KENNESAW STATE UNIVERSITY
COLLEGE OF COMPUTING AND SOFTWARE ENGINEERING
School of Data Science and Analytics

A Song of Text and Fidelity: Analyzing Textual Similarity and Critical Reception in *Game of Thrones*

Alexandra Barrett– Spring 2027
Dr. Jitendra Sai Kota

INTRODUCTION

Book adaptations represent a multibillion-dollar industry—if done right. Adapting a beloved novel brings both opportunity and risk: win over fans and you have a *Harry Potter*-level success; miss the mark and you join the ranks of *Eragon*.

Game of Thrones provides one of the most striking examples of this tension. Once celebrated for its storytelling and world-building, the series became infamous for the sharp decline in its final seasons—coinciding with the point at which the show outpaced George R.R. Martin's unfinished ***A Song of Ice and Fire*** novels**.**

This project investigates whether the series' critical success can be linked to how closely the show's dialogue aligns with the source material. By comparing subtitles from ***Game of Thrones*** episodes to the text of the original books, I quantify textual similarity and explore whether higher similarity correlates with positive critical reception.

Through this analysis, **I aim to uncover the relationship and degree to which fidelity to the source material truly predicts critical success—**or if creative adaption can stand alone.

METHODS

Data Source

- Subtitles of Seasons 1-6 of *Game of Thrones* collected from publicly available transcripts.
- Text from the released books of *A Song of Ice and Fire* by George R.R. Martin.
- Metacritic ratings used to measure critical reception.
- Eight configurations of comparison texts listed below:

	Whole Book Series (WBS)	Chapter Mapped per Episode (CM)
All Sentences	WBS Sentences	CM Sentences
	WBS Sentences SWR (Stop Words Removed)	CM Sentences SWR (Stop Words Removed)
Only Dialogue	WBS Dialogue	CM Dialogue
	WBS Dialogue SWR (Stop Words Removed)	CM Dialogue SWR (Stop Words Removed)

Similarity Calculations

Textual Fidelity:

- **Exact Match –** percent of subtitle lines appearing in book text.
- **Levenshtein distance** – edit distance thresholds used to calculate episode-level similarity.
- **Jaccard Similarity (q-gram, q = 3) –** proportion of overlapping characters per line.

Semantic Similarity:

- **Word2Vec embeddings – Skip Gram** and **CBOW** models trained on book and subtitle text: cosine similarity used to compute episode averages.
- **Word2Vec & TF-IDF** – weighting embeddings to emphasize rarer, meaningful words.
- **GloVe** embeddings – co-occurrence based word vectors.

Analysis

Chi-square tests performed between similarity and ratings categorized by tertiles (Low: 0-33%, Medium: 34–66%, High: 67–100%). Polynomial regressions assessed percent similarity versus Metacritic ratings, assumptions checked, models validated via 10-fold CV and ANOVA.

Figure 1: Direction of Percent Similarity Effects on Metacritic Ratings
All Tests vs Statistically Significant Tests

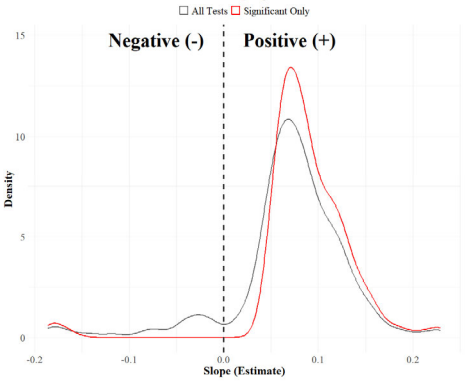


Figure 2: Average Similarity v. Metacritic Scores per Episode
Across All Comparisons and Methods

Table 1: Regression and Chi-square Analysis Across Similarity Methods

Similarity Method	Linear Regression Metrics: Mean R^2	Linear Regression Metrics: Mean Slope Est.	Linear Regression Metrics: % Models Significant (p<0.05)	Chi-Square Metrics: % Chi Sq Tests Significant (p<0.05)	Chi-Square Metrics: Mean Effect Size (Cramer's V)
Jaccard Similarity	0.23	0.11	100%	63%	0.32
Levenshtein (dist < 10)	0.17	0.10	98%	64%	0.30
Exact	0.14	0.13	88%	38%	0.27
Word2vec: CBOW & TF-IDF	0.12	0.09	50%	25%	0.26
Word2vec: CBOW	0.10	0.08	50%	13%	0.23
GloVe	0.08	0.004	50%	0%	0.19
Word2vec: Skip Gram & TF-IDF	0.07	0.05	50%	19%	0.19
Word2vec: Skip Gram	0.04	0.02	25%	0%	0.17

Regression results from linear models, Chi-square results from per case tests of independence between similarity categories and rating categories (High, Med, Low)

Figure 3: Average Statistically Significant Models' Performance
Across Methods and Book Comparison Types

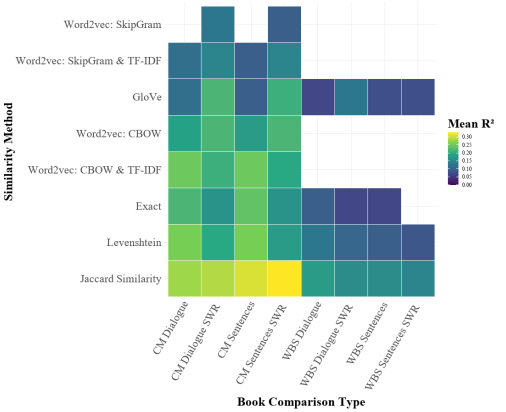
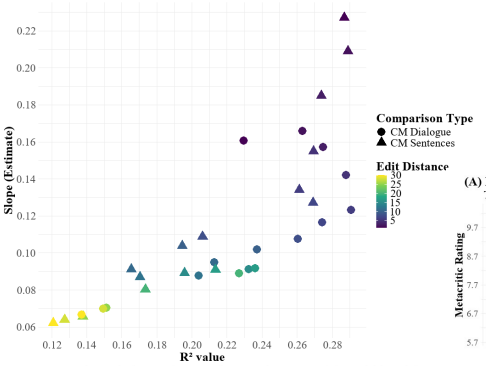


Figure 4: Percent Similarity vs Metacritic Rating Model Fit
Levenshtein Distance Threshold Effect



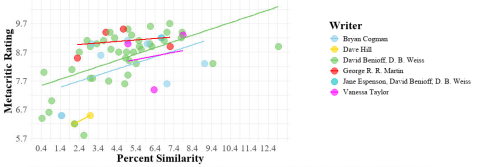
As Levenshtein edit distance decreases (becomes closer to exact) the model fit and slope increase.

Table 2: Top 10 Linear Regression Models

Similarity Method	R^2	Slope Est.	Comparison Type
Levenshtein (dist = 1)	0.29	0.23	CM Sentences
Levenshtein (dist = 2)	0.29	0.21	CM Sentences
Exact	0.24	0.23	CM Sentences
Levenshtein (dist = 3)	0.27	0.19	CM Sentences
Jaccard Similarity	0.33	0.11	CM Sentences SWR
Levenshtein (dist = 3)	0.28	0.16	CM Dialogue
Levenshtein (dist = 5)	0.29	0.14	CM Dialogue
Levenshtein (dist = 2)	0.26	0.17	CM Dialogue
Levenshtein (dist = 5)	0.27	0.16	CM Sentences
Jaccard Similarity	0.31	0.11	CM Sentences

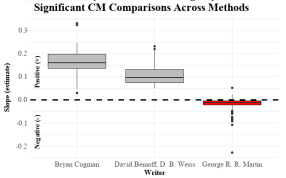
Figure 5: Writer-Level Similarity Effect on Metacritic Ratings

(A) Interaction Plot: Percent Similarity by Writer
Top Model: Levenshtein (dist = 1) CM Sentences



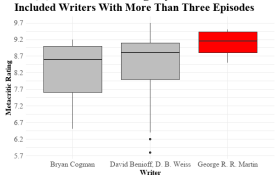
Slopes are consistently positive across majority of writers and nearly parallel, with little interaction. This aligns with the finding that only 6% of models improved when writer is interacting with percent similarity.

(B) Similarity Effects on Ratings by Writer
Significant CM Comparisons Across Methods



(B) Analyses include only writers with >3 episodes to avoid unstable slopes from very small samples. George R.R. Martin (original author) is highlighted. Slopes for other writers are positive.

(C) Metacritic Ratings by Writer
Included Writers With More Than Three Episodes



(C) George R.R. Martin's (original author) ratings are tightly clustered, giving his slope less room to vary.

Figure 5: George R.R. Martin shows the only consistent negative slopes, likely due to his low Metacritic score variability and the fact that similarity measures compare his episodes directly to his own source material.

RESULTS

Overall Trends:

- **Linear regressions favored in 83%** of cases ($p \geq 0.05$, ANOVA)
- 90% of slopes positive; **77% of models significant** ($p < 0.05$), 57% ($p < 0.01$). **97% of these slopes were positive**. (Figure 1)
- **Full-series analysis shows significant** positive correlations. (Figure 1 and 2)

Comparison Text and Similarity Method Analysis:

- **Chapter-mapped** had most statistically significant models with the highest R^2. (Figure 3)
- **Jaccard, Levenshtein, Exact strongest correlations on average,** chi-square analysis confirmed moderate association**.** (Table 1 and Figure 3)
- Levenshtein improves with stricter thresholds **(less distance = more exact)** (Figure 4)

Top models: (Table 2)

- Jaccard: Highest R^2 = **0.33**, β = 0.10 (each 1% similarity → +0.10 ratings)
- Levenshtein (dist = 2) R^2 = 0.29, β = 0.23 (Best R^2 and β)

Model Validation:

- RMSE from 10-fold CV confirms Jaccard, Levenshtein, Exact as best performers.

Writer's Effect:

- Interaction with Writer improved only 6% of models; Writer as the main effect improved 40% of models.
- **George R.R. Martin is the only writer with significant negative slopes**, and **lowest variation** in his ratings for writers > 3 episodes (Figure 5 (B) & (C)).

CONCLUSION

- **Linear models dominated (83%)** and gave clear, interpretable slopes.
- Across all models, **higher similarity between source texts and subtitles generally predicted higher Metacritic ratings**, with strongest effects observed for Jaccard, Levenshtein, and Exact match methods.
- **Closer-to-exact similarity was the strongest predictor** of Metacritic ratings across the show, **highlighting the importance of precise adaptation fidelity**.
- **Including Writer improved model fit in some cases**: 40% affected overall ratings, but only 6% changed the similarity–rating relationship**. George R.R. Martin is the only writer with significantly negative slopes**, likely due to low variation in his ratings and the fact that similarity compares his episodes to his own source material.

FUTURE WORK

- Refine word embedding models to capture greater context specific patterns.
- Investigate further George R.R. Martin's effect on models.
- Add additional variables (e.g., violence and production features) to assess impact.
- Examine other adaptations (e.g., *The Witcher* and *Dune*) to determine whether close textual fidelity yields favorable critic reception.