



The Anna Karenina Paradox

Is the title misleading? An NLP analysis

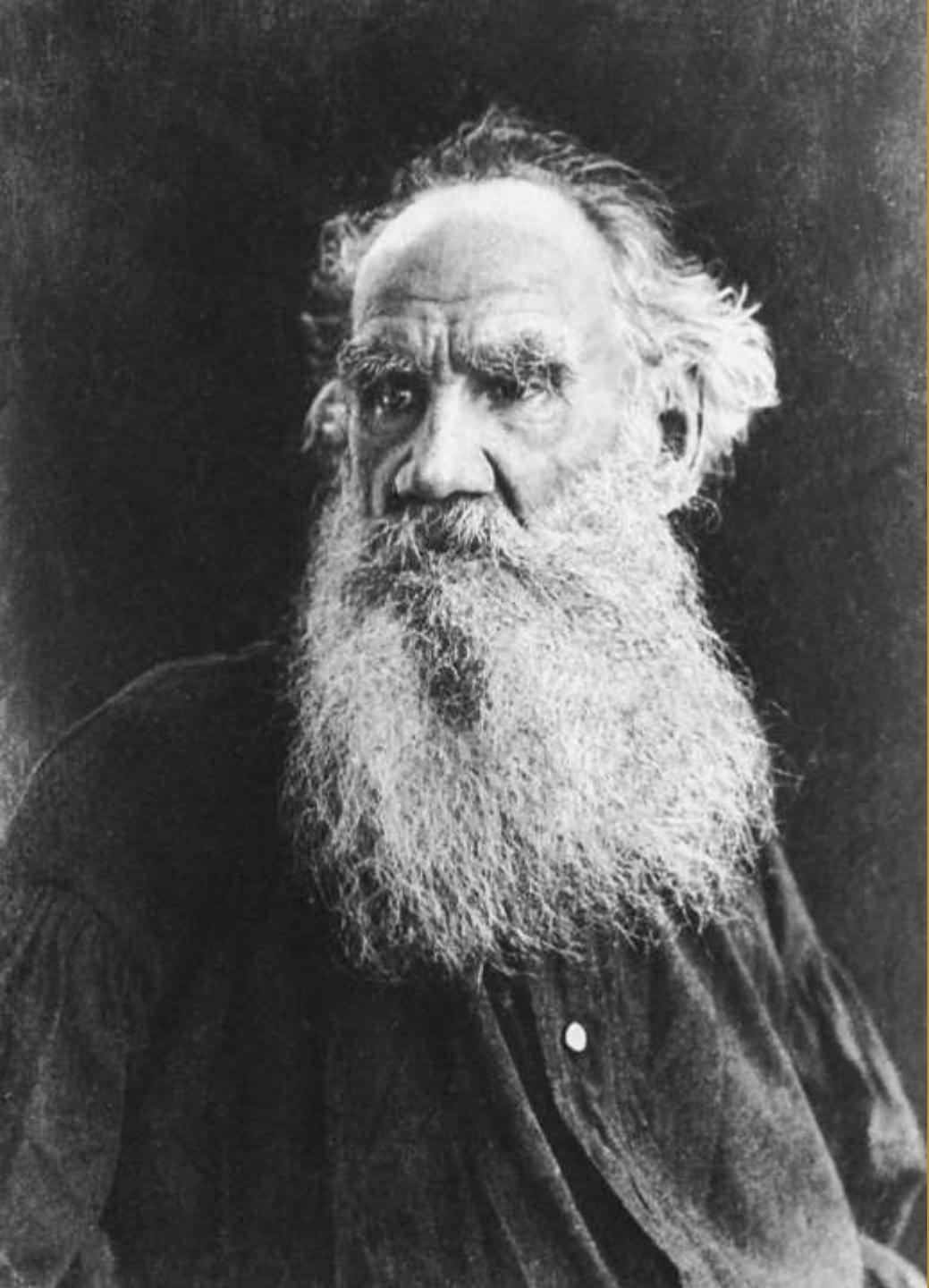
JAKHONGIR ERKINOV (100001943)
JORANAS CIGAS (100001693)
DAVED TAWDROS (100002565)
CHRISWIN BAIJU (100001868)

srh

AGENDA

- Introduction
- Context: The Anna vs Levin problem
- Our Approach
- Volume Analysis (Quantity)
- Influence analysis (Quality)
- Final Verdict



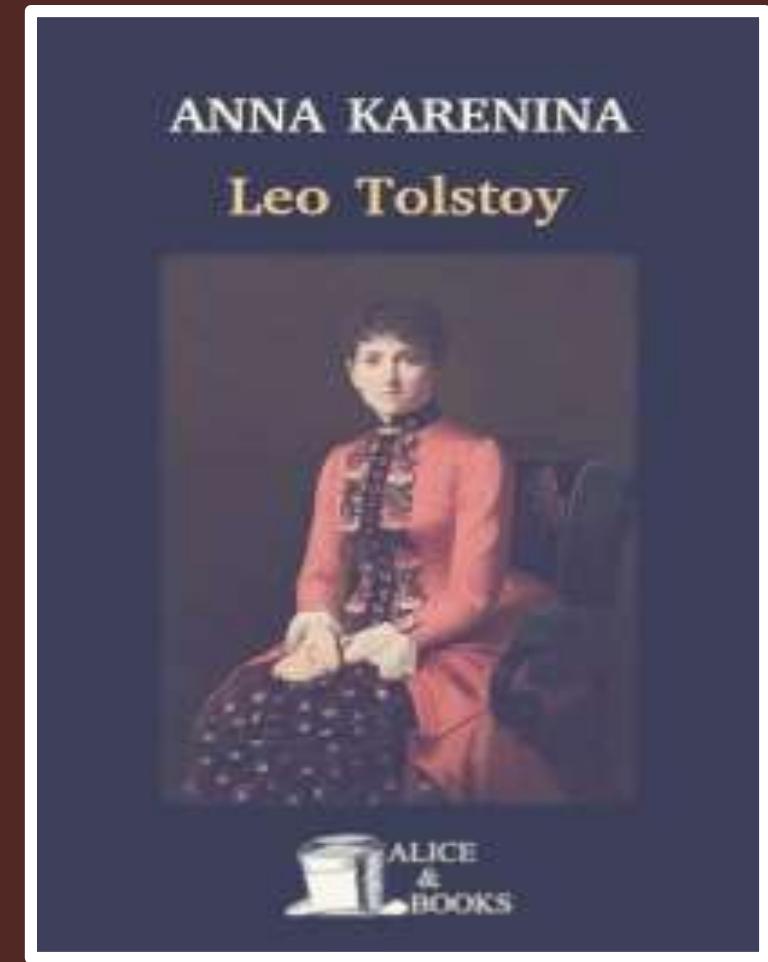


LEO TOLSTOY

- Russian author (1828–1910)
- Wrote epic realist novels
- Known for psychological depth + huge casts

CONTEXT: THE ANNA VS LEVIN PROBLEM

- The Expectation: Title suggests Anna is the sole protagonist.
- The Reality: Konstantin Levin appears just as often.
- The Twist: Levin is Tolstoy's semi-autobiographical self-portrait.
- Research Question: Does the data support the title?



OUR APPROACH: NLP & MAPPING

- **The tool:** Python Natural Language Processing (NLP) pipeline.
- **The Challenge:** Russian names are complex and inconsistent.
- **The solution:** We built a dictionary to map them..
- **Metrics:** Volume (Counts) vs. Influence (Network).

Name	Variations Mapped
Levin	Konstantin, Kostya
Anna	Arkadyevna, Karenina

ANNA

"Anna": "Anna", "Karenina": "Anna",
"Anna", "Arkadyevna": "Anna",
"Anna Arkadyevna": "Anna",
"Anna Karenina": "Anna",

VRONSKY

"Vronsky": "Vronsky", "Count Vronsky": "Vronsky", "Alexey Kirillovitch": "Vronsky",

LEVIN

"Levin": "Levin",
"Konstantin": "Levin",
"Konstantin Levin": "Levin",
"Dmitrievitch": "Levin",

A close-up portrait of a man with long, wavy hair, looking slightly to the side with a serious expression.

THE EVIDENCE FOR LEVIN

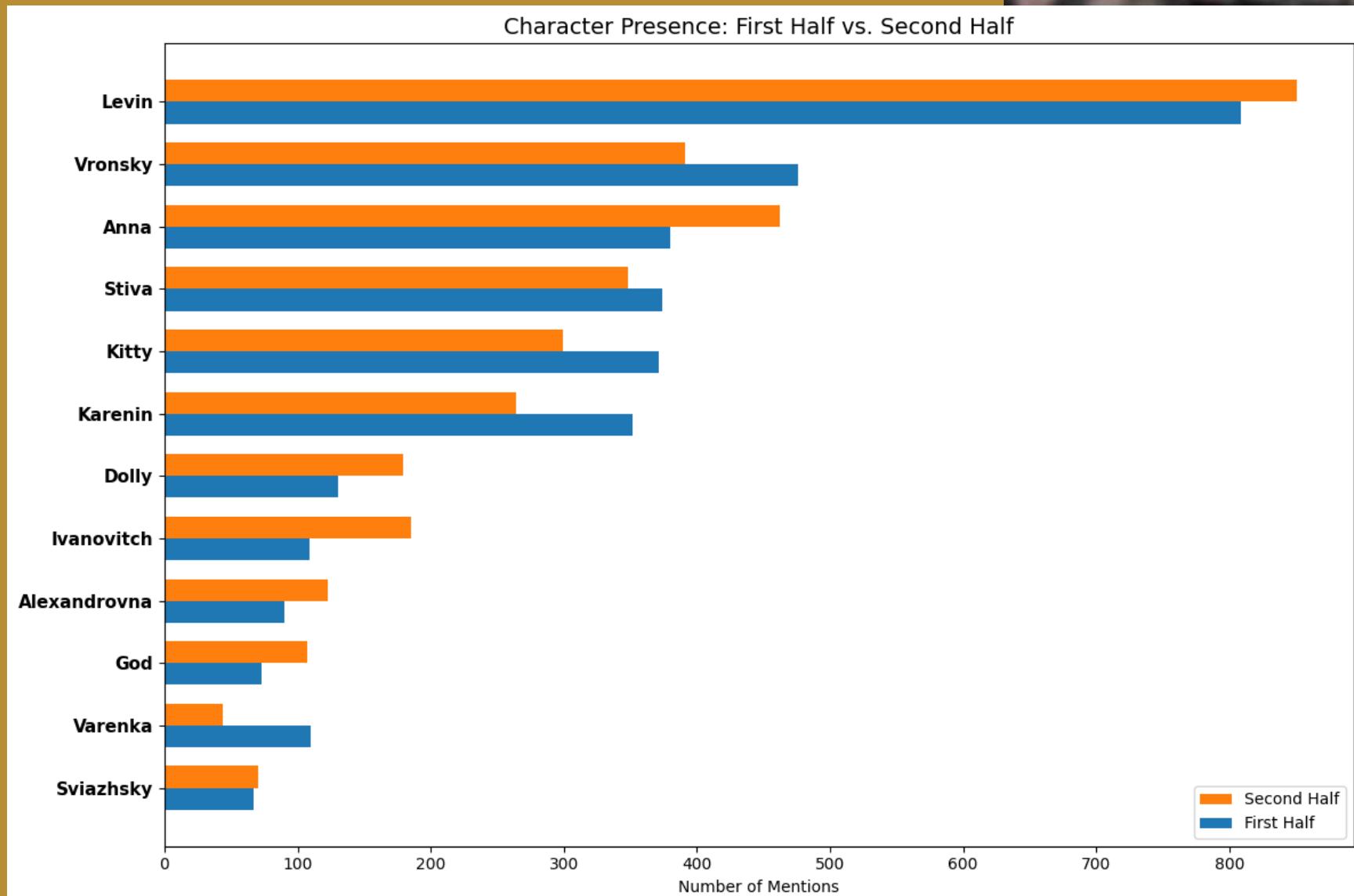
- *Quantity Analysis (Volume)*
-

IST METRIC

TOTAL MENTIONS

- **Bullet points:**
- Levin ~1,600 mentions
- Anna ~1,100 mentions
- Levin appears significantly more often

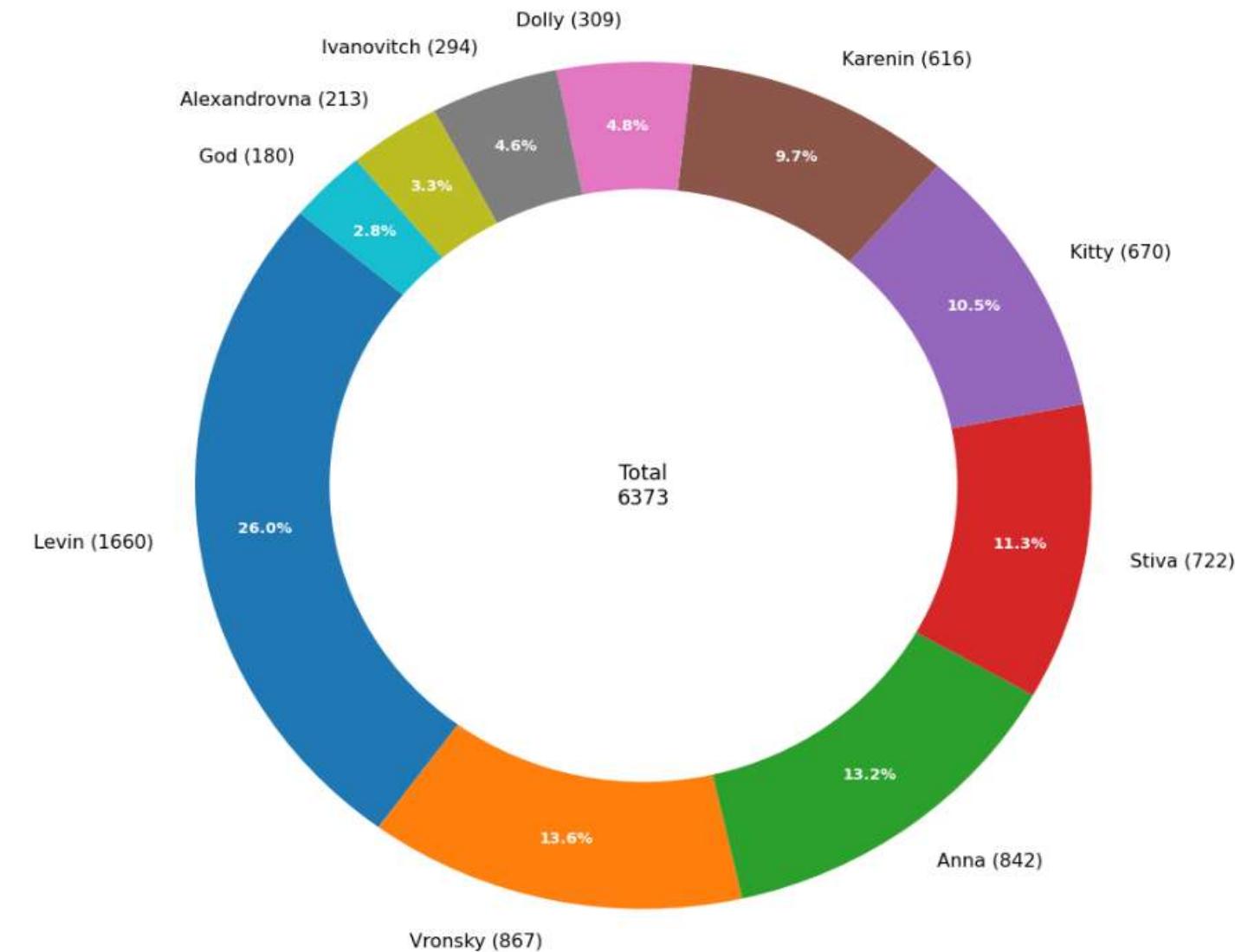
Tools: pandas, NLTK/spaCy, regex, matplotlib



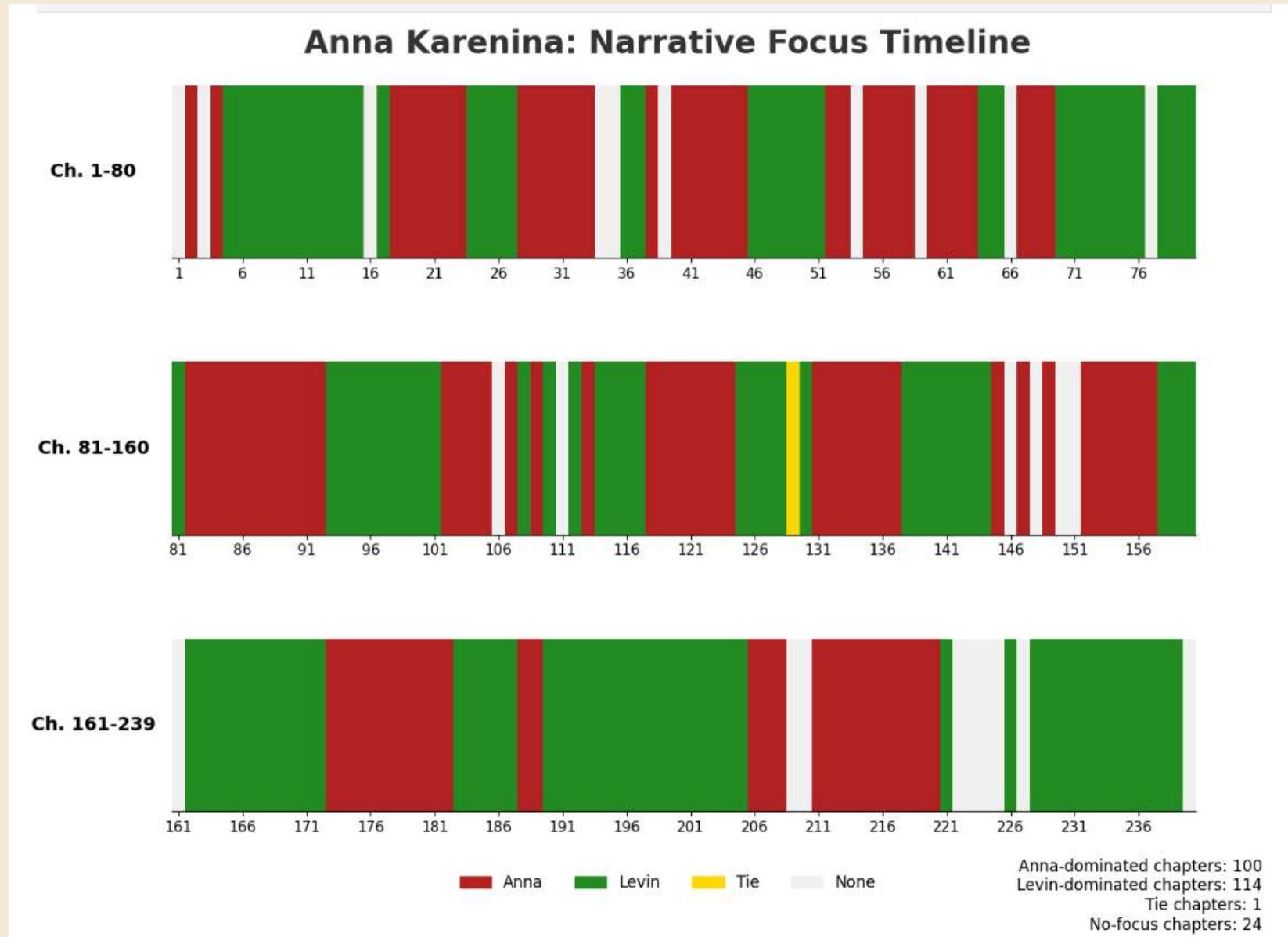
• Share of Mentions (%) by Character

- Total mentions: 6373"
- Levin has the largest share (26%)
- Anna is 13%

"Data processed with Python"
using pandas, NLTK/spaCy, regex,
matplotlib'



FOCUS TIMELINE



Bullet points:

- Legend:
- Red = Anna
- Green = Levin
- Yellow = Tie
- White = None

- Tools: numpy, pandas, pathlib, re, matplotlib

A close-up portrait of a woman with short, dark hair styled in a voluminous, textured updo. She has a serious expression, looking slightly upwards and to her right. Her skin tone is warm, and she is wearing small, light-colored stud earrings. The background is dark and out of focus.

THE EVIDENCE FOR ANNA

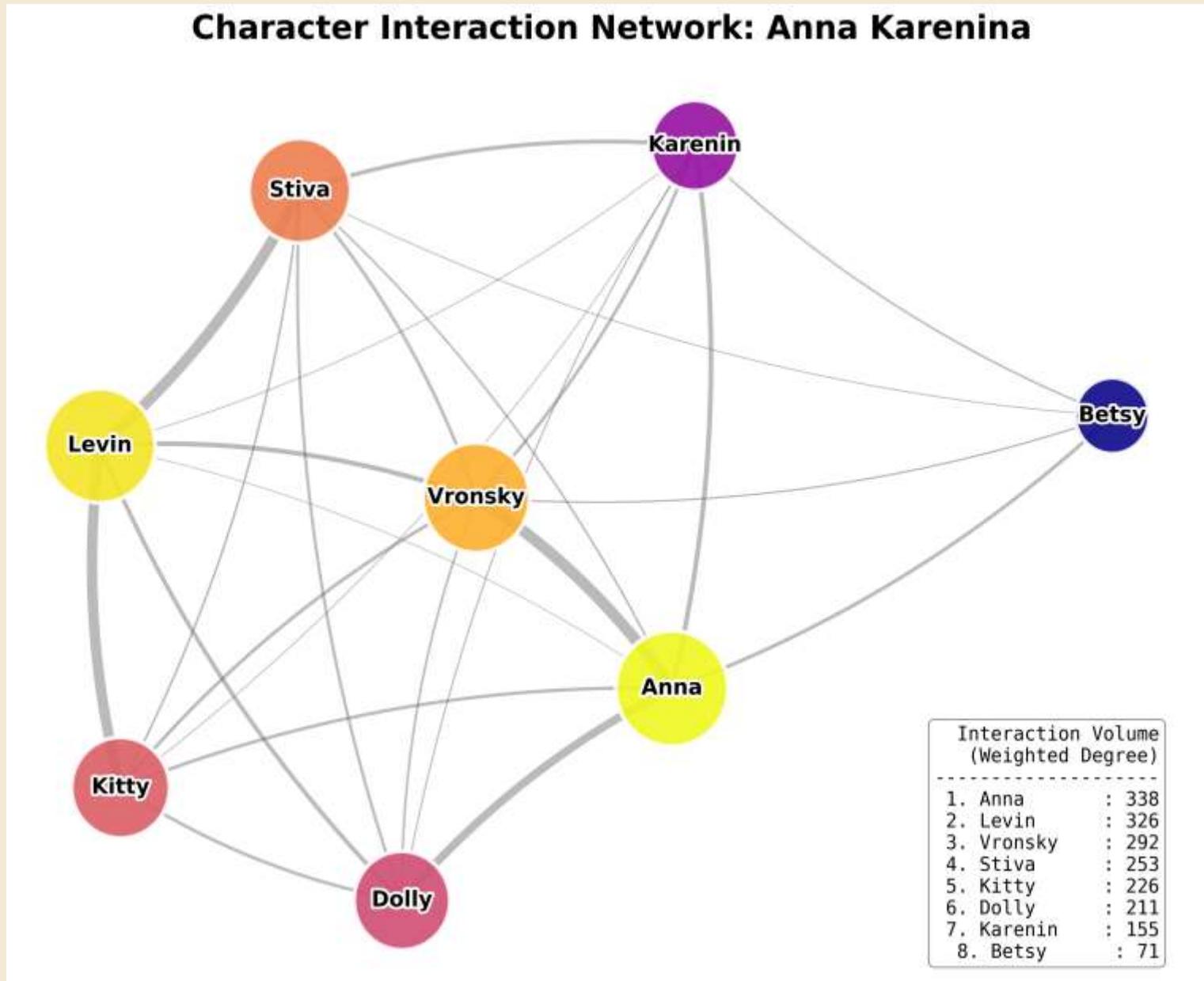
- *Influence analysis (Quality)*

2ND METRIC

THE NETWORK PARADOX

- **The Metric:** We measured **Interaction Volume** (Weighted Degree).
- **The Surprise:** Levin (326 interactions) is statistically almost as active as Anna (338). He is a heavyweight.
- **The Hub:** Anna is the **Structural Center**. She connects the disparate worlds (Politics, Society, Family). Without her, the network breaks.

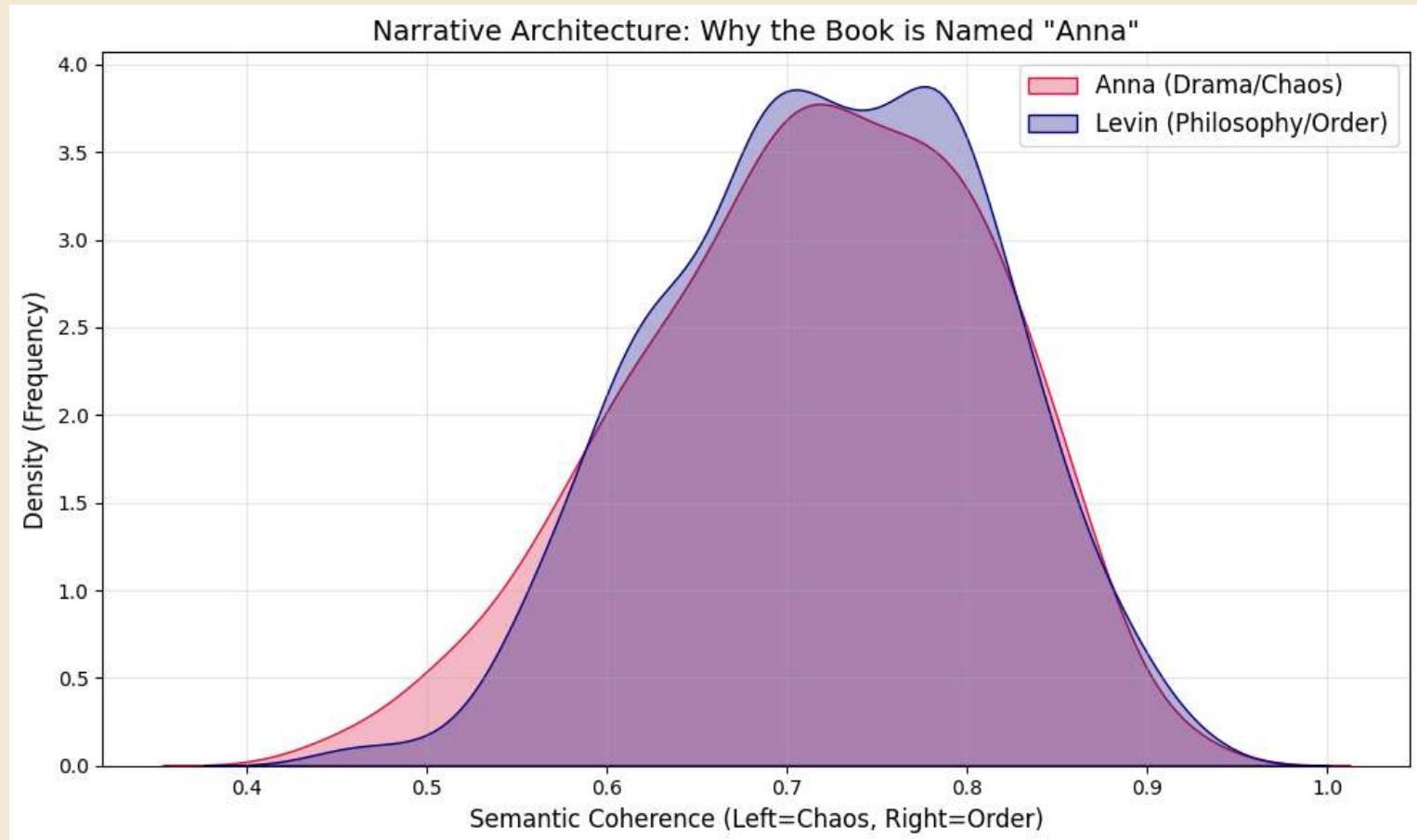
Tools: itertools, networkx, nltk, pandas, IPython, matplotlib



Paragraph Semantic Coherence

- **The Metric:** We measured **Semantic Coherence** to determine if the narrative voice was "stable" or "chaotic."
- **The Data:** Levin (**0.730**) scores higher than Anna (**0.719**). The difference is statistically significant.
- **The Insight:** Anna represents **Entropy and Drama** (lower coherence). Levin represents **Order and Philosophy**. The novel needs Anna's chaos to drive the plot.

Tools: datetime, matplotlib, numpy, pandas, regex, seaborn, sentence_transformers, sys, PyTorch, transformers



A woman with curly hair, wearing a dark turtleneck, is looking down at a book she is holding. The background is a soft-focus landscape with rolling hills and a cloudy sky.

FINAL VERDICT

WHAT DO WE UNDERSTAND SO FAR?

FINAL VERDICT

The Case for Levin (Quantity):

- **Winner:** Total Mentions & Focus Timeline.
- **Role:** He dominates the **Volume** of the text. He represents the novel's *Expanse* (Philosophy, Agriculture, Russia).

The Case for Anna (Quality):

- **Winner:** Paragraph Semantic Coherence & Network Centrality.
- **Role:** She dominates the **Structure** of the text. She represents the novel's *Focus* (Drama, Society, Connection).

Conclusion:

- **The Title is Accurate.**
- While Levin provides the *voice*, Anna provides the *gravity*. She is the structural hub that connects the novel's disparate worlds; without her, the network collapses.

About us

Jakhongir Erkinov
Jak.infosec@gmail.com

Joranas Cigas
jorceebusiness@gmail.com

Chriswin Baiju
chriswinbaiju25@gmail.com

Daved Tawdros
davidayad.2006@gmail.com