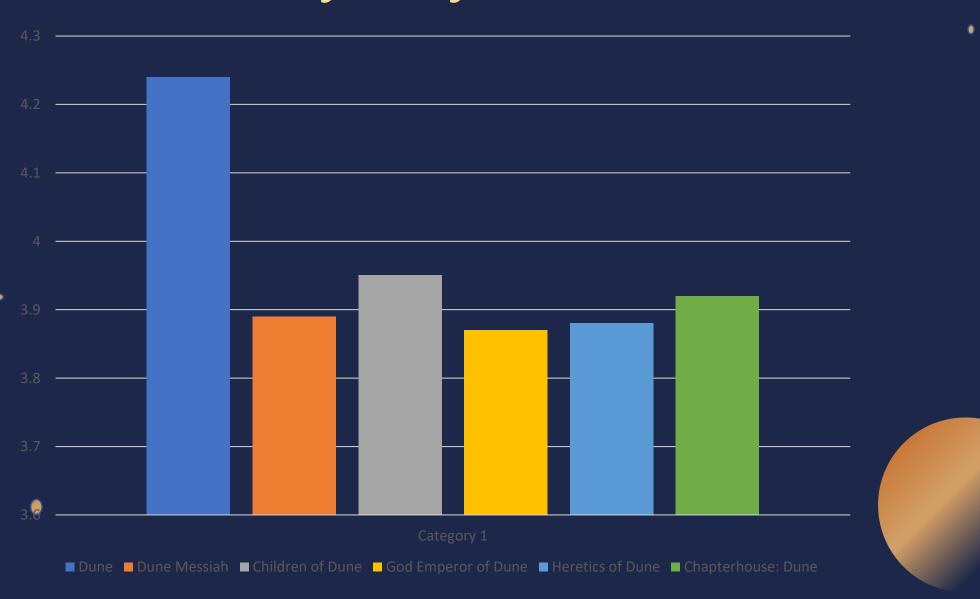# introduction

Dune: the
foundation of
modern sci-fi

ratings on goodreads

question

Can we model thematic exploration across books?

Can topic trends over books explain book ratings?

data and pre-processing

1)
Dune e-book PDF's

3)
Removal of pre-chapter quotes/general cleaning

2)
Splitting of chapters into docs
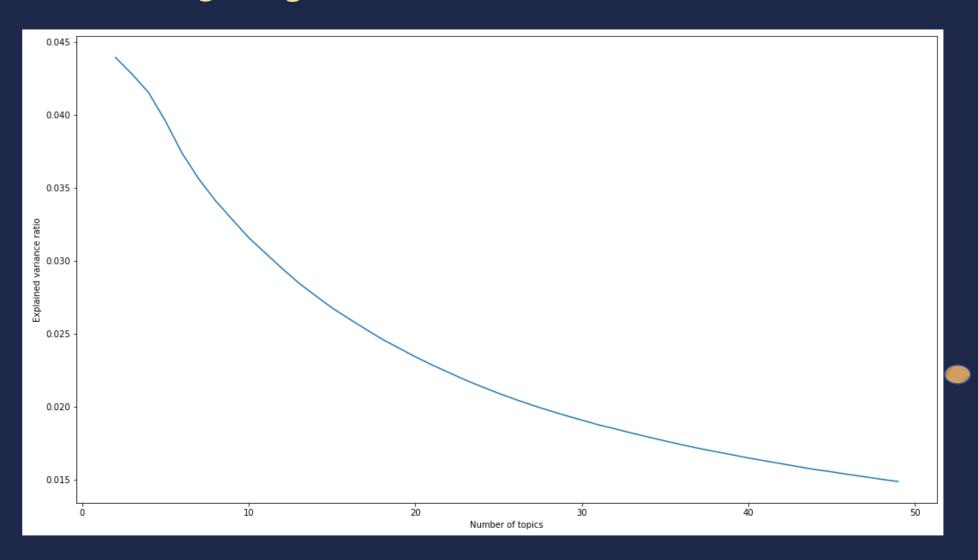
4)
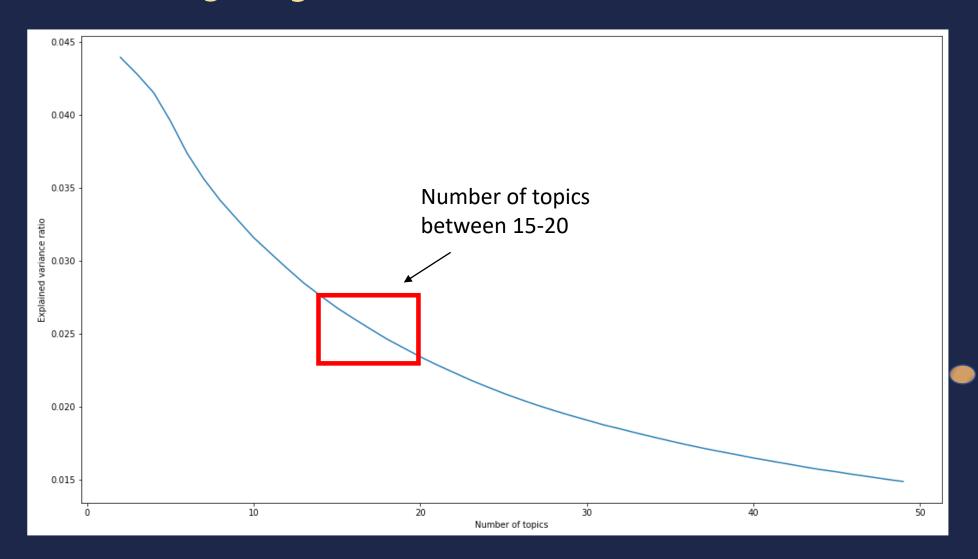Stop-word removal/lemmatization

model evaluation

Qualitative

Intuition eye-test

Quantitative

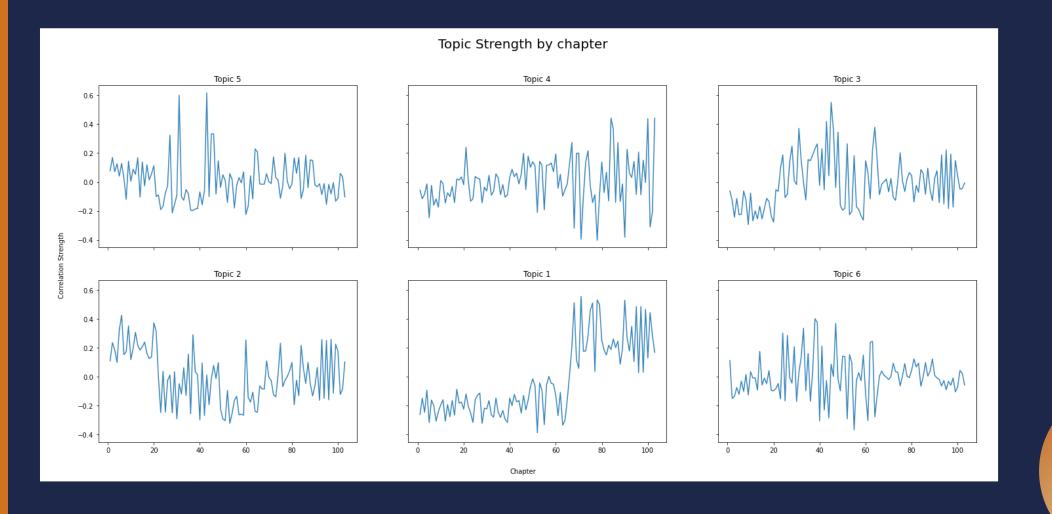Explained Variance Ratio

tuning hyper-parameters

tuning hyper-parameters

Number of topics between 15-20
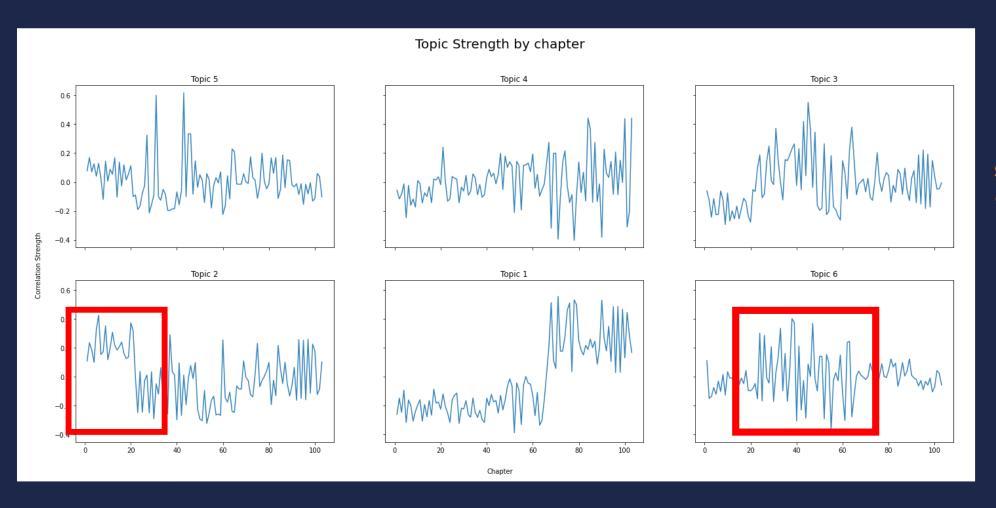
– Settle on number of topics = 20

– Average explained variance = 2.3%

our results

Topic Strength by chapter

top 6 topics by prevalence

1. God Emperor Leto II
2. Paul and Life in the Desert
3. House Corrino on Salusa Secundus
4. Emperor Leto II's Court
5. Farad and the Preacher
6. Paul as the Preacher

top 6 topics by prevalence

1. God Emperor Leto II
2. Paul and Life in the Desert
3. House Corrino on Salusa Secundus
4. Emperor Leto II's Court
5. Farad and the Preacher
6. Paul as the Preacher

a topic in action- paul and life in the desert

said, paul, jessica, stilgar, baron, duke

"PAUL STOOD outside the stilltent in the late afternoon. The crevasse where he had pitched their camp lay in deep shadow. He stared out across the open sand at the distant cliff, wondering if he should waken his mother, who lay asleep in the tent.... Within the tent, Jessica awakened, turned onto her back and peered sidelong
out the transparent end at Paul. He stood with his back to her and something about his stance reminded her of his father. She sensed the well of grief rising within her and turned away.
Presently she adjusted her stillsuit, refreshed herself with water from the tent's catchpocket, and slipped out to stand and stretch the sleep from her muscles."

- Dune, Frank Herbert

insights

- Topics that explain most variance are book-plot tied

- Books that are most plot-driven and related to Paul and Jessica are highest rated

future work

- Dialogue?
  - Topic modeling by punctuation use
- Focus more on order of words

THANK YOU

appendix

Initial modeling:

```
Topic: ' Paul and Jessica in desert '
said, paul, jessica, stilgar, baron, duke, thought, hawat, chani, gurney, mother, ll, man, fremen, ve, kynes, yueh, water, san
d, saw

Topic: ' Harkonnens and imperial politics '
baron, feyd, piter, hawat, rautha, nefud, lord, emperor, duke, yueh, rabban, sardaukar, uncle, leto, nephew, mentat, count, ah,
fenring, said

Topic: ' House Atreides on Dune '
duke, hawat, yueh, kynes, halleck, table, leto, sire, room, father, idaho, tuek, mapes, thufir, harkonnens, fremen, ll, men, do
or, said

Topic: ' Atreides and the Fremen '
gurney, paul, stilgar, halleck, hawat, fremen, kynes, men, sardaukar, duke, sire, lord, feyd, worm, rautha, man, fedaykin, equi
pment, storm, sand

Topic: ' Palace intrigue '
yueh, paul, stilgar, tooth, book, jamis, gurney, leto, ring, wanna, sardaukar, house, ll, tattoo, read, hall, filmbook, seeker,
room, diamond

Topic: ' Desert ecosystem '
kynes, sand, yueh, paul, tent, pack, lump, desert, piter, worm, baron, father, rock, fissure, thopter, surface, worms, storm, h
ole, lecturing

Topic: ' Paul '
paul, feyd, gurney, rautha, mother, old, reverend, emperor, tent, woman, seeker, haderach, kwisatz, box, pain, bene, gesserit,
count, choam, slave

Topic: ' Desert adventure '
hawat, sand, lump, tent, pack, rock, cliff, fremen, mapes, seeker, fissure, basin, night, stars, thumper, room, dunes, sound, f
oam, filter

Topic: ' Sietch Tabr '
alia, harah, baron, sardaukar, emperor, tent, sand, mother, reverend, rabban, tharthar, child, fremen, chani, piter, hangings,
sietch, old, son, woman

Topic: ' Jessica and the Harkonnens '
piter, jessica, gurney, czigo, jamis, baron, duke, scarface, paul, kinet, leto, litter, lord, stilgar, pilot, banker, woman, vo
ice, knife, mentat
```