

Extracting Sentiment Networks from Shakespeare’s Plays

Eric T. Nalisnick and Henry S. Baird
Dept. of Computer Science and Engineering
Lehigh University
Bethlehem, Pennsylvania 18015, USA
Email: {etn212, hsb2}@lehigh.edu

Abstract—Automatic methods for analyzing sentiment and its movement through a play’s social network are investigated. From structured dialogue we can algorithmically determine who is speaking and guess at who is listening or being directly addressed. Knowing who is speaking to whom allows the flow of sentiment to be tracked between characters and, within plays with clear time-lines, permits tracking the development of emotional relationships. We hypothesize that changing polarities between characters can be modeled as edge weights in a dynamic social network— a “sentiment network”—which can be used to distinguish a document’s genre (tragedies versus comedies), detect a given character’s enemies and allies, and model the overall emotional development of a play. Experiments on Shakespeare’s plays are presented along with discussion of improvements and further applications.

I. INTRODUCTION

Modeling and understanding the emotional content and relationships described in a document remains a daunting technical challenge that is nearly as hard as generalized machine reading. Humans bring deep contextual knowledge to everything that they read. Although machines can quickly access huge databases, these remain too incomplete and unstructured to support the rich inferences that humans routinely make [14], thus leaving shallow statistical analysis as the only scalable alternative. Syntactic analysis, *i.e.* tagging each word with its part of speech (POS) and checking conformity to syntactic rules, has been achieved for many languages including English [21]. And some progress has been made in extracting the semantics of short passages of text [17]. One increasingly effective recent approach adding another dimension to shallow reading is *sentiment analysis*, in which words are mapped to *valences*, numbers intended to represent a range of negatively and positively ranked emotions [18]. For example, the WordNet Affect Lexicon (WAL) associates words with varying degrees of the six Ekman emotions: joy, sadness, anger, fear, disgust, and surprise [18].

Insightful analysis of literary fiction often challenges trained human readers and has long baffled computerized methods. Automatic analysis of literary documents has, so far, been shown useful in applications that either do not consider the content of the story (*e.g.* attempting merely to attribute authorship [23]), or that require the intervention of human experts (*e.g.* making inferences based on word occurrence statistics [16]). There appears to have been little progress in robust, automatic methods to extract and make use of plot, setting, and relationships among characters. We hope that

progress might be possible by focusing systematically on the automatic analysis of interactions among characters, and we begin by applying methods for sentiment analysis on long text passages including entire plays.

We discuss related work in the next section. The third section describes how we estimate character-to-character sentiment within plays and presents our extraction of sentiment networks (among all the characters). Lastly, in sections IV and X, we suggest future work and present a few tentative conclusions.

II. SENTIMENT ANALYSIS AND RELATED WORK

Algorithmic sentiment analysis is now widely used commercially to infer user opinions from product reviews [19] and social-media messages [15]. Most such systems accumulate word-occurrence statistics, driven by fixed lists associating words with “valences” (signed integers representing positive and negative feelings) [5]. Some also are sensitive to syntax commonly associated with expressions of dis/pleasure [19], and others make inferences via link analysis (*e.g.* one’s views may be influenced by one’s friends’) [6]. And some even refine sentiments into categories associated with specific emotions [18]. Distinguishing topics as well as emotions can matter [20]: *e.g.* when a user writes, “I love the interface but hate the customer support,” the two cases need to be teased apart to attribute dis/pleasure accordingly. There is also serious interest in detecting fraudulent reviews [7]. Unsurprisingly, methods which, like these, lack deep understanding often work more reliably as the length of the input text increases.

A. Sentiment Analysis in Literature

Turning our attention now to the automatic analysis of literary fiction, we find fewer clear successes. Perhaps the most intensively studied application has been authorship attribution (AA). Typically, after deleting “stop words”¹, an AA system tags each remaining word with a POS and groups these tags into N-grams [23]. Machine learning methods model the N-gram distributions separately for each author of interest and the resulting classifiers attempt to validate authorship claims and suggest authors for anonymous works. AA systems have, famously, been wielded in debates over Shakespeare’s true identity [23].

¹Oversaturating words whose use is motivated more by proper syntax than by meaning: *e.g.* ‘the’, ‘a’, ‘to’.

While robust analysis of semantic content (*i.e.* plot, setting, characters, etc.) of works of fiction is still open, this problem has been attacked by using networks to model plot as character interactions. Mutton [2004] adapted methods for extracting social networks from Internet Relay Chat (IRC) to mine Shakespearean plays for their networks [10]. Novels would seem far harder, but Elson et. al. [2010] developed a reliable method for speech attribution in unstructured texts and then succeeded also in extracting social networks [3][2][1].

The only sentiment analysis of Shakespeare we know of was done by Mohammad [2011]. He tracked emotions, over time, in several genres of English fiction including fairy tales and novels [9]. He has proposed, as future applications of sentiment analysis, search and summarization of literary texts. We now extend his work by performing sentiment analysis at the character level.

III. SENTIMENT ANALYSIS WITHIN SHAKESPEARE'S PLAYS

Our primary goal has been to explore what sentiment analysis, applied fully automatically—with a minimum of hand-crafted adaptations—can reveal about the narrative content of fiction. Can it distinguish among literary genres? Can it map friendships, enmities, and alliances and track their evolution? Can it shed fresh light on characters and their motivations? Can it offer insights into plot structure?

We chose the *AFINN* word list for sentiment analysis [13], which contains 2477 English words, labeling each with a *valence*, an integer between -5 and +5 (-5 denoting the most negative words and +5 the most positive). For example, *catastrophic* has a value of -4, *outstanding* has a value of +5. We did not modify this modern word-list before applying it to 17th C. literature in the hope that our methods might work at least as well on any play written since.

A. Full-Text Analysis

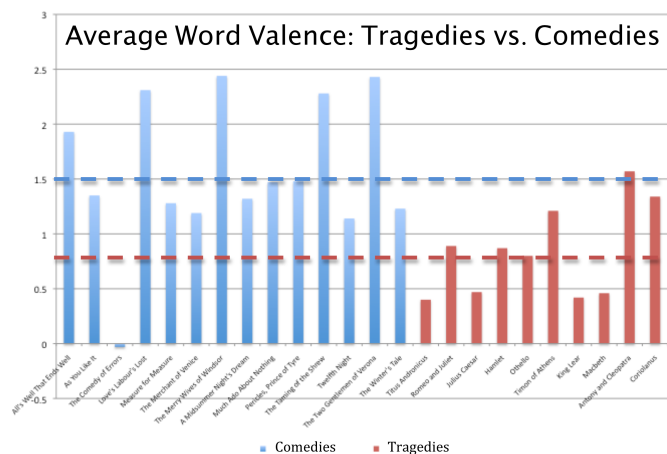


Fig. 1. For each play, the valence of each word was summed and then divided by the number of words in both the word list (AFINN) and the respective play. The result is the average word valence (AWV), which seems to discriminate well between tragedies and comedies. The blue line marks the average AWW for comedies and the red line represents the same value for tragedies.

Our first experiment tested if sentiment could differentiate comedies from tragedies. For each play, we calculated the

average word valence (AWV) by summing the valence values for all words in the play and then dividing by the number of the play's words in AFINN. Fig. 1 shows that comedies generally have a higher AWW than tragedies (as we expected). The lighter, higher dotted line marks the average value for comedies, 1.55, which is almost double the average AWW for tragedies, 0.84, as marked by the other dotted line.

There are several notable observations in the results. *Titus Andronicus* has the lowest AWW of the tragedies (0.40), which is a success since the play is widely considered to be Shakespeare's bloodiest and most violent work. But also notice that *The Comedy of Errors* (COE) has an even lower AWW of -0.03. There are several possible explanations for this outlying value. The play was one of Shakespeare's earliest works (1st to 3rd) and therefore might be stylistically immature. Furthermore, upon examining the play's content, the plot features farcical but dark elements such as beatings, infidelity, theft, and madness, which probably have been taken at 'face value' by the shallow analysis. Yet there are some critics that might say COE's AWW value is not an outlier. Gwyn Williams claims, in her essay *The Comedy of Errors Rescued from Tragedy*, that "A careful analysis of this play, however, shows that it might easily have worked out as a tragedy" [22]. The sentiment analysis seems to support her claim.

B. Character-to-Character 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

Fig. 2. The characters in *Hamlet* are ranked by Hamlet's emotions towards them. Every time Hamlet speaks, his words' valence values are summed and assumed to capture his feelings towards the previous speaker. Claudius draws the most negative emotion, as expected.

Although document-wide sentiment analysis can be useful, as shown above, we believe a better model lies in analyzing the flow of sentiment from character to character because fiction is fundamentally a description of character interactions. When literature is analyzed only as a whole, the interactions and character details are glossed over and missed. We mined for character-to-character sentiment by summing the valence values over each instance of continuous speech and then assumed that sentiment was directed towards the character that spoke immediately before the current speaker. This assumption doesn't always hold, but we've found that the positivity or negativity each character feels towards one-another at the culmination of the play can be tracked with some accuracy via this method. For example, Hamlet's sentiment rankings upon the conclusion of the play are shown in Fig. 3. Not

surprisingly, Claudius draws the most negative sentiment from Hamlet, receiving a score of -27. We investigate the potential of static and dynamic character-to-character analysis thoroughly in [11].

C. Combining Sentiment Analysis with Social Networks

As discussed in Section II, extracting social networks from documents has been shown to be valuable. Yet no previous work has given edge weights to these networks via sentiment analysis. Doing so produces a model that better captures the natural properties of literature: the polarity of all character relationships is captured at once.

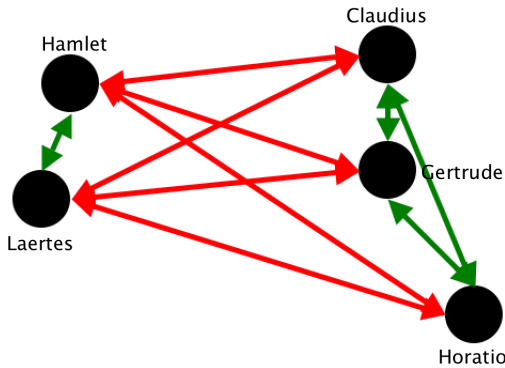


Fig. 3. The above network shows Structural Balance Theory’s prediction of how the characters in *Hamlet* will split allegiances when undergoing trauma. Green lines mark positive sentiment between characters, red lines mark negative sentiment. The network was obtained by repeatedly squaring the sentiment network’s adjacency matrix until the polarity (sign) of edges stabilized.

Similar work has been done in extracting social networks from emails and rating connections. Groh and Hauffa used a variety of features to characterize social relations [4]. Their features include: relative message frequency, message frequency balance, message length, use of first name, and word valence among others. We analyzed *Hamlet* using these features but the results did not seem as useful as just word valence. In literature, interaction between a majority of characters is bound to occur (or why would they be in the play in the first place?). Groh and Hauffa analyzed a domain (emails) in which interaction is sparse, which most likely explains their attributes’ more successful application.

As shown in the previous section, the results of character-to-character analysis can be done manually, but we also wanted to explore the possibility that these networks can be explained by a general sociological model. We tested if Shakespeare’s plays obey *Structural Balance Theory* (SBT), the idea that a friend of a friend is also your friend (i.e. transitivity of sentiment) [8]. This theory is captured by triangles in the graph. If two sides of a triangle have positive sentiment, the third side is likely to have positive sentiment or become positive over time, or vice versa for negative connections.

Marvel et al. showed that SBT can not only be modeled by a graph, but the evolution of relationships under stress can be simply modeled [8]. If X is an $n \times n$ adjacency matrix with nonzero entries that represent real-valued un/friendliness

scores towards other nodes in the network, $\frac{dX}{dt} = X^2$. That is, repeatedly squaring the adjacency matrix simulates stress on the network, eventually causing the signs of X ’s entries to stabilize. Marvel et al. demonstrated that their method could accurately predict observed, real-world network fracturing. For instance, their method correctly predicted how countries sided during World War II.

We ran Marvel et al.’s procedure on the adjacency matrices of our Shakespearean sentiment networks to test if we could predict how a play’s characters will split into factions just from looking at the state of the sentiment network after Act II. We saw varied results. Some plays converged to an arguably reasonable division. *Hamlet*’s results showed Gertrude, Claudius, and Horatio in one group, Hamlet and Laertes in an adversarial group (Figure 5.7), and Ophelia and Polonius as occupying middle ground, having positive and negative connections to both sides. But *Othello*’s matrix did not converge to a reasonable division. After every time step, Othello and Iago had a positive edge between them, which contradicts the enmity the reader knows Iago has for Othello. But since that hate is never revealed to Othello’s face until the end of the play and Iago provides his commander with nothing but flatteries until that point, the sentiment analysis does not and cannot detect the deceit. This problem shows a clear inadequacy in our shallow SA methods. Overall, we did not find the factions calculated to be compelling enough to support the idea that Shakespeare’s plots can be accurately modeled mathematically by SBT.

D. Inadequacies of Shallow Sentiment Analysis

Given the problems with sentiment analysis seen first hand in our SBT experiments, we looked for some way to identify subtle rhetorical elements such as irony and deceit. We wanted some shallow method that could identify that Iago is lying and does not have the positive sentiment evident from superficial analysis.

Psychologist James W. Pennebaker has had some success in this area. By studying the rhetorical tendencies of individuals in various mental states, he pinpointed several shallow linguistic features or patterns that can reveal deeper emotions. For example, Pennebaker compared debates in which students supported an argument that they agreed with and one that they did not, essentially lying in the latter case. The truthful debates showed an increased number of self references, adjectives, and non-negative words when compared to the untruthful debates [12]. He reasoned that liars tend to distance themselves from their falsehoods (lack of self reference) and provide less details since lying is cognitively more challenging than telling the truth.

Although Shakespeare’s characters and their emotions are artificial, artistic constructions and thus not likely to reveal psychological ‘tells’ the way Pennebaker’s specimens did, we tested Shakespeare’s plays for Pennebaker’s markers of deceit, hoping the analysis would show Shakespeare’s villains as more untruthful than the heroes. We analyzed Shakespeare’s three most famous tragedies: *Othello*, *Hamlet*, and *Macbeth*. For each character who spoke more than 900 words, we averaged the percentage of their words that were self-references (I, me, myself, etc.), that had a non-negative valence value, and that were adjectives or adverbs (to capture descriptiveness).

The resulting percentages were then normalized for each play, forcing the ‘most truthful’ character’s score to 100 and the ‘most deceptive’ character’s to 0. The results are shown in Fig. 6.

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

Fig. 4. Above is the table showing the truthfulness rankings (as determined by Pennebaker’s features) for Shakespeare’s three most famous tragedies. The rankings were determined by averaging the percentage of self-references, non-negative words, and descriptors a character uses: the higher the average, the more truthful a character is. The results have been normalized, forcing the most truthful character’s score to 100 and the least truthful’s score to 0. The character rank-order aligns with popular interpretations of the three plays.

Pennebaker’s features produced an ordering that is consistent with readings of these tragedies. Iago, Emilia, Claudius, and Lady Macbeth all were determined to be the least truthful in their respective plays. Emilia may seem to be an outlier here, but actually she tells a significant lie about the origin of the handkerchief, the most influential piece of ‘evidence’ in convincing Othello of his wife’s infidelity. While these rankings alone are probably not enough to discriminate heroes from villains—as Horatio, who is certainly not a villain, is near the bottom of *Hamlet*’s rankings—there does seem to be some worth in the analysis. An alternate explanation for why the Pennebaker analysis seems to work is that the villains, at least in the case of *Hamlet* and *Macbeth*, are heads of state and thus would be inclined to use *we* (in reference to the state) in their speeches, which would lower their score.

IV. FUTURE WORK

We believe character-to-character sentiment analysis can, in its current state, be useful in education and to theorists’ search for new perspectives on a work. Yet to judge our work’s steps towards automatic, deep literary analysis or towards a use in automated, machine reading, more work must be done. Firstly, a shallow method for detecting subtle rhetorical features (irony, humor, deceit, etc.) must be better studied before standard sentiment analysis becomes more reliable. And furthermore, more experiments on more types of literature are needed. Our next step will be to subject novels to sentiment analysis by using a speech attribution method similar to Elson et al.’s [3].

V. CONCLUSIONS

As demonstrated, shallow sentiment analysis can be used with some success to recognize and analyze human relationships described within a document. Word valence is clearly linked to document genre, and rankings of enemies and allies that match reader expectations can be found through our character-to-character method. Despite some success, the sentiment networks extracted were not observed to be governed by Structural Balance Theory, as we hypothesized. One reason our sentiment networks are problematic is that deceit, an essential and pivotal element of many literary works, is hardly detectable through word lists and causes erroneous edge weights. Psychological models have made the most progress identifying suppressed emotions in words and provided some surprising insights into Shakespeare’s three famous tragedies. But we doubt that these models can scale and generalize well.

REFERENCES

- [1] A. Agarwal, A. Corvalan, J. Jensen, and O. Rambow, “Social network analysis of alice in wonderland,” *NAACL-HLT 2012*, p. 88, 2012.
- [2] D. Elson, N. Dames, and K. McKeown, “Extracting social networks from literary fiction,” in *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2010, pp. 138–147.
- [3] D. Elson and K. McKeown, “Automatic attribution of quoted speech in literary narrative,” in *Proceedings of AACL*, 2010.
- [4] G. Groh and J. Hauffa, “Characterizing social relations via nlp-based sentiment analysis,” in *Proceedings of the Fifth international aaii conference on Weblogs and Social media (icWSm-2011)*, 2011.
- [5] S.-M. Kim and E. Hovy, “Determining the sentiment of opinions,” in *Proceedings of the 20th international conference on Computational Linguistics*, ser. COLING ’04. Stroudsburg, PA, USA: Association for Computational Linguistics, 2004. [Online]. Available: <http://dx.doi.org.ezproxy.lib.lehigh.edu/10.3115/1220355.1220555>
- [6] J. Krauss, S. Nann, D. Simon, K. Fischbach, and P. Gloor, “Predicting movie success and academy awards through sentiment and social network analysis,” in *ECIS European Conference on Information Systems*, 2008.
- [7] R. Maranzato, A. Pereira, M. Neubert, and A. P. do Lago, “Fraud detection in reputation systems in e-markets using logistic regression and stepwise optimization,” *SIGAPP Appl. Comput. Rev.*, vol. 11, no. 1, pp. 14–26, Jun. 2010. [Online]. Available: <http://doi.acm.org.ezproxy.lib.lehigh.edu/10.1145/1869687.1869689>
- [8] S. Marvel, J. Kleinberg, R. Kleinberg, and S. Strogatz, “Continuous-time model of structural balance,” *Proceedings of the National Academy of Sciences*, vol. 108, no. 5, pp. 1771–1776, 2011.
- [9] S. Mohammad, “From once upon a time to happily ever after: Tracking emotions in novels and fairy tales,” in *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*. Association for Computational Linguistics, 2011, pp. 105–114.
- [10] P. Mutton, “Inferring and visualizing social networks on internet relay chat,” in *Information Visualisation, 2004. IV 2004. Proceedings. Eighth International Conference on*. IEEE, 2004, pp. 35–43.
- [11] E. Nalisnick and H. Baird, “Character-to-character sentiment analysis in shakespeare’s plays,” in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics: Short Papers*. Association for Computational Linguistics, 2013.
- [12] M. Newman, J. Pennebaker, D. Berry, and J. Richards, “Lying words: Predicting deception from linguistic styles,” *Personality and Social Psychology Bulletin*, vol. 29, no. 5, pp. 665–675, 2003.
- [13] F. Å. Nielsen, “Afinn,” Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby, March 2011. [Online]. Available: <http://www2.imm.dtu.dk/pubdb/p.php?6010>

- [14] A. Peñas and E. Hovy, "Filling knowledge gaps in text for machine reading," in *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, ser. COLING '10. Stroudsburg, PA, USA: Association for Computational Linguistics, 2010, pp. 979–987. [Online]. Available: <http://dl.acm.org.ezproxy.lib.lehigh.edu/citation.cfm?id=1944566.1944679>
- [15] B. Poblete, R. García, M. Mendoza, and A. Jaimes, "Do all birds tweet the same?: characterizing twitter around the world," in *Proceedings of the 20th ACM international conference on Information and knowledge management*, ser. CIKM '11. New York, NY, USA: ACM, 2011, pp. 1025–1030. [Online]. Available: <http://doi.acm.org.ezproxy.lib.lehigh.edu/10.1145/2063576.2063724>
- [16] T. Rommel, "so soft, so sweet, so delicately clear.a computer assisted analysis of accumulated words and phrases in lord byron's epic poem don juan," *Literary and linguistic computing*, vol. 9, no. 1, pp. 7–12, 1994.
- [17] A. Smith, "Automatic extraction of semantic networks from text using leximancer," in *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology: Demonstrations-Volume 4*. Association for Computational Linguistics, 2003, pp. 23–24.
- [18] C. Strapparava and A. Valitutti, "Wordnet-affect: an affective extension of wordnet," in *Proceedings of LREC*, vol. 4, 2004, pp. 1083–1086.
- [19] T. T. Thet, J.-C. Na, C. S. Khoo, and S. Shakthikumar, "Sentiment analysis of movie reviews on discussion boards using a linguistic approach," in *Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion*, ser. TSA '09. New York, NY, USA: ACM, 2009, pp. 81–84. [Online]. Available: <http://doi.acm.org.ezproxy.lib.lehigh.edu/10.1145/1651461.1651476>
- [20] I. Titov and R. McDonald, "Modeling online reviews with multi-grain topic models," in *Proceedings of the 17th international conference on World Wide Web*, ser. WWW '08. New York, NY, USA: ACM, 2008, pp. 111–120. [Online]. Available: <http://doi.acm.org.ezproxy.lib.lehigh.edu/10.1145/1367497.1367513>
- [21] K. Toutanova and C. Manning, "Enriching the knowledge sources used in a maximum entropy part-of-speech tagger," in *Proceedings of the 2000 Joint SIGDAT conference on Empirical methods in natural language processing and very large corpora: held in conjunction with the 38th Annual Meeting of the Association for Computational Linguistics-Volume 13*. Association for Computational Linguistics, 2000, pp. 63–70.
- [22] G. Williams, "The comedy of errors rescued from tragedy," *Review of English Literature*, vol. 5, pp. 63–71, 1964.
- [23] Y. Zhao and J. Zobel, "Searching with style: Authorship attribution in classic literature," in *Proceedings of the thirtieth Australasian conference on Computer science-Volume 62*. Australian Computer Society, Inc., 2007, pp. 59–68.