

Computational Analysis of Biblical Source Criticism

AN INTERDISCIPLINARY PROJECT IN CS + ANCIENT TEXTS

ARUNIMA TRIPATHY

Rutgers University

Monday, March 31, 2025

Background

1

Authorship theory for the first five books of the Bible. They are also referred to as the Pentateuch, or the Torah in the Jewish tradition.

2

Documentary Hypothesis:

- J - Yawhist, stories and personal interactions
- E - Elohist, moral and ethical themes
- D - Deuteronomy, emphasizes the law
- P - Priestly, rituals, rules, and religious structure

3

Competing scholarly views

Goals

My own experiment...

Can we explore thematic and stylistic structures in the Pentateuch to assign authorship?

Approach:

No labeling, no assumptions

Observe patterns as they emerge

Connect code to content, if possible

Initial Considerations

1

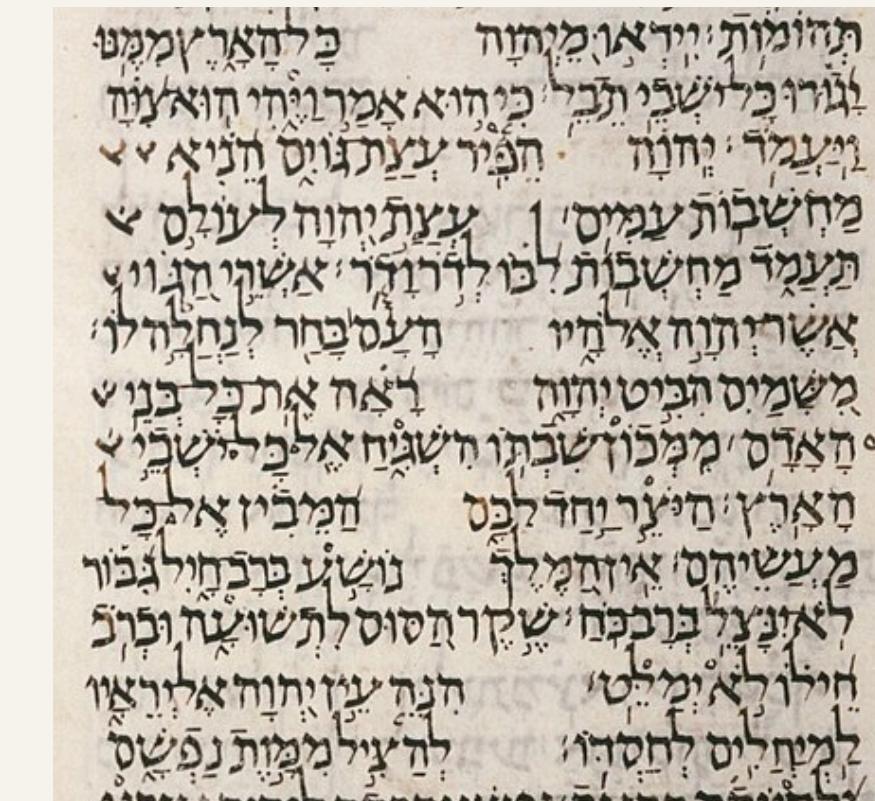
Which version to use?

2

Ancient Hebrew vs.
Modern Hebrew

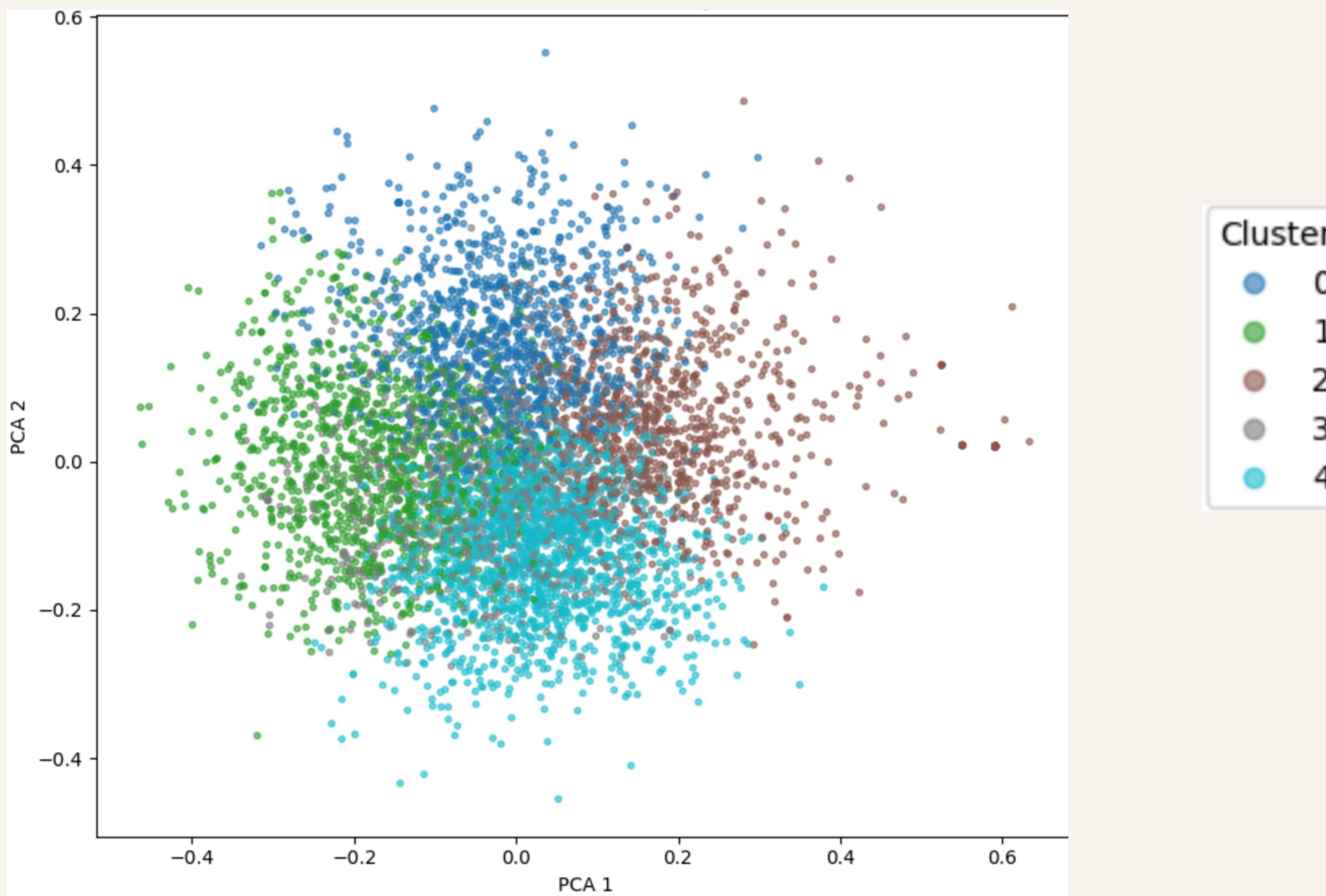
3

Custom pre-
processing pipeline



Verses from Leningrad Codex

2D PCA of Verses by Clusters



Clusters Across Books



Representative Verses

From Cluster 0:

Leviticus 10:10

“You must distinguish between what is sacred and what is common, between what is ceremonially unclean and what is clean.”

Exodus 28:41

“After you put these clothes on your brother Aaron and his sons, anoint and ordain them. Consecrate them so they may serve me as priests.”

P - Priestly Source

- “Holy”
- “Priests”
- “Clean”
- “Unclean”

Challenges

Documentation and
reproducability

Interpreting results: signal or
noise?

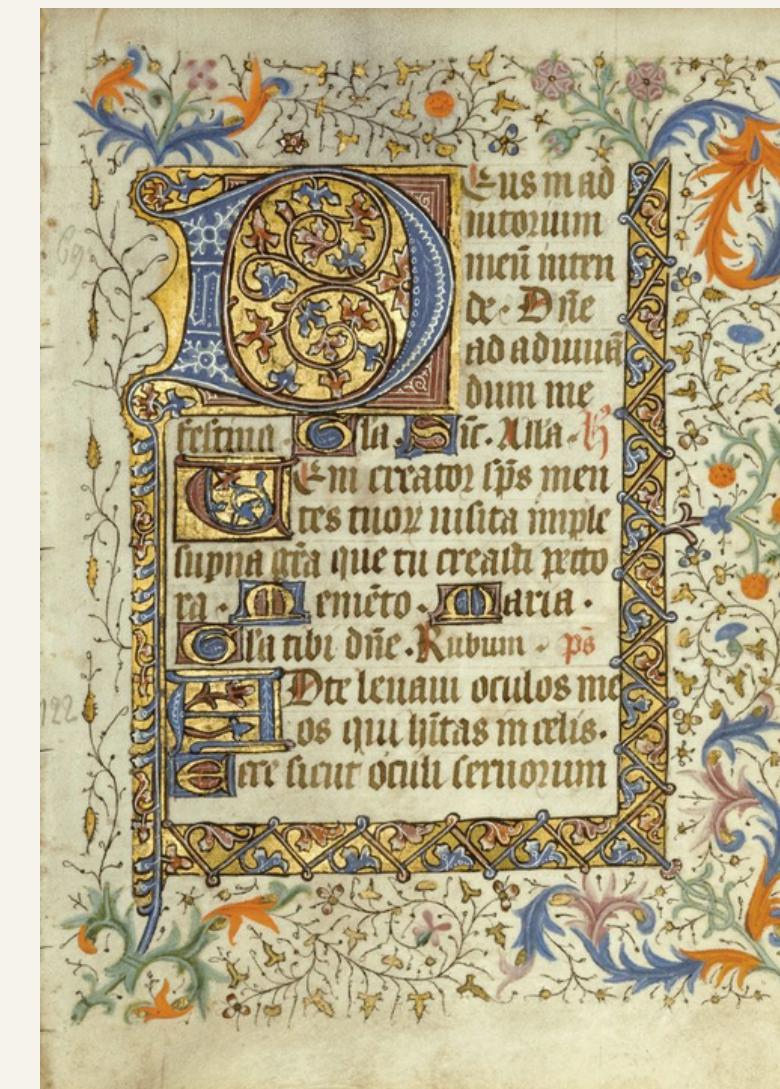
Relevance to Scholarship

Aligning results to current divisions

- Cluster 0 → Priestly
- Cluster 4 → Narrative

Reflecting on strict divisions

- Sections are fluid



Illuminated Manuscript

Future Direction

