Natural Language Processing

Final Week

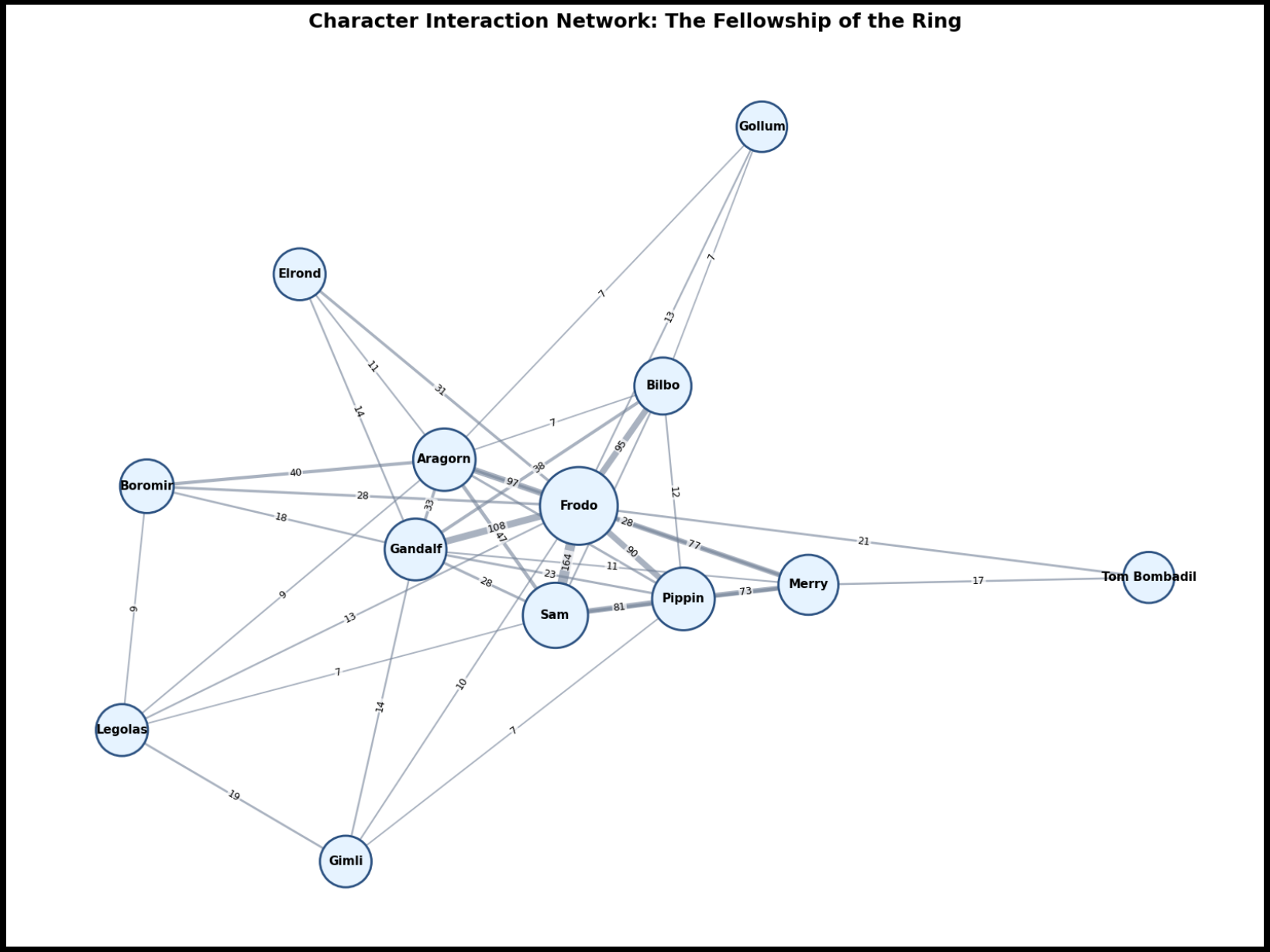
J. R. R. Tolkien, Final Project

Adilya Korkmaz, Vladyslav Stupak, Daniil Zykov

Why Analyse Tolkien?

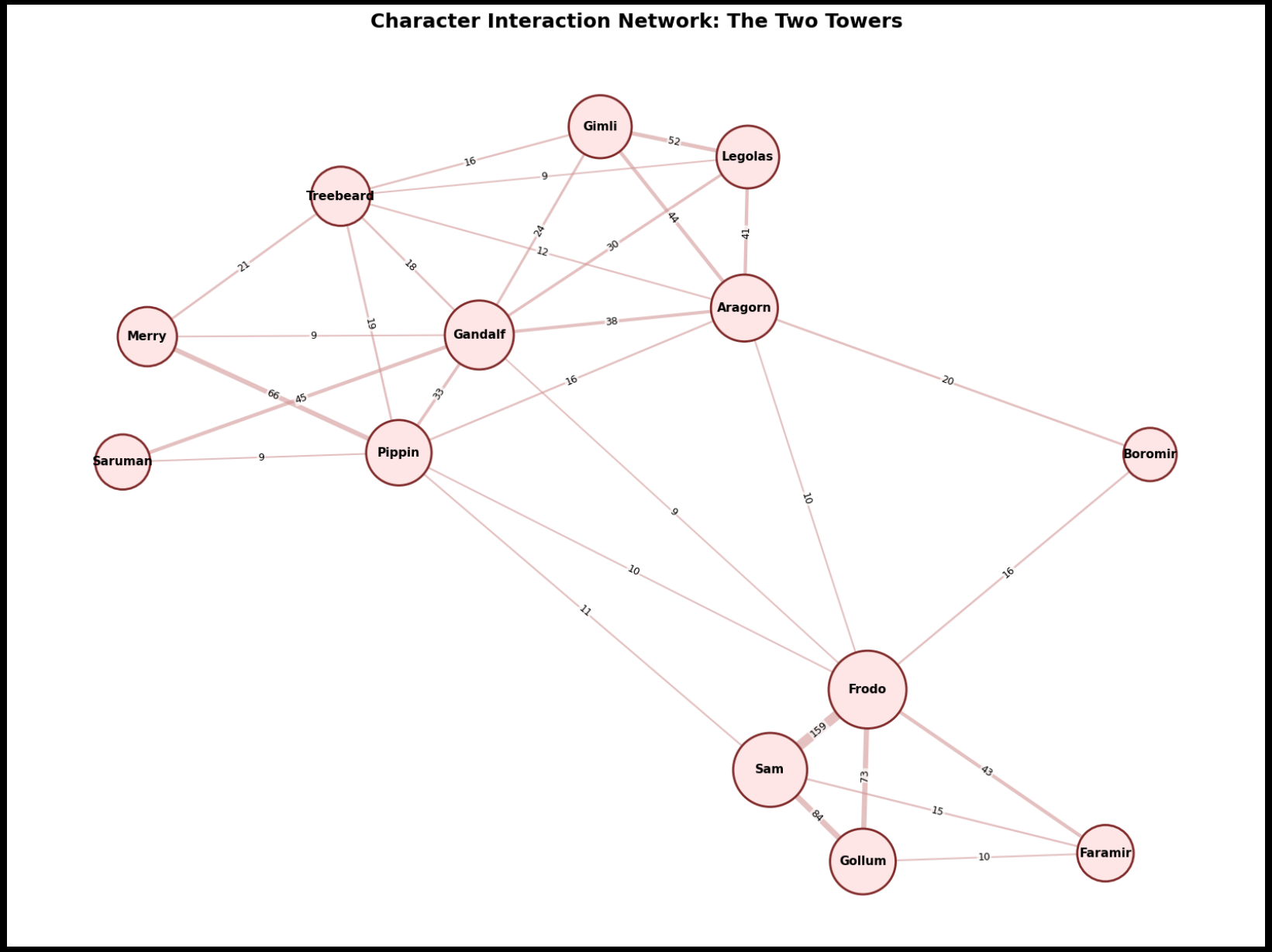
- **Native English writing**
  We wanted to make sure nothing was missed or misinterpreted in translation
- **Generally known**
  A popular series with several media adaptations that most people at least heard about
- **Author‘s linguistical background**
  J.R.R. Tolkien was a linguist before he was a writer

Character Interactions
The Fellowship of the Ring



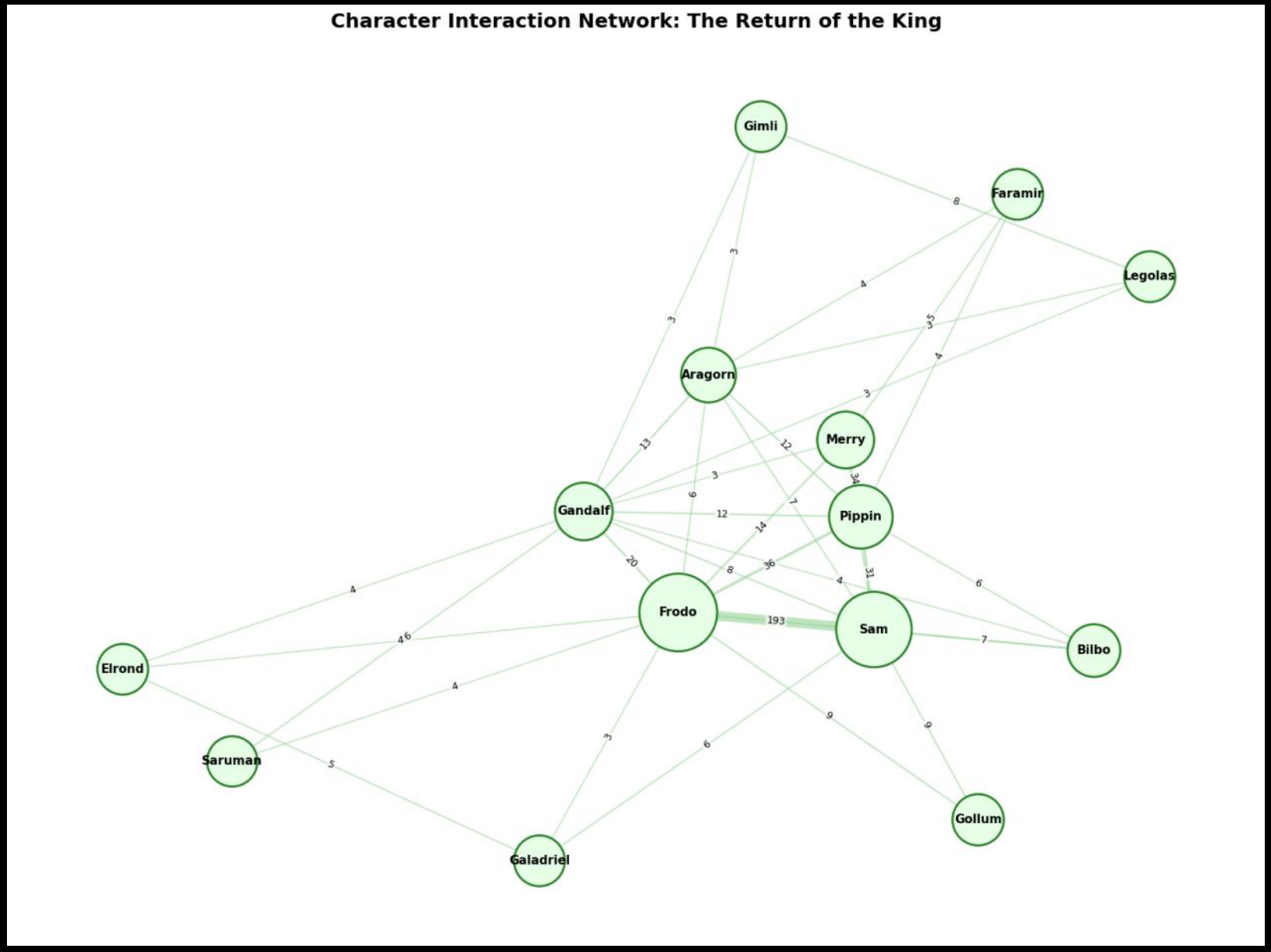
- Characters that appear within 15 words of each other are considered to be "interacting" - fundamental assumption
- Bigger nodes = more important/central characters Thicker edges = stronger relationships Spatial clustering = characters who share similar social circles
- Grouping - runs a **simulation** where Every node acts like a magnet repelling all other nodes Every edge acts like a spring pulling its two nodes together The simulation runs 100 times, each time moving nodes a tiny bit

Character Interactions
The Two Towers



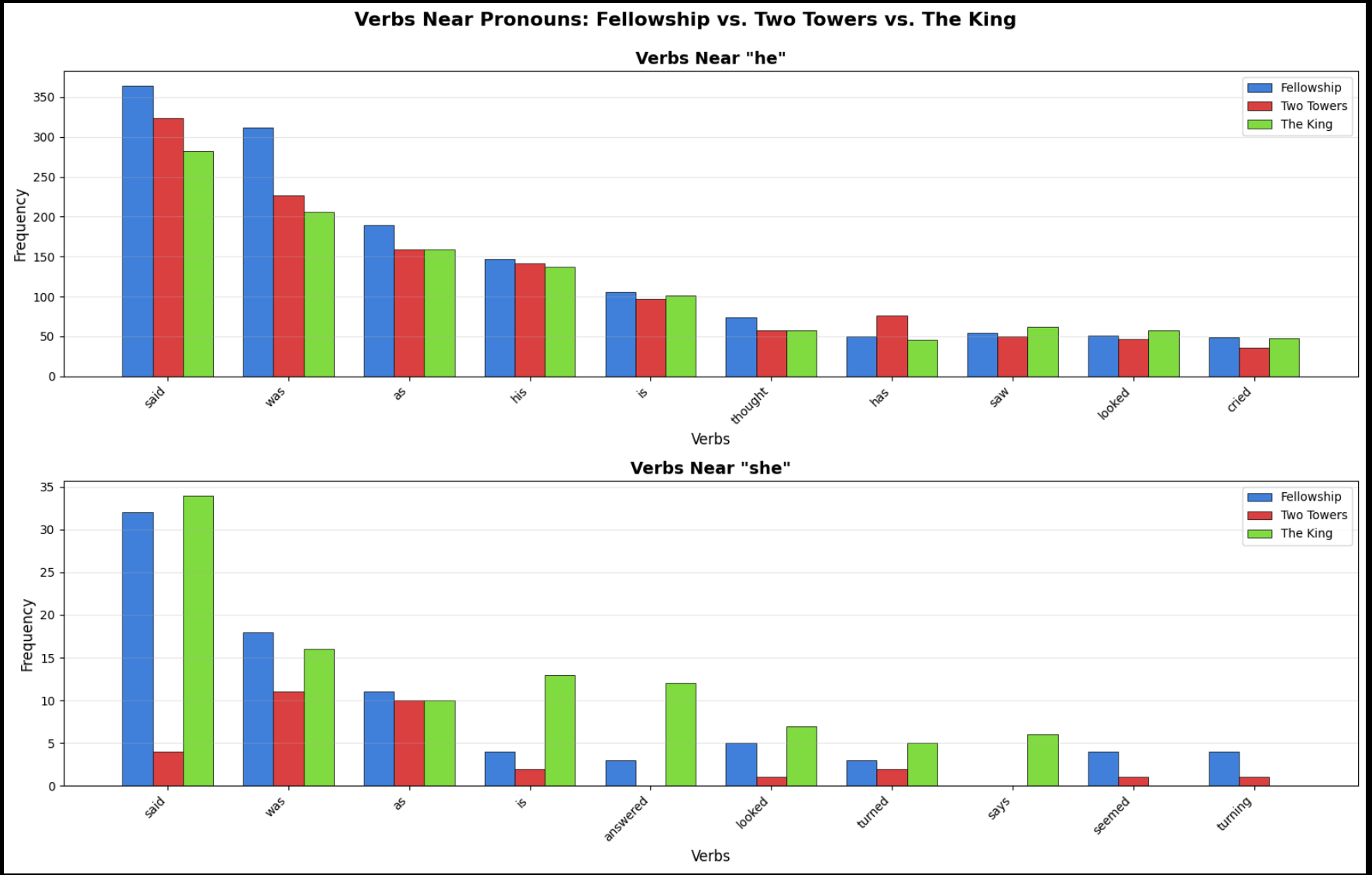
- Characters that appear within 15 words of each other are considered to be "interacting" - fundamental assumption
- Bigger nodes = more important/central characters Thicker edges = stronger relationships Spatial clustering = characters who share similar social circles
- Grouping - runs a **simulation** where Every node acts like a magnet repelling all other nodes Every edge acts like a spring pulling its two nodes together The simulation runs 100 times, each time moving nodes a tiny bit

Character Interactions
The Return of the King



- Characters that appear within 15 words of each other are considered to be "interacting" - fundamental assumption
- Bigger nodes = more important/central characters Thicker edges = stronger relationships Spatial clustering = characters who share similar social circles
- Grouping - runs a **simulation** where Every node acts like a magnet repelling all other nodes Every edge acts like a spring pulling its two nodes together The simulation runs 100 times, each time moving nodes a tiny bit

Use of Gendered Pronouns Whole Trilogy Compared

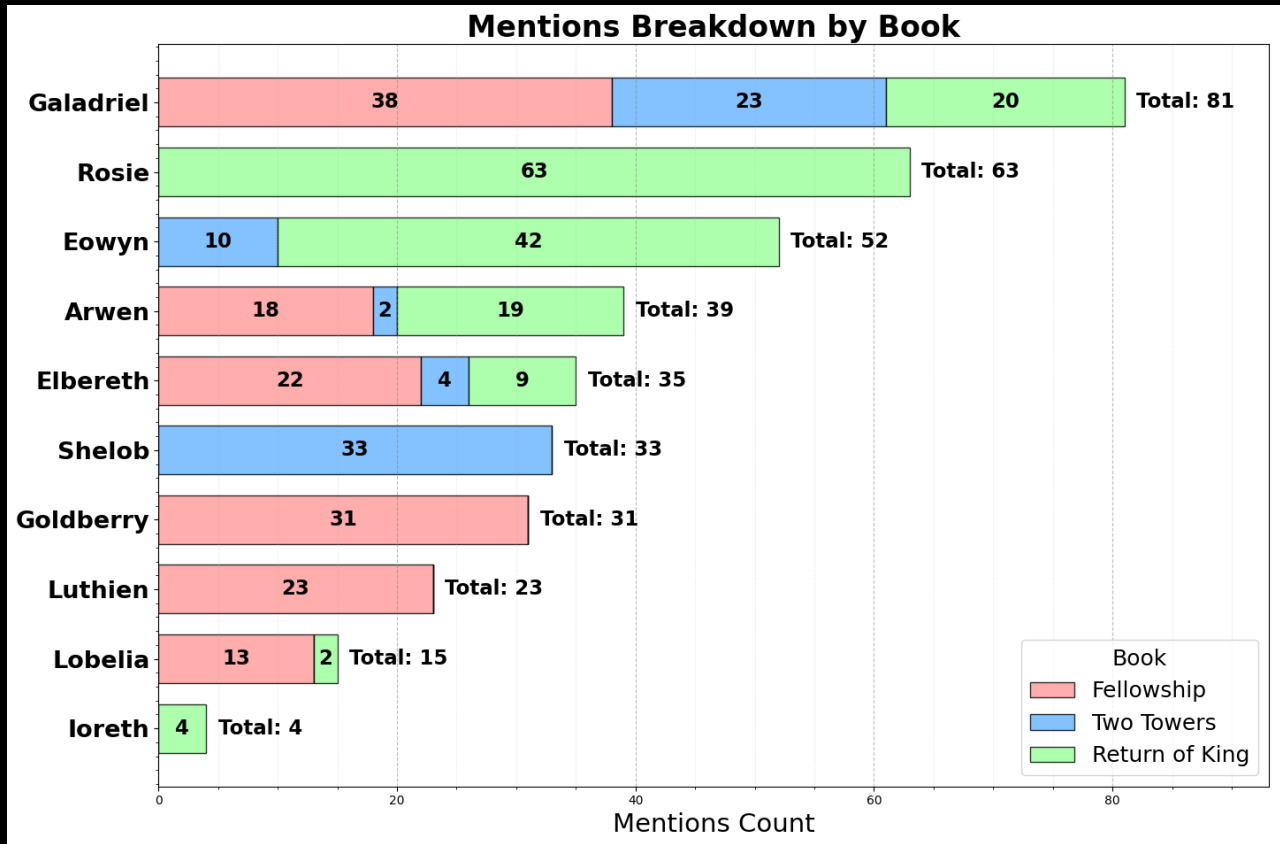


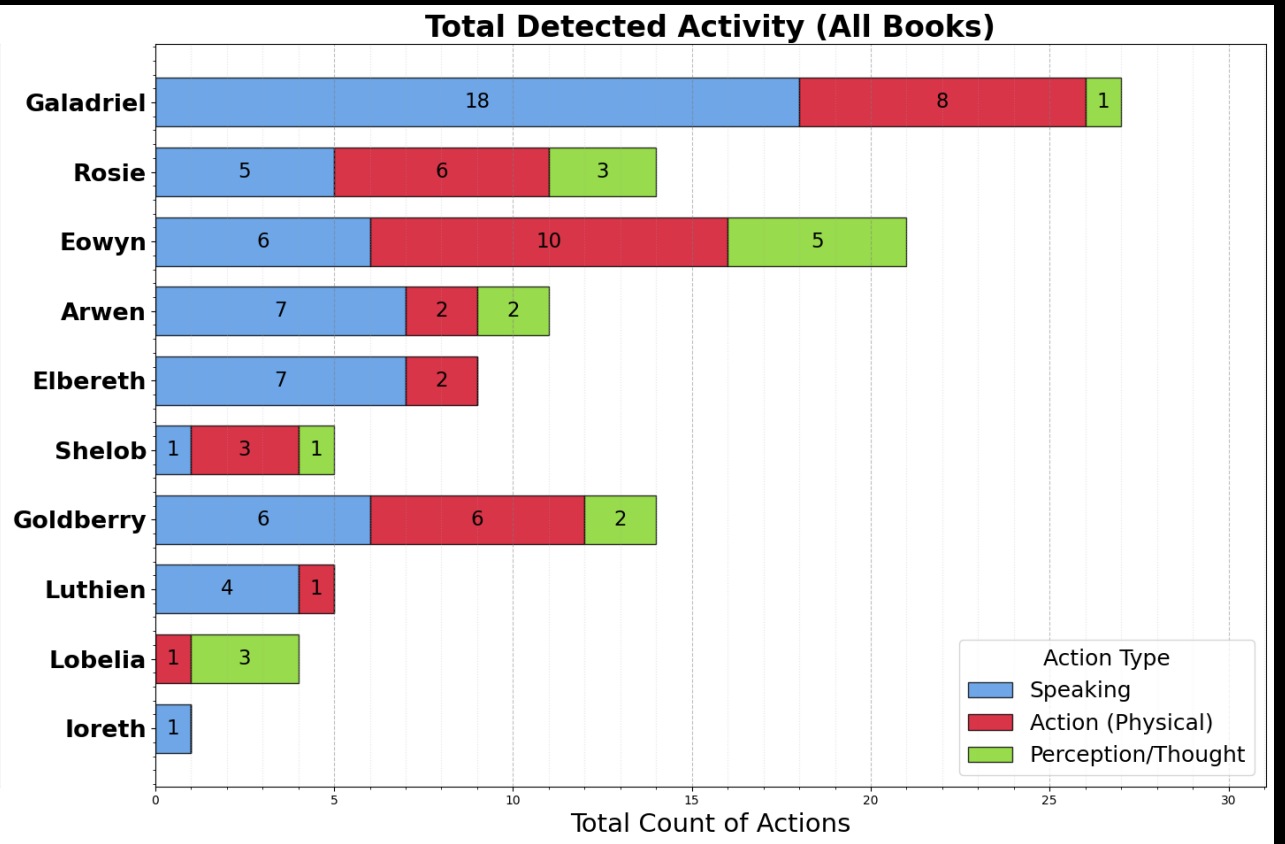
- The algorithm seeks gendered pronouns and looks at the next word in the sentence, counting every instance

- It is worth noting that the trilogy does not pass the Bechdel test with the second book of the trilogy having only three female characters in total one being Shelob

Mentions & Actions
Whole trilogy compared

- The first method is a tally counter seeking characters‘ names in the trilogy. The histogramm is divided by book for clarity
- The second method is more complicated, judging the characters‘ appearances based on context

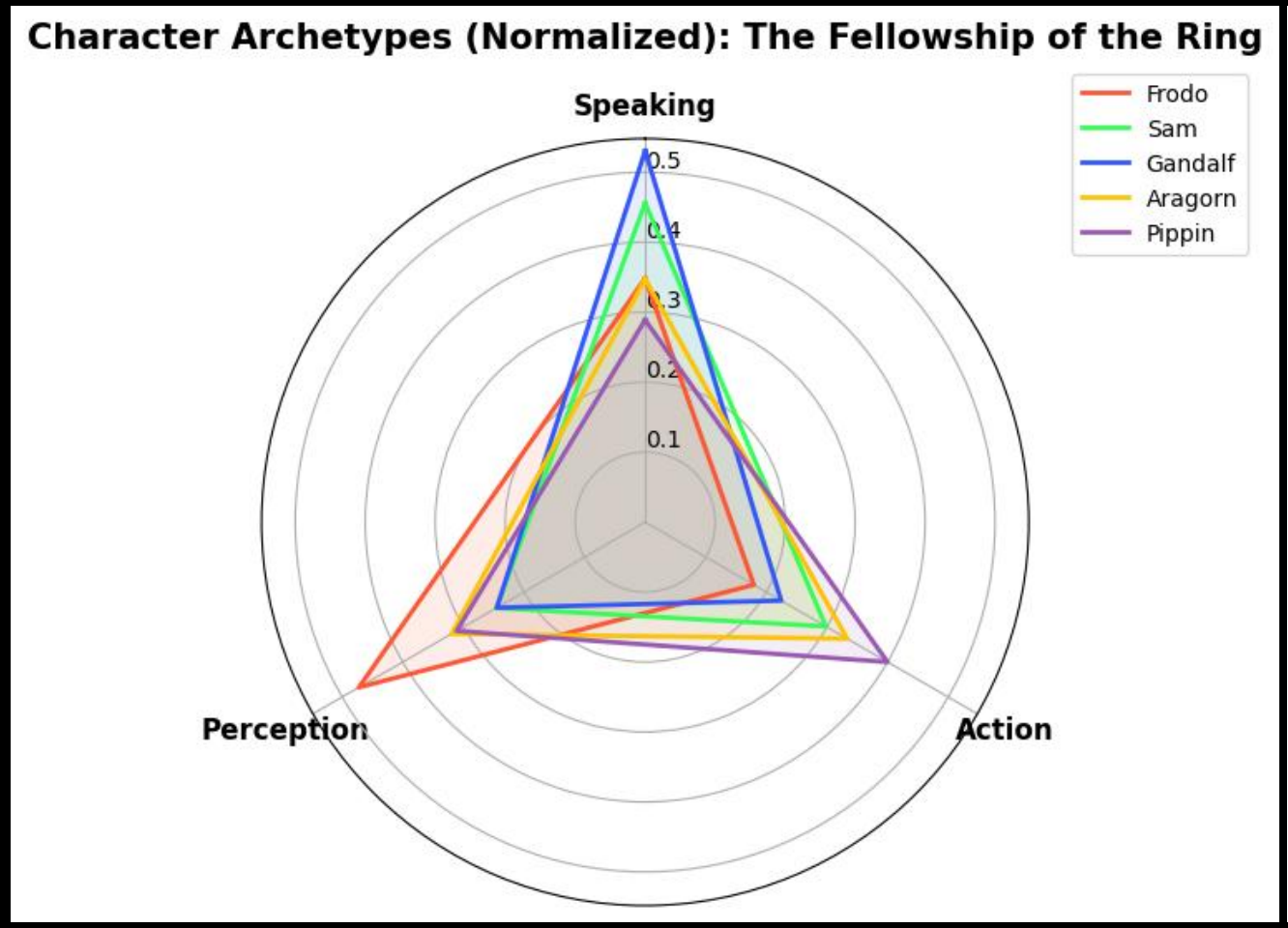


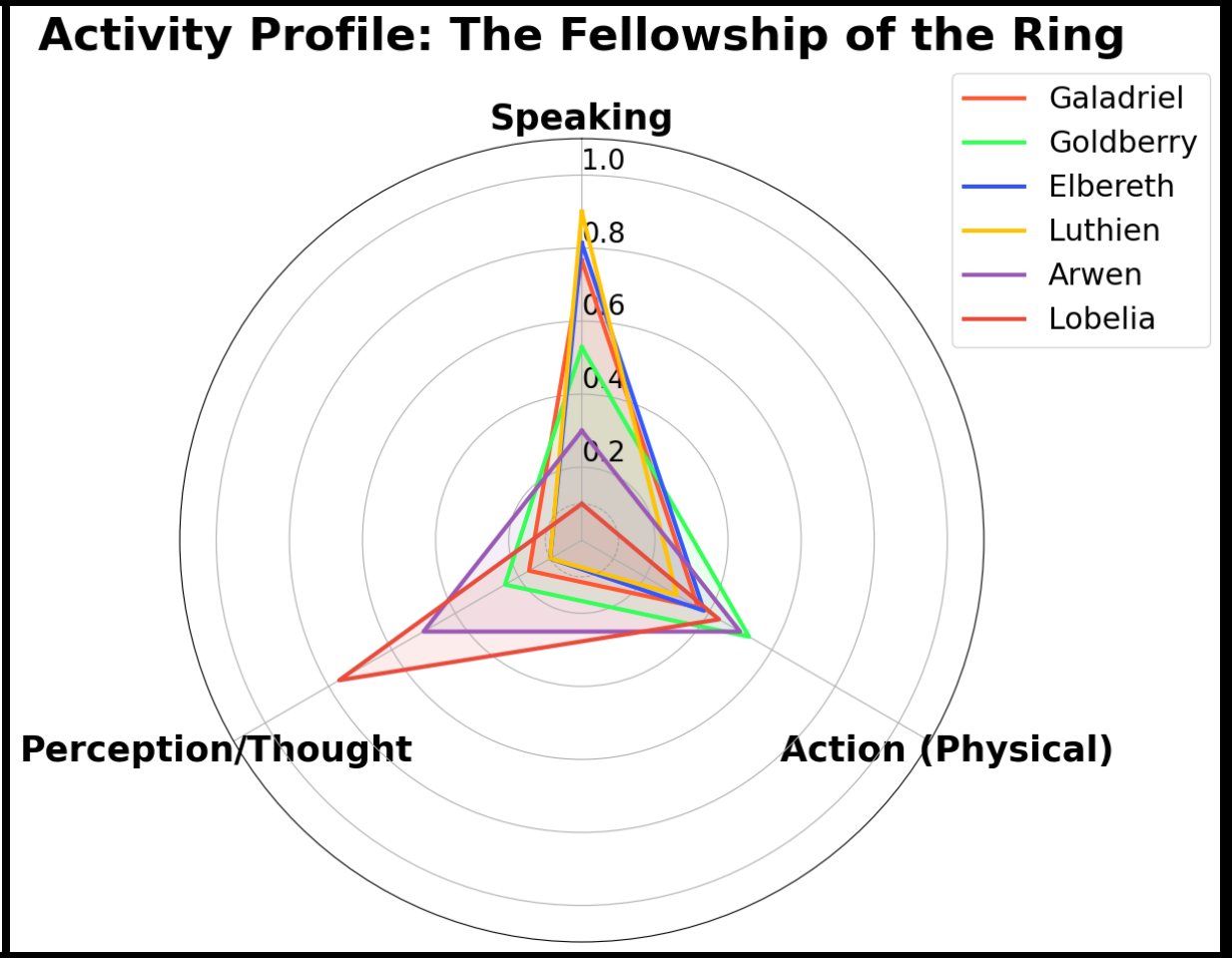


APS Ratios Compared
The Fellowship of the Ring

Data Collection

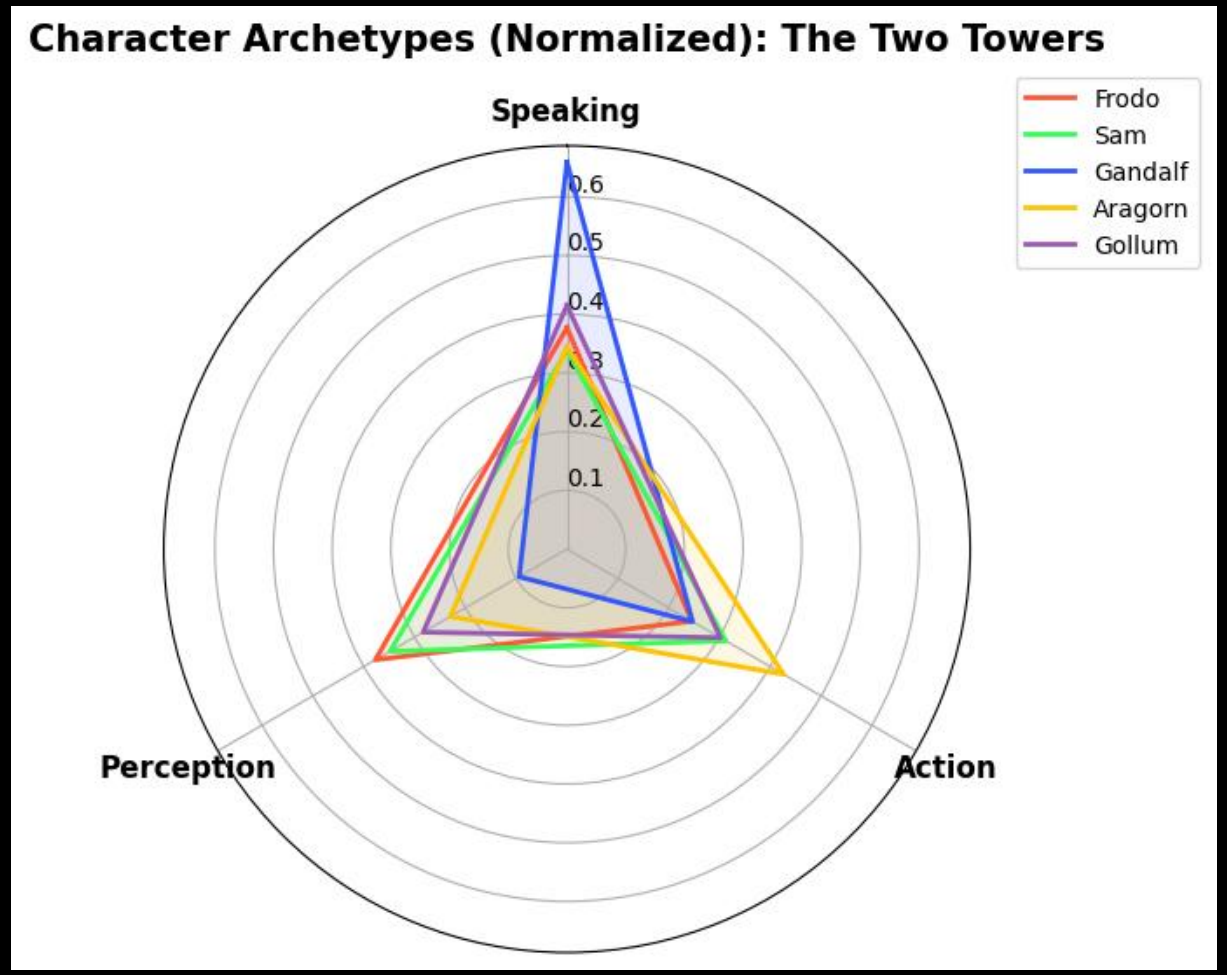
The code analyzes character activity by scanning through the book text and identifying when character names appear followed by action verbs within a 6-word window, The following 6 words are examined to find verbs in three categories: **Speaking**, **Action (Physical)**, **Perception/Thought**

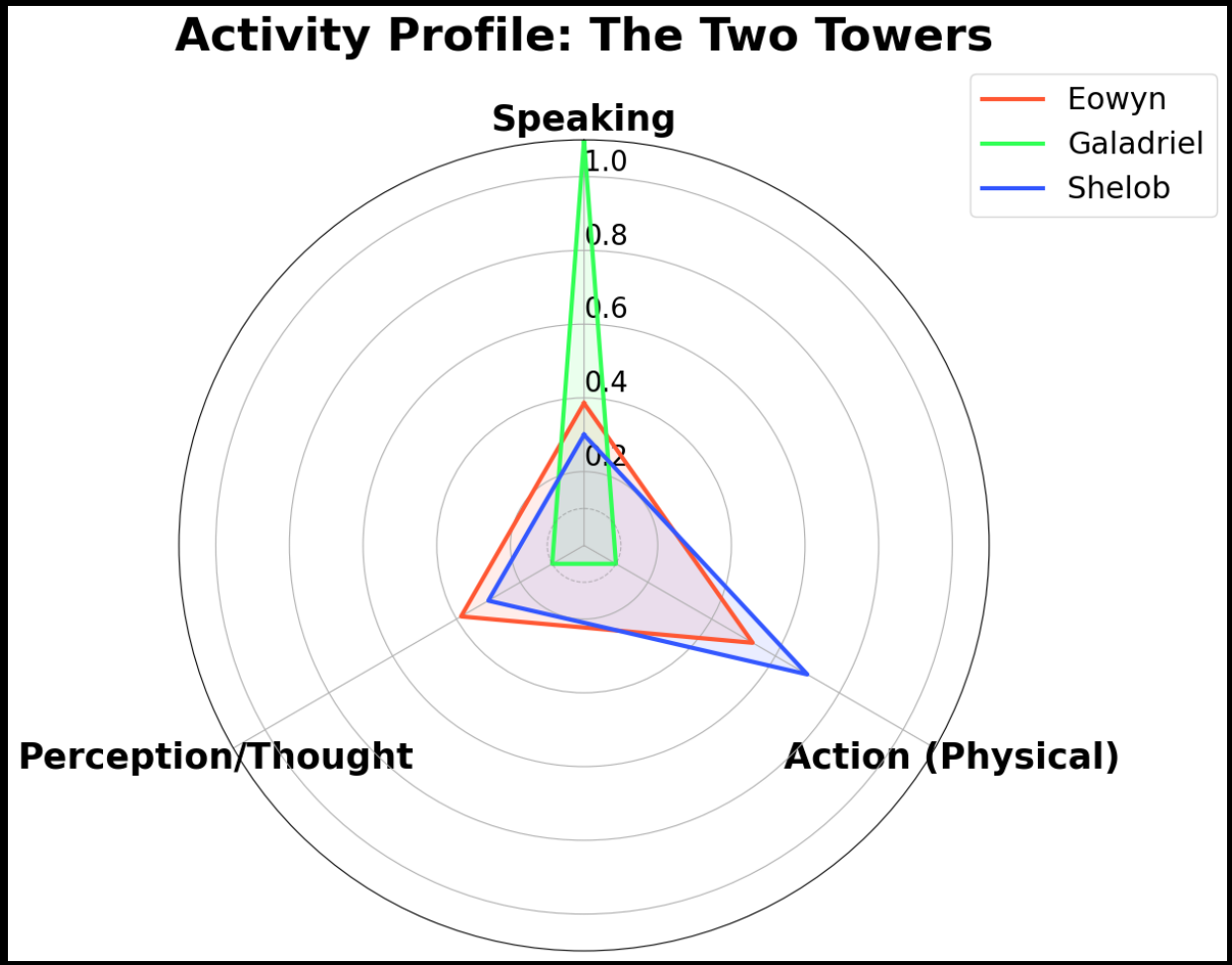




APS Ratios Compared
The Two Towers

Data Collection
The code analyzes character activity by scanning through the book text and identifying when character names appear followed by action verbs within a 6-word window, The following 6 words are examined to find verbs in three categories: **Speaking**, **Action (Physical)**, **Perception/Thought**

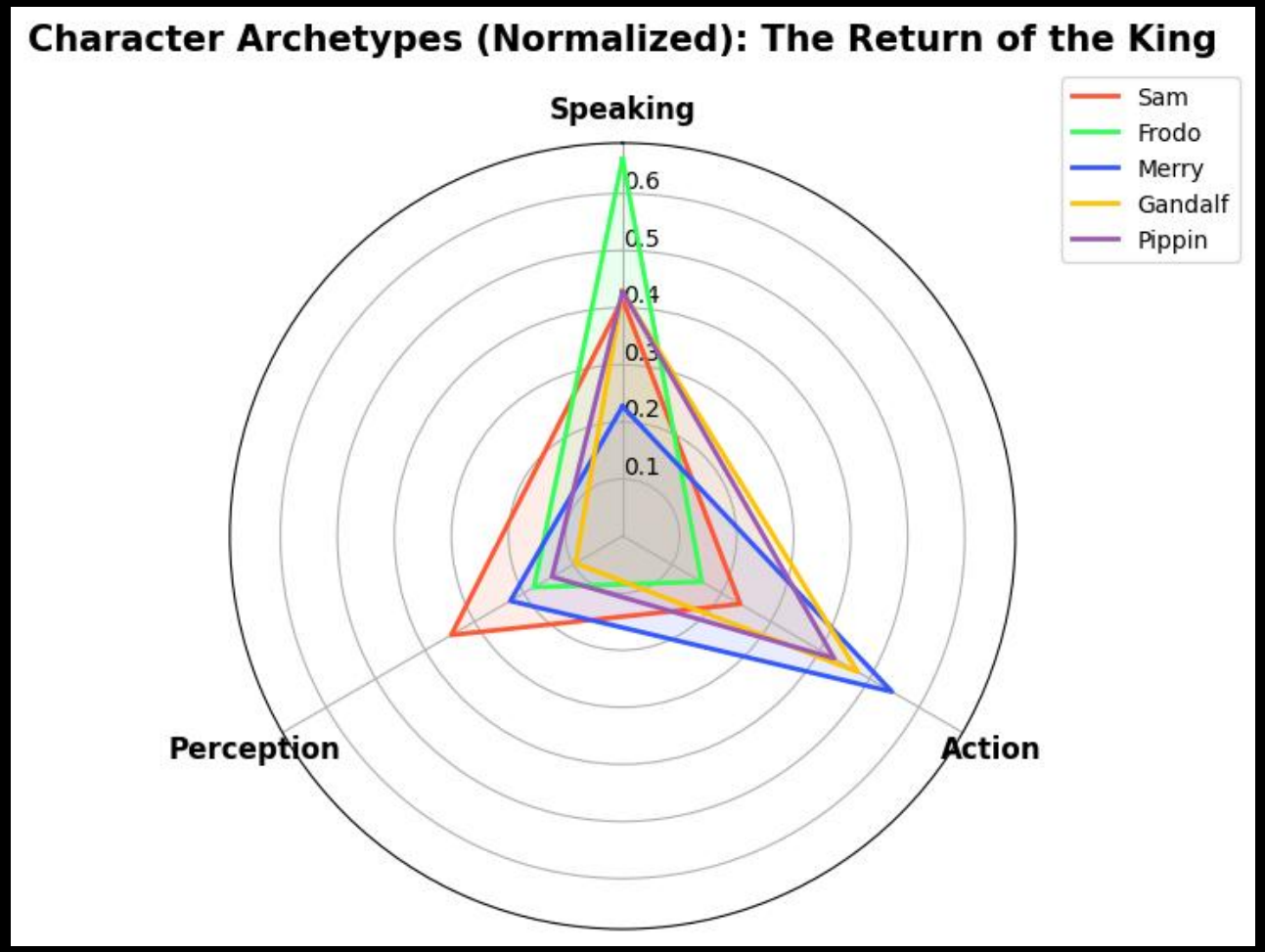


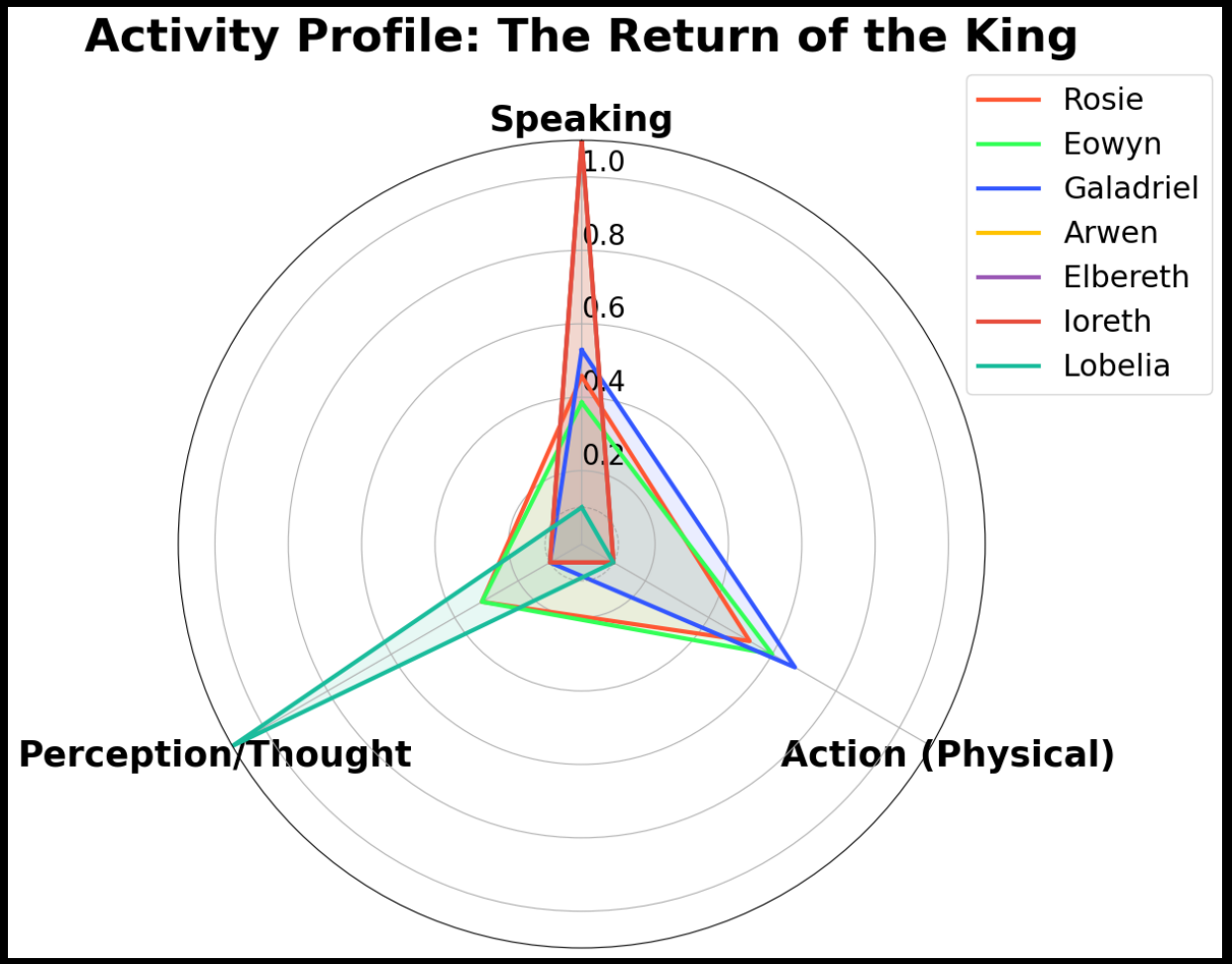


APS Ratios Compared
The Return of the King

Data Collection

The code analyzes character activity by scanning through the book text and identifying when character names appear followed by action verbs within a 6-word window, The following 6 words are examined to find verbs in three categories: **Speaking**, **Action (Physical)**, **Perception/Thought**





Team DVA

J. R. R. Tolkien, Final Project

Adilya Korkmaz
100001930

Adilya.Korkmaz@stud.srh-university.de

Vladyslav Stupak
100001664

Vladyslav.Stupak@stud.srh-university.de

Daniil Zykov
100003596

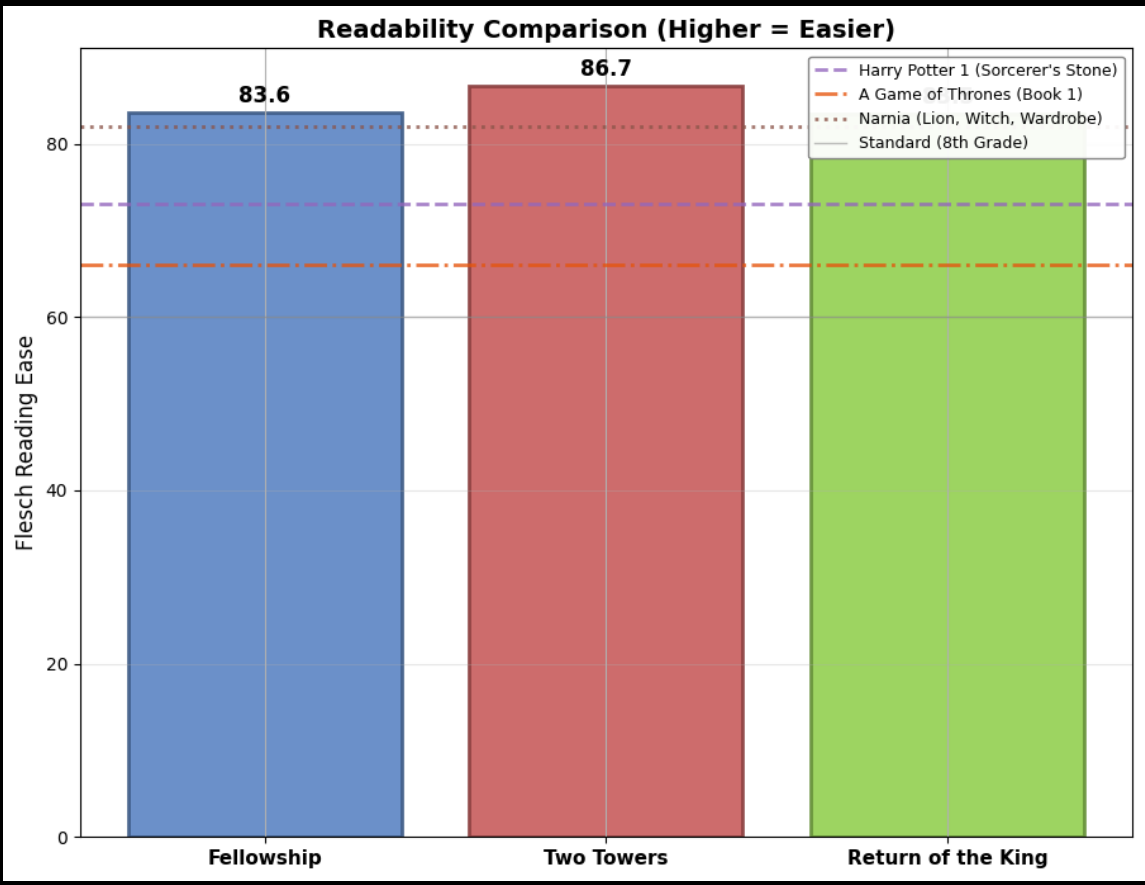
Daniil.Zykov@stud.srh-university.de

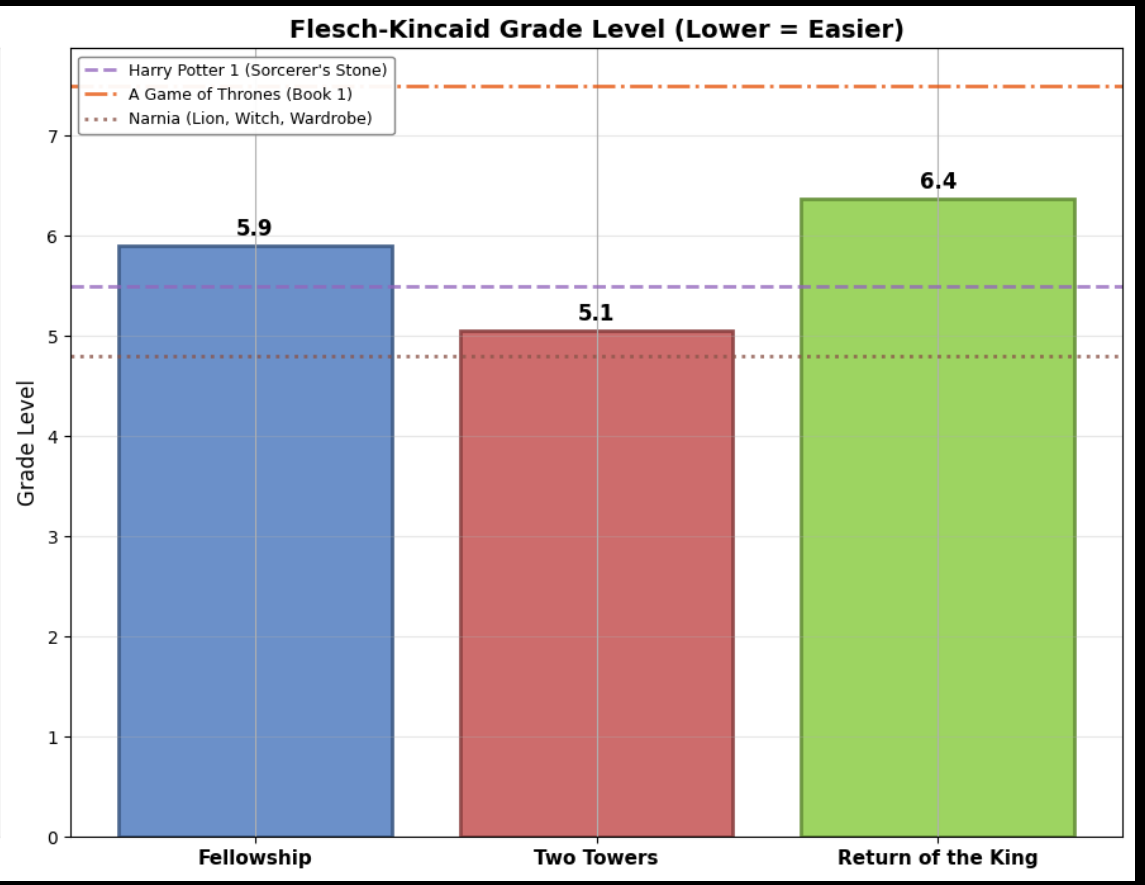
Readability Comparison

Methods include:

- **Extraction of readability metrics** (Flesch-Kincaid Grade Level) from a DataFrame.
- **Value labeling** on bars with `bar_label()` for clear numerical comparison.

- The charts compare readability across the three books using Flesch Reading Ease and Flesch-Kincaid Grade Level. Higher Flesch scores and lower grade levels indicate easier text

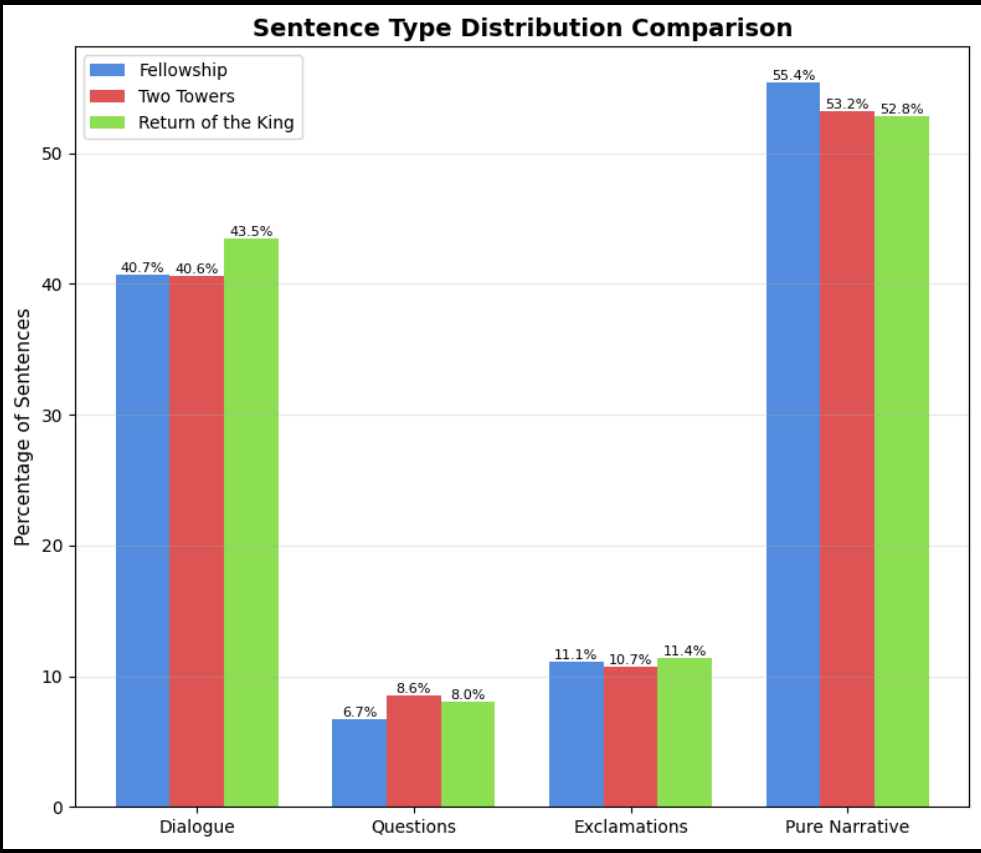


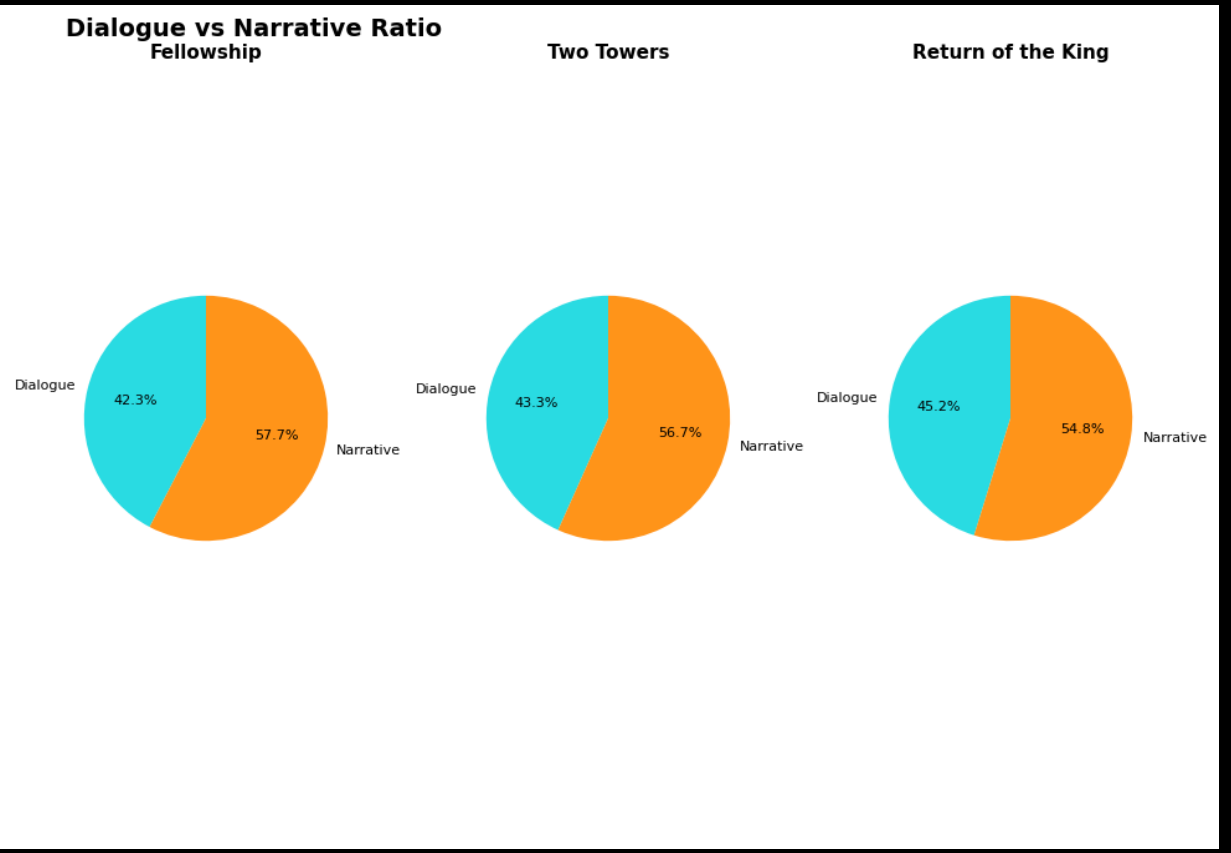


Distribution Comparison

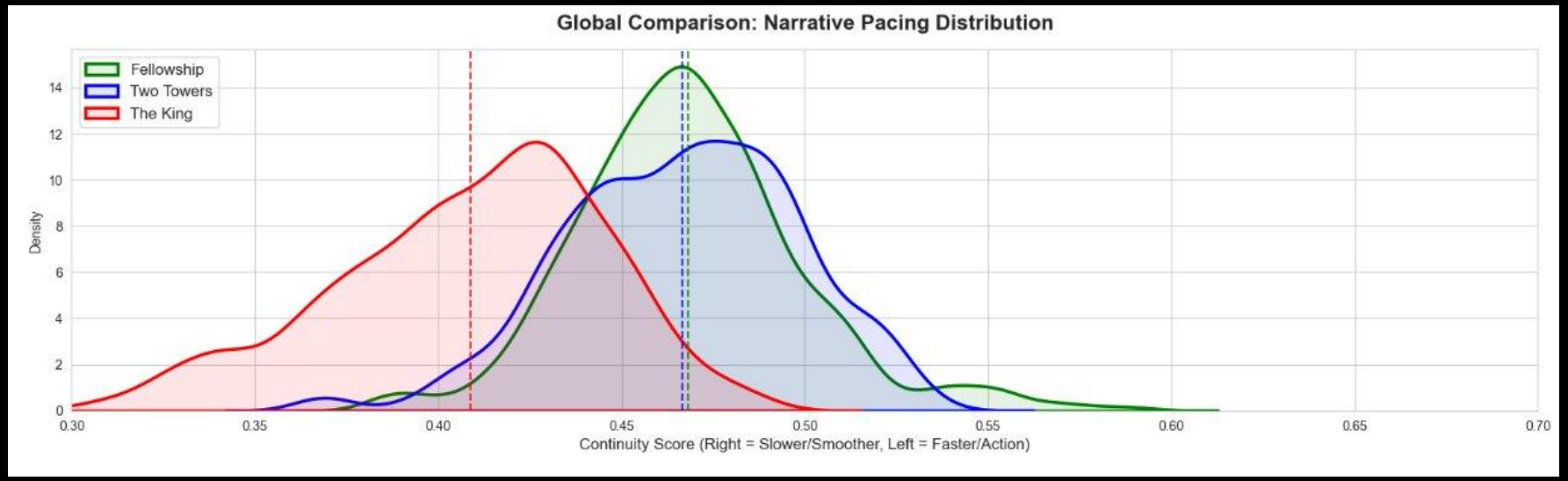
Methods include:
Custom subplot positioning using fig.add_axes() for precise control of the pie chart layout.

- The charts compare how often each book uses dialogue, questions, exclamations, and pure narrative. The pie charts highlight the overall balance between dialogue and narration in each book





Narrative Pacing & Topic Shift



- Sentence-Transformers turns sentences into multidimensional vectors and calculates cosine similarity between them. The closer the vectors – the more similar the sentences. We then create graphs sentence by sentence