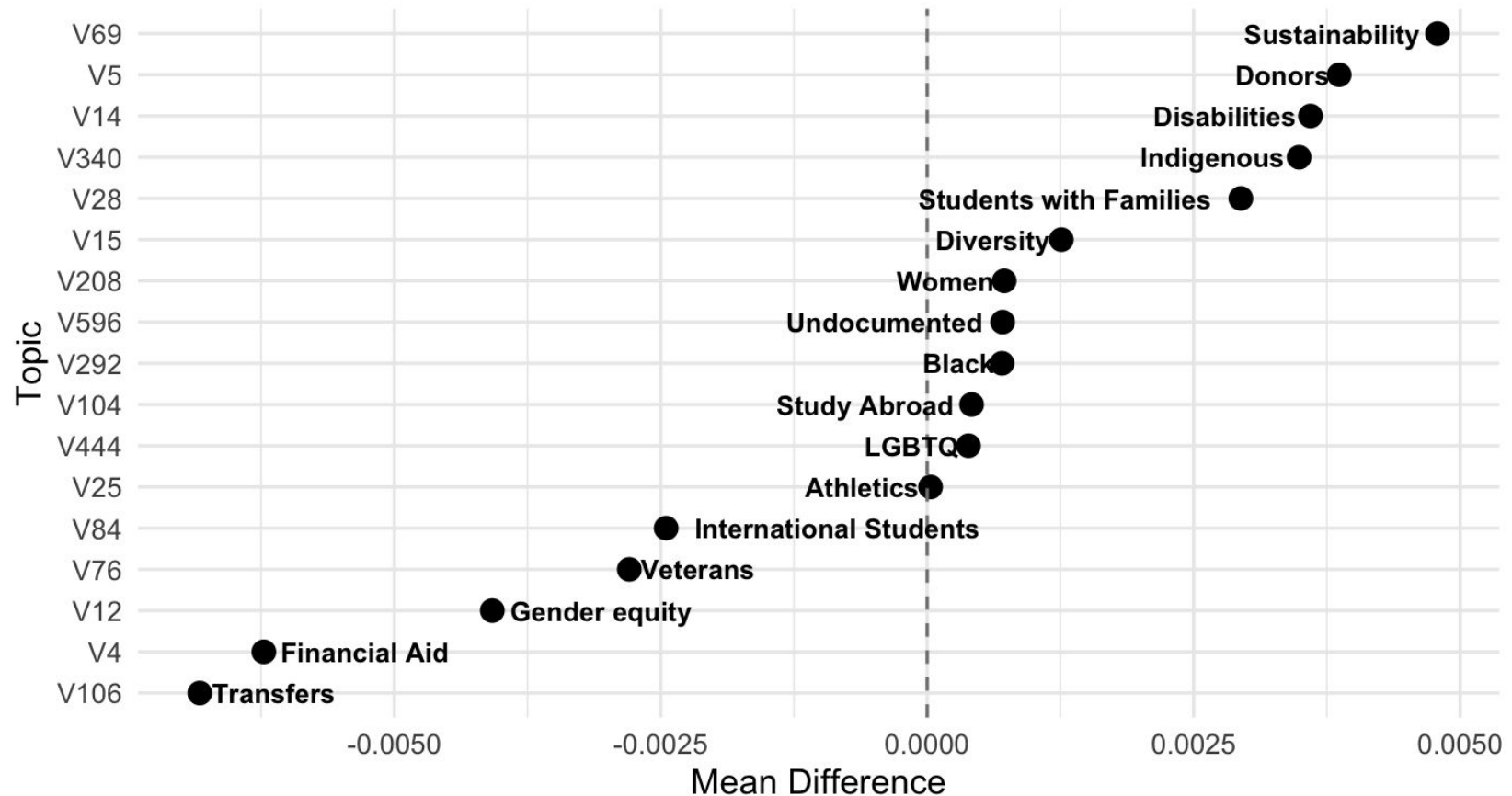| Correlation	Topic name	Associated phrases
0.53	Equality and diversity	['equality charter', 'equality diversity', 'equality charters', 'equality act', 'diversity inclusion', 'gender equality', 'workplace equality', 'race equality', 'discrimination', 'diversity']
0.51	Black celebration	['celebrating black', 'celebrate black', 'black month', 'black heritage', 'black students', 'month black', 'black community', 'equality diversity', 'black females', 'black women']
0.48	Languages and cultures	['cultures languages', 'materialisms', 'humanities early', 'linguistics', 'ancestors exhibition', 'documentary', '2021 blog', '2020 blog', '2019 blog', 'languages area']
0.47	Social justice and activism	['social justice', 'youth activism', 'society social', 'inequalities injustices', 'activism housing', 'activism', 'social policy', 'sociology', 'promote social', 'welfare social']
0.45	Sustainability and partnerships	['schools unitslocationsfacilitiesequality', 'participation faculties', 'awards sustainabilityour', 'sustainabilityour peopleengagement', 'partnerships volunteering', 'youth lifecourse', 'activism housing', 'voluntary community', 'sustainabilityour', 'injustice partnerships']
0.42	International partnerships and support	['botswana', 'students applicants', 'tourism hospitality', 'lifestyles', 'scholarships support', 'partners abroad', 'schools colleges', 'student blog', 'mentoring', 'international partners']
0.40	Gender diversity	['gender diversity', 'gender equality', 'female leadership', 'female participation', 'equality awards', 'masculinities', 'culture postdoctoral', 'diversity', 'opportunities culture', 'students overview']
0.39	Global challenges and interdisciplinary studies	['global challenges', 'interdisciplinary centres', 'interdisciplinary', 'director interdisciplinary', 'social inclusion', 'cultural policy', 'inclusion cultural', 'global culture', 'academic collaboration', 'diversity']

India: topics most closely related to "diversity" + "discrimination"

Correlation	Topic name	Associated phrases
0.54	Youth Engagement and Talent Development	['celebrated youth', 'youth', 'literacy youth', 'youth free', 'talent showcase', 'fest', 'corporates young', 'workshop', 'workshop principals', '2016 invitation']
0.48	Student Welfare Services	['opportunity cell', 'cell website', 'sundaranar', 'youth welfare', 'welfare youth', 'navigation', 'sundaranar rights', 'ugc net', 'nss cell', 'placement cell']
0.44	Academic Departments and Faculty Governance	['education department', 'student ku', 'studies department', 'faculty council', 'faculty councils', 'interdisciplinary syllabi', 'faculty biochemistry', 'coursework', 'faculty', 'career counseling']
0.33	Academic Events and Competitions	['seminar', 'workshop', 'dec workshop', 'fest inspire', 'annual fest', 'awareness', 'sessional examination', 'moot competition', 'competition', 'fest']
0.66 (discrimination)	Sexual Harassment and Discrimination Policies	['harassment policy', 'sexual harassment', 'discrimination sexual', 'harassment complaint', 'harassment women', 'complaints sexual', 'complaint sexual', 'harassment', 'policy sexual', 'discrimination']
0.64 (discrimination)	Grievance Handling Procedures	['grievance status', 'submit grievance', 'register grievance', 'grievance cell', 'grievance grievance', 'students grievance', 'grievance type', 'grievance students', 'grievance', 'grievance redressal']
0.49 (discrimination)	Academic Administration and Support Services	['details students', 'committee sgrc', 'naac committees', 'discrimination nirf', 'students admissions', 'staff distinguished', 'college prospective', 'rti staff', 'counselling students', 'internal committee']
0.41 (discrimination)	Special Support and Accommodations	['challenged students', 'facility disability', 'disabilities', 'disability', 'disability support', 'persons disabilities', 'visitor chairperson', 'examination proctor', 'teaching special', 'chairperson']

Focus back on the USA

file:///Users/lmarbach/Desktop/Diversity%20in%20HE/USA/heatmap.html

Differences in Means between Elite and Non-Elite Universities USA

For each university, the mean of the probabilities of each topic for each webpage was computed.

Next steps

- Adding more covariates (ranking, location, status, etc.)

- More countries

- Over time

Feedback

- If you had to review this proposal, what would you say?

- What other covariates should we look at?

- How to compare diversity discourse and diversity practice?

Get in touch!

lmarbach@stanford.edu