

# Extracting Sentiment Networks from Shakespeare's Plays

Paper authors: Eric Nalisnick and Henry Baird

Presentation by: Haris Memic and Rob Johnston

8 min. notification

# Presentation Structure

- Sentiment Analysis (General Perspective)

Linked

- Paper: Research Problem
- Paper: Methodology
- Paper: Results
  - a. Full-Text Sentiment Analysis
  - b. Character-to-Character Sentiment Analysis
  - c. Combining Sentiment Analysis with Social Networks
  - d. Subtle Rhetorical Analysis
- Paper: Limitations and Future Work

# General Perspective

# Sentiment Analysis: Applications

- Products & Services
  - benchmarking, reviews
  - marketing intelligence
  - purchase/consume decisions
- Public sentiment
  - consumer confidence
  - predictions (e.g. election results)

# Sentiment Analysis: Introduction

- Sentiment analysis = detection of attitudes
- Attitudes: enduring, affectionate, coloured beliefs, dispositions towards objects or persons [1]
- Components
  - Source
  - Target
  - Type
    - set of types: love, like, hate, value, desire,...
    - (un)weighted polarity: +, -, neutral
  - Text (document, paragraph, sentence, ...)

[1] Scherer. 1984. Emotion as a Multicomponent Process: A model and some cross-cultural data. In Review of Personality and Social Psych 5: 37-63.

# Sentiment Analysis: Algorithmics

- Tokenization
  - tokenizers
- Feature extraction
  - negation
  - all words (N.B.) vs subset of words (e.g. adjectives)
  - lexicons
- Classification (predicting binary or ordinal label)
  - Naïve Bayes (binarized multinomial)
  - Maximum Entropy
  - SVM

# Sentiment Analysis: Issues

- Negation
  - I **didn't** find this move **amazing**.
- Sarcasm
  - "What a **great** watch, stopped working after a week"
- Intention sentences
  - "I am **dying** to see that movie"

# Sentiment Analysis: Lexicons (1)

- Harvard's General Inquirer
  - <http://www.wjh.harvard.edu/~inquirer/homecat.htm>
  - +/- ; power/submission ; pleasure/pain/virtue/vice ; overstatement/understatement; ...
- Bing Liu's Lexicon
  - <https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html#lexicon>
  - list of + and - opinion /sentiment words
- AFINN Lexicon
  - <https://github.com/fnielsen/afinn>
  - list of words with valence intensities
- MPQA Lexicon
  - [http://mpqa.cs.pitt.edu/lexicons/subj\\_lexicon/](http://mpqa.cs.pitt.edu/lexicons/subj_lexicon/)
  - list of words with intensity
- VADER Lexicon
  - <https://github.com/cjhutto/vaderSentiment>



# Sentiment Analysis: Lexicons (2.a)

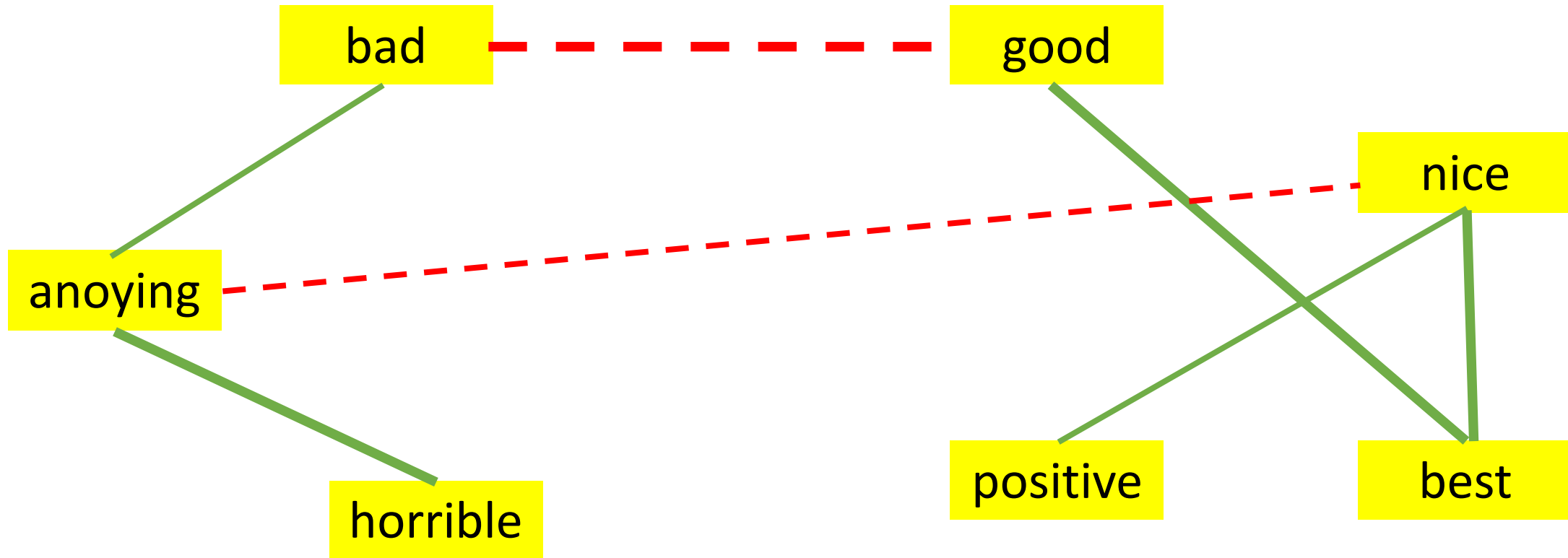
- No appropriate dictionary available – domain specific knowledge: bootstrap from few labeled/hand-build samples [2]:
  - a) Manually label a “seed” sample set of words
  - b) Expand with additional words:
    - words conjoined by “and” have same sentiment: “his car is old and ugly”
    - words conjoined by “but” have different sentiment: “the judge was fair but brutal”
    - thesaurus: synonyms (same sentiment), antonyms (opposite sentiment)
    - use nearby words
  - c) Classifier assigns “polarity similarity” for all pairs of words => graph
  - d) Cluster the graph
- Extensions:
  - Phrases [3], N-grams
  - Thesaurus
  - Deep learning [4]

[2] Hatzivassiloglou, McKeown. "Predicting the semantic orientation of adjectives." Association for Computational Linguistics (ACL), 1997.

[3] Turney. "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews." ACL, 2002.

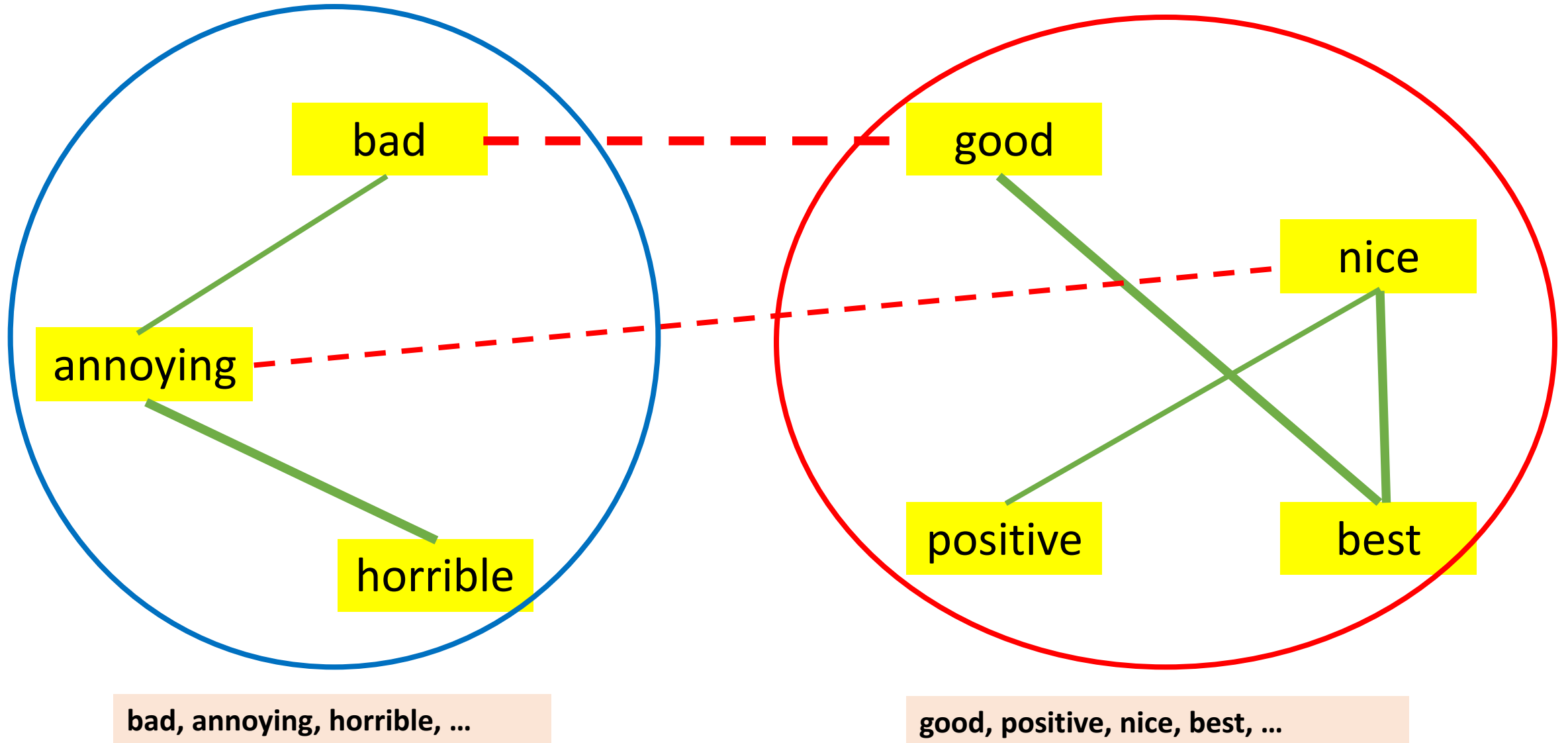
[4] Socher. "Recursive deep models for semantic compositionality over a sentiment treebank." *EMNLP*, 2013.

# Sentiment Analysis: Lexicons (2.b)



count (AND)  
count (BUT)

# Sentiment Analysis: Lexicons (2.b)



Paper

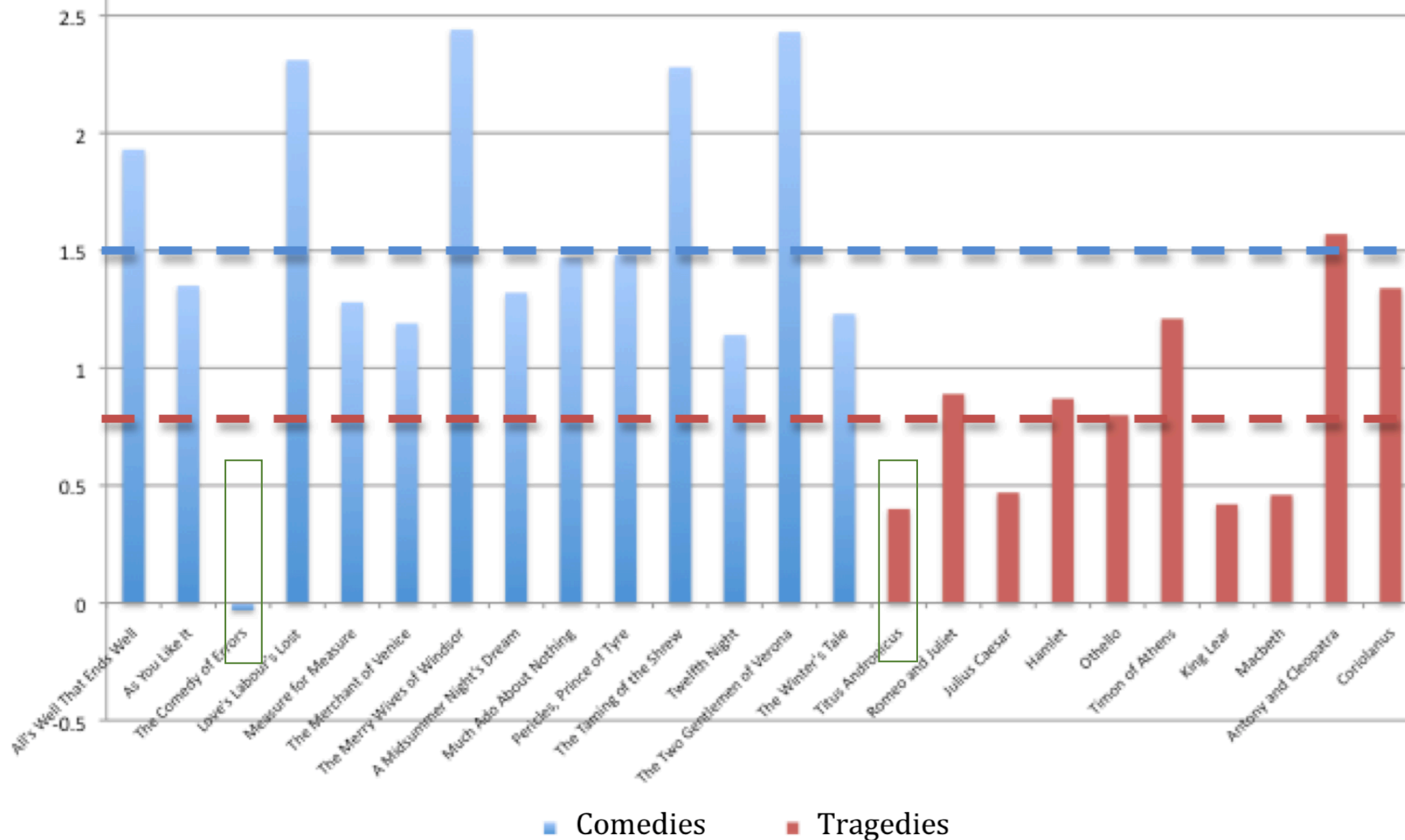
# Research Problem(s)

- Sentiment Analysis of Shakespeare's plays
  - Distinguish literary genres
  - Character-to-character sentiments (friendships, enmities)
- Social Network Analysis (SNA) of plays
  - Top-view of character relationships' polarity
  - Character motivations

# Methodology

- Dictionary approach
  - AFINN dictionary of valence-labeled words (2477)
    - scale: -5 to 5
      - “outstanding” = +5; catastrophic = -4
- Sentiment of individual plays
  - Average word valence = 
$$\frac{\sum valence(word_i)}{\# \text{ play's words in AFINN}}$$
- Social Network Analysis
  - nodes = play’s characters
  - edges = sentiment (valence values) towards immediate prior speakers
  - testing “Structural Balance Theory” by Marvel’s stress method [5]
  - square X adjacency matrix of un/friendliness scores between actors, until signs in X stabilize

# Results: Average Word Valence: Tragedies vs. Comedies

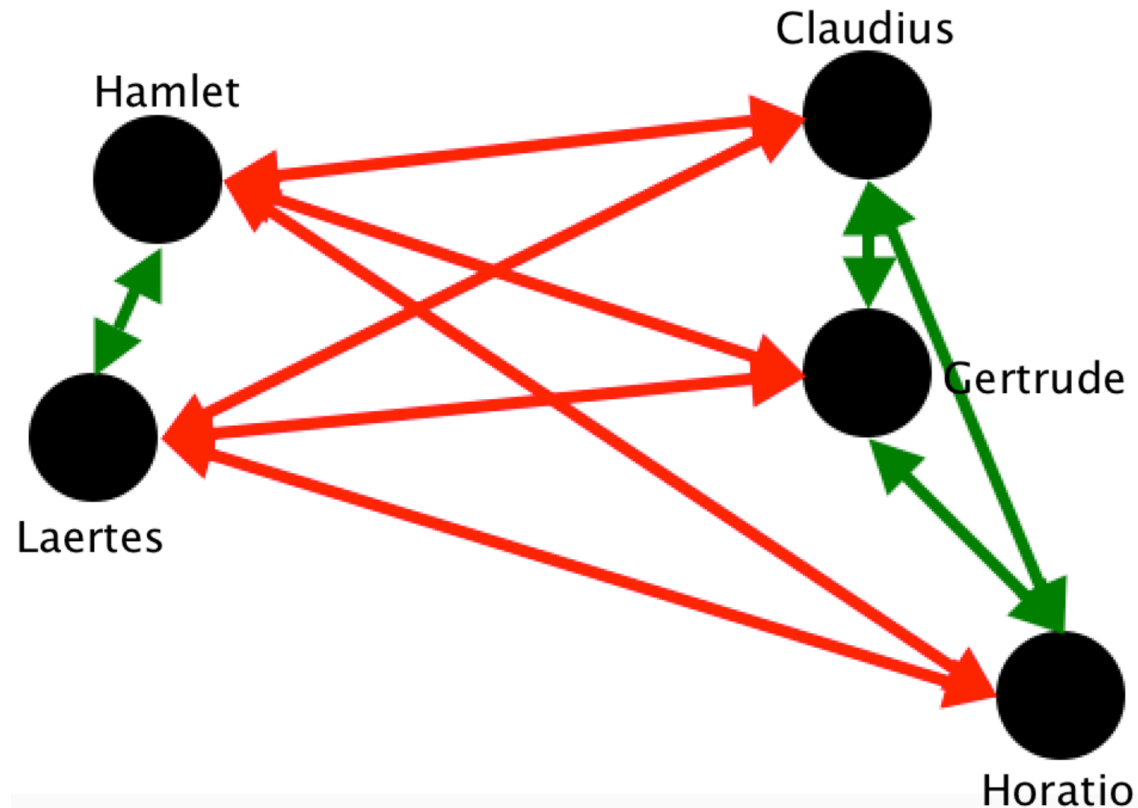


# Results: Character-to-Character Sentiment Analysis

Character	Hamlet's Sentiment Valence Sum
Guildenstern	31
Polonius	25
Gertrude	24
Horatio	12
Ghost	8
Marcellus	7
Osric	7
Bernardo	2
Laertes	-1
Ophelia	-5
Rosencrantz	-12
Claudius	-27



# Results: Sentiment Analysis & Social Networks



- Hamlet after Act II.
- Structural Balance Theory = friend of a friend is a friend
- Some SNA matrices/plays converged nicely, others (Othello) didn't

# Results: Subtle Rhetorical Analysis

- Tested Shak. plays for Pennebaker's markers of deceit [6]
  - Truthful debates ~ more self-references, adjectives, non-negative words
  - Avg % of words that were self-references, adjectives, adverbs, non-negative valence val.
    - Normalized scores (0-100) for each play
- Mixed results

Character	Truthfulness Score
OTHELLO	
1. Desdemona	100.00
2. Cassio	62.50
3. Othello	37.50
4. Emilia	0.00
5. Iago	0.00
HAMLET	
1. Polonius	100.00
2. Ophelia	81.25
3. Gertrude	68.75
4. Hamlet	62.50
5. Laertes	56.25
6. Horatio	37.50
7. Claudius	0.00
MACBETH	
1. Macduff	100.00
2. Malcolm	73.68
3. Macbeth	15.79
4. Lady Macbeth	0.00

# Limitations and Future Work

- Subtle rhetorical elements (deceit, irony, humor) not captured well
  - Sentiment networks didn't work for Structural Balance Theory
- Limited to Shakespeare literature
- 17<sup>th</sup> century English
- Edges = sentiment (valence values) towards immediate prior speakers
  - Sentiment analysis with speech attribution

# Exploratory Points

- Main: Paper Implementation (Python)
- Bonus: Expanded version of Udemy's "Text Mining, Scraping and Sentiment Analysis with R" (R)

# References

- Yan, Bo. “Sentiment Analysis of Twitter Data”, Master’s Thesis, Rensselaer Polytechnic Institute, New York, 2016.
- Jurafsky, Dan. Lecture materials for “CS 124: From Languages to Information”, Stanford University, 2016.
- Bing Liu. “Sentiment Analysis: Mining Opinions, Sentiments, and Emotions”, Sentiment Analysis Essentials, NLPCC, Shenzhen, 2014.