



Different Religions, Different Posts?

How do posts in the subreddits r/christianity, r/buddhism, r/islam, and r/atheism differ in language use?

By: Ami, Gian, Jenny, Mattéo and Mimi

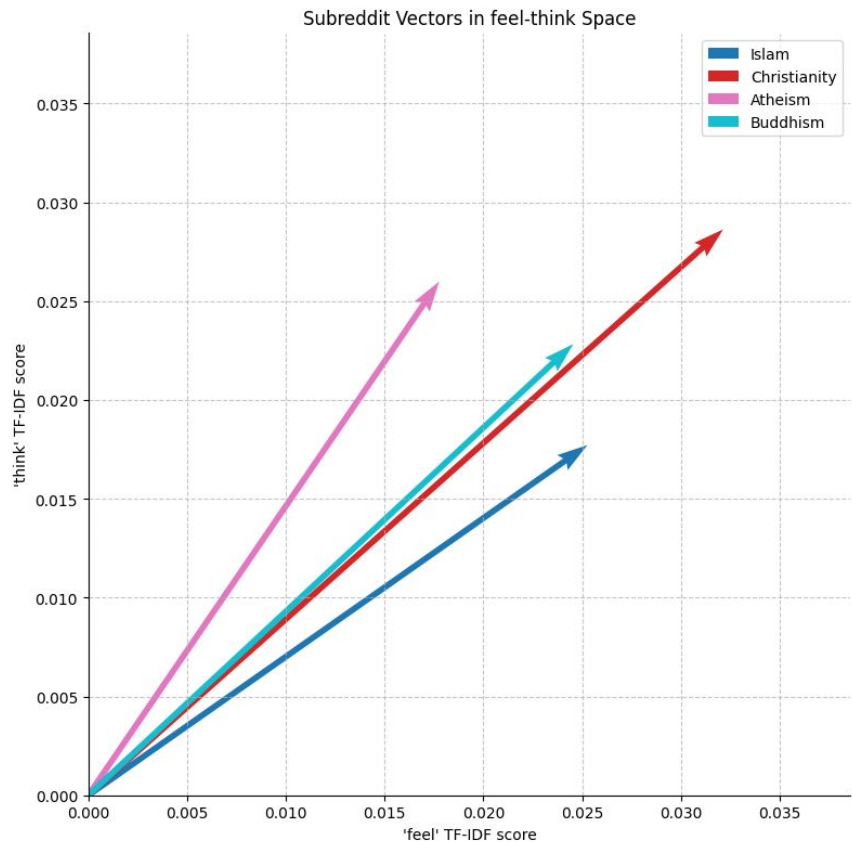
Part I: Analytical vs. Emotional Language

Theoretical background:

- Pennycook et al. (2017) find a negative correlation between religious belief and performance on analytical thinking measures in their meta analysis

H1: Posts in r/atheism use more analytical language, while religion subreddits are associated with more emotional language

Posts in the r/atheism subreddit are closest to “think”, whereas posts in the r/islam subreddit are closest to “feel”



Angles between vectors			
Christianity	1.2°		
Islam	7.9°	6.7°	
Atheism	12.7°	14°	20.6°
	Buddhism	Christianity	Islam

Applying a pre-defined lexicon for scoring “analytical” and “emotional” sentiments across subreddits (1/2)

Lexicon creation

Analytical words

explain, rational, although, despite, knowledge,
evidence, proof, reason,...

Emotional words

sad, love, fear, regret, guilty, hate, emotional, angry,
terrible, concern,...

Example sentences

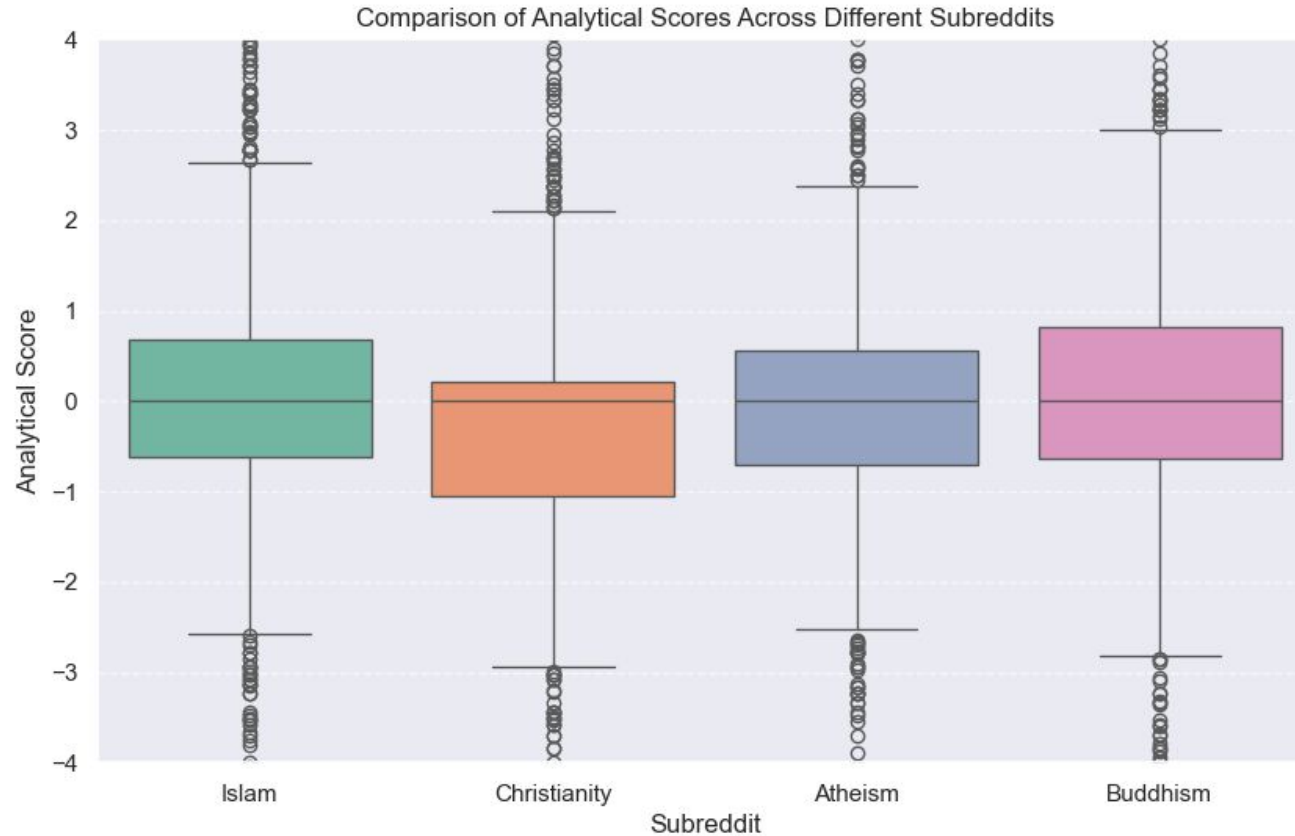
Sentence with high analytical score

“1. Who wrote the Quran? (Provide evidence for the Quran) 2. Why do you think islam is the truth?”

Sentence with high emotional score

“I could be in the worst of the worst situations and I read maghrib then POOF its all gone, people who genuinely follow the religion are the best people.”

Applying a pre-defined lexicon for scoring “analytical” and “emotional” sentiments across subreddits (2/2)



Part II:

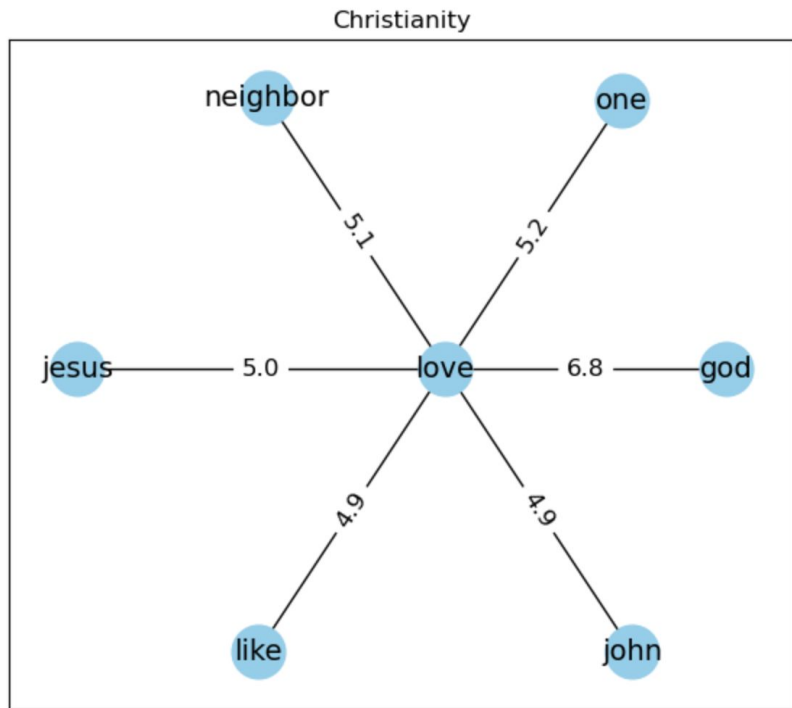
Word Associations

Theoretical background:

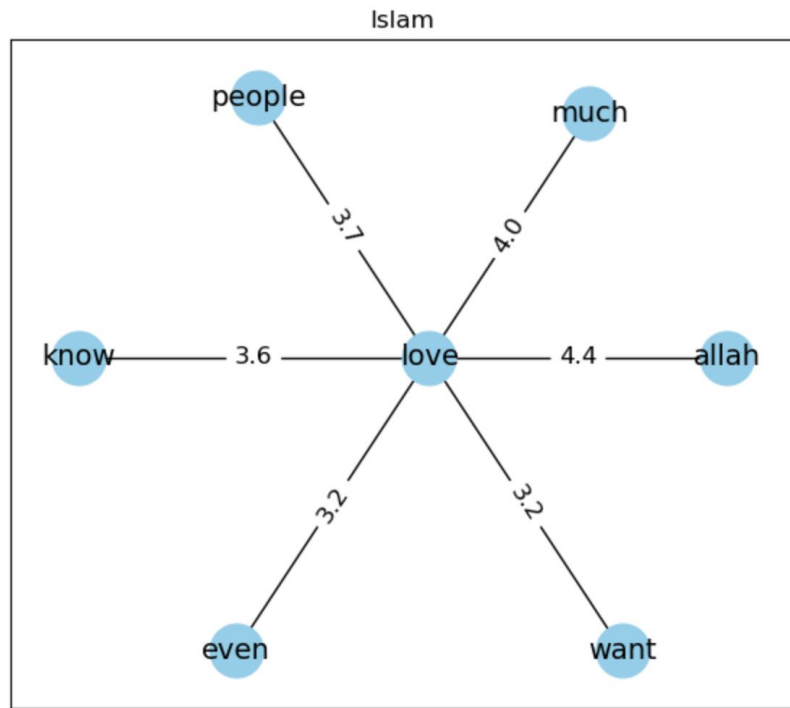
- Schmidt et al. (2020) find that different subreddits talk differently about key religious concepts

H2: Collocations surrounding the words “love”, “death” and “god” within subreddits vary across subreddits

“Love” - Word Association using Pointwise Mutual Information (PMI)

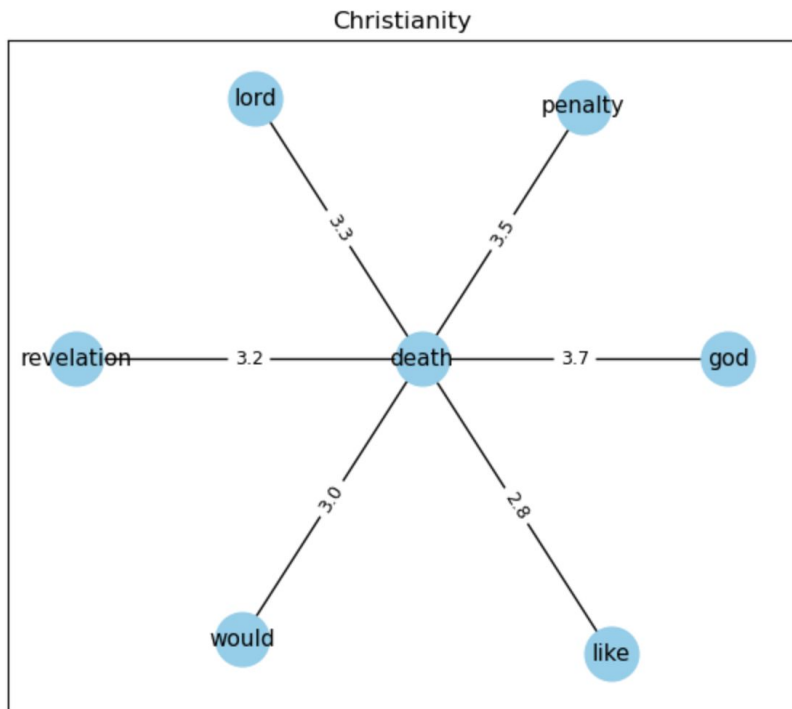


- Very strong PMI between “love” & “god”
- “Love” in association with “neighbor”
- Biblical references (“john”)

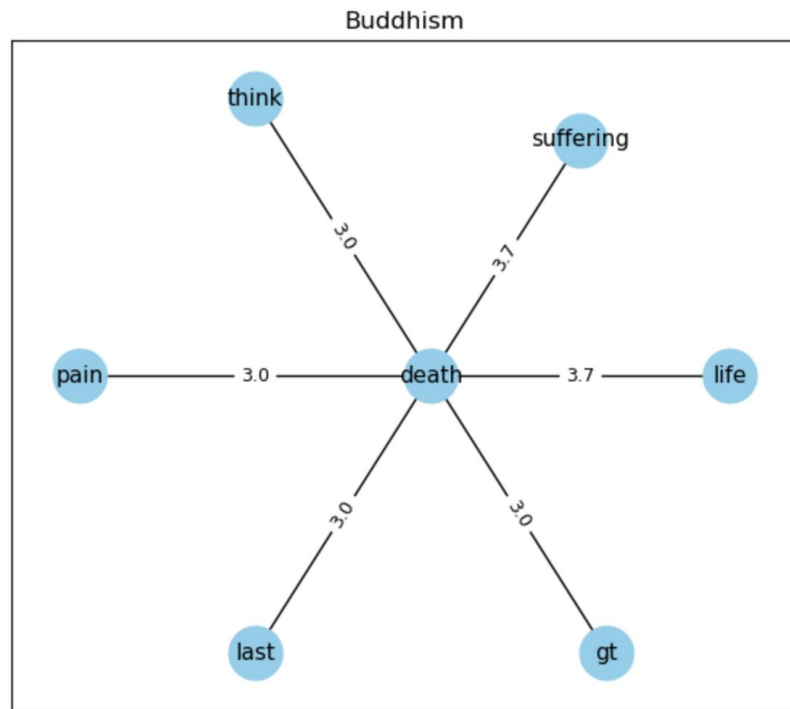


- Strong PMI between “love” & “Allah”
- “Love” in association with “people”

“Death” - Word Association using PMI

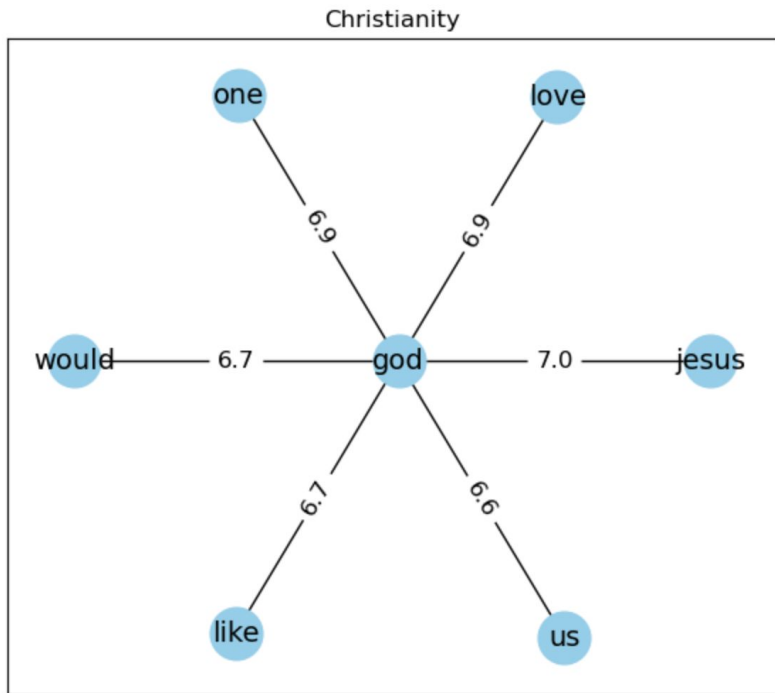


- Strong PMI with “penalty”
- Strong PMI with “god” and “lord”

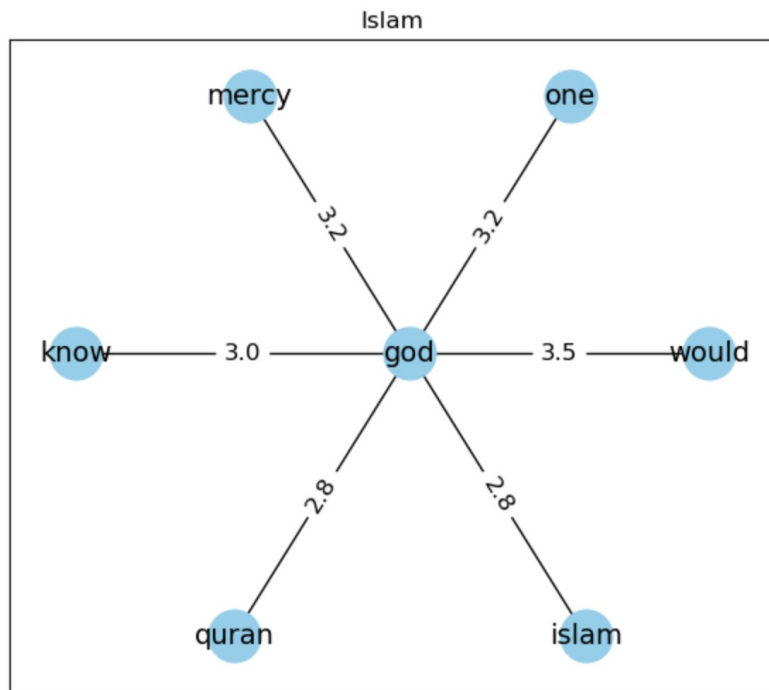


- Strong PMI with “suffering”, “pain”, and “life”
- Association with “life” could reflect reincarnation

"God" - Word Association using PMI



- "One" god corresponds to Christian narratives
- Strong PMI with "Jesus" and commonly associated with "love"



- "One" god corresponds to Islam narratives
- Association with religious text "Quran"

Part III - Machine Learning

Classification of Posts

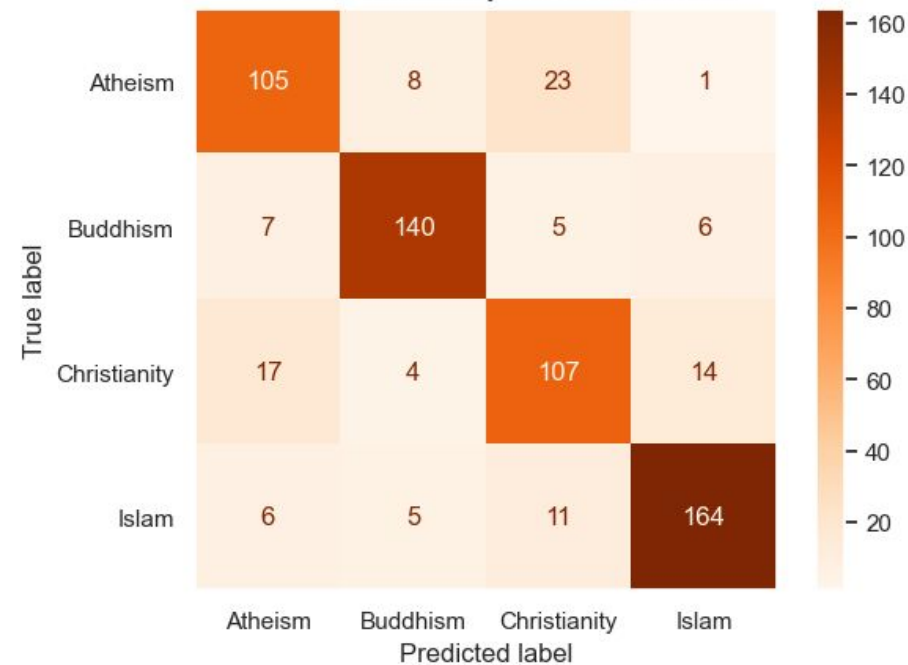
Theoretical background:

- Patel & Meehan (2021) found that Multinomial Naive Bayes achieved best accuracy rate as compared with other models for fake news detection on Reddit posts

H3: Language use across subreddits is distinct enough to allow accurate classification using supervised Machine Learning

Classification of subreddits with Multinomial Naive Bayes

Multinomial Naive Bayes Confusion Matrix



MNB Classification Report

Subreddit	Precision	Recall	F1-score
Atheism	0.78	0.75	0.76
Buddhism	0.84	0.89	0.86
Christianity	0.70	0.71	0.71
Islam	0.82	0.80	0.80

Overall model accuracy: 0.78

Top ten words for each class sorted by log probability

Results	Atheism	Buddhism	Christianity	Islam
1	god	buddhism	god	allah
2	religious	buddha	jesus	islam
3	religion	buddhist	christian	quran
4	trump	like	just	muslim
5	people	life	like	know
6	just	practice	don	like
7	christian	know	bible	dua
8	atheist	meditation	know	haram
9	church	just	people	just
10	like	mind	love	help

Limited text processing and quantification constitute areas of improvement

Research design	<ul style="list-style-type: none">• Lack of quantification of differences between subreddits possible for word associations
Sample selection	<ul style="list-style-type: none">• Limited external validity of Reddit data usage
Data	<ul style="list-style-type: none">• Omission of posts that are not in English and problems with spelling mistakes• Limited text processing (e.g., lemmatisation, stemming)
Operationalisation	<ul style="list-style-type: none">• Creation of lexicon with limited words• Operationalisation of rational vs. emotional through “feel” and “think”• We infer a connection between words through PMI without considering context

We have gathered partial support for our hypotheses, however, there is potential for future research

H1: Posts in r/atheism use more analytical language, while religion subreddits are associated with more emotional language

Limited support

H2: Collocations surrounding the words “love”, “death” and “god” within subreddits vary across subreddits

Supported

H3: Language use across subreddits is distinct enough to allow accurate classification using supervised Machine Learning

Supported

Potential for further research

- Seasonality analysis using posts throughout whole year (re holidays)
- Incorporating comments for analysis

Bibliography

- Patel, A. and Meehan, K. (2021). Fake News Detection on Reddit Utilising CountVectorizer and Term Frequency-Inverse Document Frequency with Logistic Regression, MultinomialNB and Support Vector Machine. 32nd Irish Signals and Systems Conference (ISSC), Athlone, Ireland, 2021, pp. 1-6, doi: 10.1109/ISSC52156.2021.9467842
- Pennycook, G. & Ross, R.M. & Koehler, D.J. & Fugelsang, J. A. (2017). Correction: Atheists and Agnostics Are More Reflective than Religious Believers: Four Empirical Studies and a Meta-Analysis. PLOS ONE 12(4): e0176586. <https://doi.org/10.1371/journal.pone.0176586>
- Schmidt, T., Kaindl, F., & Wolff, C. (2020). Distant reading of religious online communities: A case study for three religious forums on reddit.