Beyond Canonical Texts: A Computational Analysis of Fanfiction

Smitha Milli
Computer Science Division
University of California, Berkeley
smilli@berkeley.edu

David Bamman
School of Information
University of California, Berkeley
dbamman@berkeley.edu

Abstract

While much computational work on fiction has focused on works in the literary canon, user-created fanfiction presents a unique opportunity to study an ecosystem of literary production and consumption, embodying qualities both of large-scale literary data (55 billion tokens) and also a social network (with over 2 million users). We present several empirical analyses of this data in order to illustrate the range of affordances it presents to research in NLP, computational social science and the digital humanities. We find that fanfiction deprioritizes main protagonists in comparison to canonical texts, has a statistically significant difference in attention allocated to female characters, and offers a framework for developing models of reader reactions to stories.

1 Introduction

The development of large-scale book collections—such as Project Gutenberg, Google Books, and the HathiTrust—has given rise to serious effort in the analysis and computational modeling of fiction (Mohammad, 2011; Elsner, 2012; Bamman et al., 2014; Jockers, 2015; Chaturvedi et al., 2015; Vala et al., 2015; Iyyer et al., 2016). Of necessity, this work often reasons over historical texts that have been in print for decades, and where the only relationship between the author and the readers is mediated by the text itself. In this work, we present a computational analysis of a genre that defines an alternative relationship, blending aspects of literary production, consumption, and communication in a single, vibrant ecosystem: fanfiction.

Fanfiction is fan-created fiction based on a previously existing, original work of literature. For clarity we will use the term CANON to refer to the original work on which a fanfiction story is based (e.g. Harry Potter) and the term STORY to refer to a single fan-authored story for some canon.

Although stories are based on an original canonical work and feature characters from the canon, fans frequently alter and reinterpret the canon—changing its setting, playing out an alternative ending, adding an original character, exploring a minor character more deeply, or modifying the relationships between characters (Barnes, 2015; Van Steenhuyse, 2011; Thomas, 2011).

In this work, we present an empirical analysis of this genre, and highlight several unique affordances this data presents for contemporary research in NLP, computational social science, and the digital humanities. Our work is the first to apply computational methods to fanfiction; in presenting this analysis, we hope to excite other work in this area.

2 Fanfiction data

Our data, collected between March–April 2016, originates from `fanfiction.net`.[1] In this data, AUTHORS publish stories serially (one chapter at a time); REVIEWERS comment on those chapters.

A summary of data is presented in table 1. The scale of this data is large for text; at 55 billion to-

[1] While terms of service prohibit our release of this data, tools to collect and process it can be found here: `http://github.com/smilli/fanfiction`.

Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pages 2048–2053,
Austin, Texas, November 1-5, 2016.

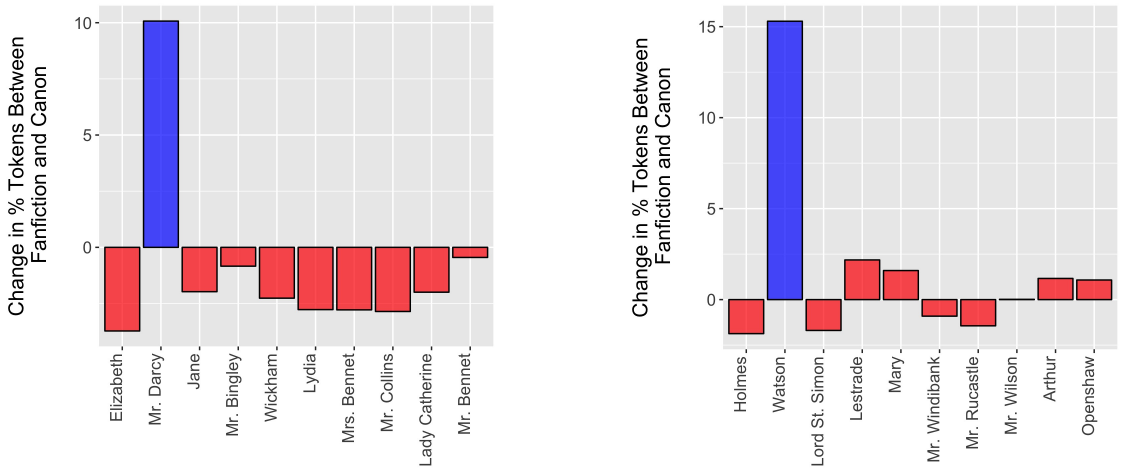

Figure 1: Difference in percent character mentions between fanfiction and canon for *Pride and Prejudice* (left) and *Sherlock Holmes* (right).

kens, it is over 50 times larger than the BookCorpus (Zhu et al., 2015) and over 10% the size of Google Books (at 468B tokens).

The dataset is predominantly written in English (88%), but also includes 317,011 stories in Spanish, 148,475 in French, 102,439 in Indonesian, and 73,575 in Portuguese. In total, 44 different languages are represented.

Type	Num of Type
Canons	9,246
Stories	5,983,038
Tokens	55,264,185,653
Reviews	159,914,877
Users	2,093,601
–Authors	1,364,729
–Reviewers	1,438,721
Languages	44

Table 1: Summary of the fanfiction.net corpus

3 Analysis of fanfiction

3.1 Differences between canons and fanfiction

The systematic ways in which fanfiction stories differ from their canonical works can give insight into the characteristics of a story that are desired by fans but may be missing from the mainstream canon. We investigate two questions: 1.) Is there a difference between the characters emphasized in fanfiction compared to the original canon? And 2.) Is gender presented differently in these two sources?

Character differences. In order to explore the differing attention to character, we consider fanfiction from ten canons whose original texts appear in Project Gutenberg; we selected the ten canons from unique authors with the most fanfiction stories associated with them.[2] To extract and compare characters, we run BookNLP (Bamman et al., 2014) on both the canonical work and the top 100 stories (by number of reviews) from each canon, and pair characters across canon/fanfiction with the same name.

To measure how the importance of individual characters varies between canon and fanfiction, we calculate the change in the percent of all character mentions a character has in the canon to the average percent of character mentions that same character has in fanfiction.

Across all canons we find that the most prominent character in the canon had at most a small increase in percent character mentions, while less prominent characters received large increases. The results for two illustrative examples, *Pride and Prejudice* and *Sherlock Holmes*, are shown in Figure 1. The percent of character mentions for the main protagonists (Elizabeth and Holmes) decreases in fanfiction, but the secondary characters of Mr. Darcy and Watson

[2] *Les Miserables* (3996 fanfiction stories), *Sherlock Holmes* (3283), *Pride and Prejudice* (3084), *Peter Pan* (2431), *Alice in Wonderland* (1446), *Anne of Green Gables* (620), *Jane Eyre* (331), *Little Women* (286), *The Scarlet Pimpernel* (255), and the *Secret Garden* (222).

Labels	Terms
Author encouragement	read story one reading chapters time best ever review long update please love soon story amazing really hope continue writing chapter great good keep really work story job forward awesome ca wait next chapter see na happens gon great read like well really story love chapter way one see interesting
Requests for story	would like know get think going could something really even
Emotional reactions	wow better beautiful getting fight adorable keeps team birthday tears oh god yes man yay damn hell dear yeah got poor lol cute howl evil bad hate baby feel lord xd loved funny love haha sweet lol ah cute aww

Table 2: Top 10 terms in the 10 manually grouped LDA topics.

show a large increase. These findings confirm the results of Xu et al. (2011), who find a greater increase in mentions of Mr. Darcy relative to Elizabeth in a different corpus of *Pride and Prejudice* fanfiction, and supports claims that fanfiction authors may delve deeper into characters that receive less attention in the canon (Jwa, 2012; Thomas, 2011).

Gender differences. Fanfiction has a predominantly female authorship and readership base (Barnes, 2015); these stories often oppose traditional gender norms present in the canon and showcase stronger female characters (Handley, 2012; Scodari and Felder, 2000; Leow, 2011; Busse, 2009).

In order to test whether fanfiction allocates more attention to female characters than canonical stories, we compare the percent of mentions of male and female characters using the same collection of stories from Gutenberg canons as above. 40.1% of character mentions in the canons are to women; in fanfiction, this ratio increases to 42.4%. This effect size is small (2.3% absolute difference), but in a bootstrap hypothesis test of the difference (using 10^6 bootstrapped samples), the difference is highly significant ($p < 0.001$), suggesting that fanfiction does indeed devote more attention to female characters.

3.2 Interaction between users

A unique characteristic of this data is the chapter-by-chapter reader reviews; any user can leave a review for a chapter of a story. Authors are also frequently reviewers of other stories (Thomas, 2011), forming an ecosystem with qualities of a social network.

709,849 authors in this data (52%) are also reviewers; if we define a network node to be a user and edge to be a reviewing relationship (a directed edge exists from $A \rightarrow B$ if A reviews one of B's works), this network contains 9.3M such directed edges, with an average outdegree of 13.2 and indegree of 15.6 (each author on average reviews for 13 other authors, and is reviewed by 16).

To explore the content of these reviews computationally, we sampled one story with more than 500 reviews from 500 different canons and ran Latent Dirichlet Allocation (Blei et al., 2003) with 10 topics on the text of the reviews (excluding names as stopwords). This is an exploratory analysis, but can give insight into the broad functions of reader responses in this domain.

Table 2 presents the results, grouping the topics into three exploratory categories: positive encouragement and pleas for updates, requests to the author about the progression of the story, and emotional reactions to the story. Prior studies that have examined the role of reviews as a form of interaction between the reader and the author have documented the first two categories extensively (Campbell et al., 2016; Magnifico et al., 2015; Lammers and Marsh, 2015; Black, 2006). However, despite a significant portion of the reviews consisting of the reader's emotional responses to the story, the way in which readers use reviews as a means of emotional engagement with the story itself has yet to be examined in such detail.

3.3 Predicting reader responses to character

The presence of reader reviews accompanying each fanfiction chapter presents a unique opportunity to develop a predictive model of how readers respond to text—given the text of a chapter, can we predict

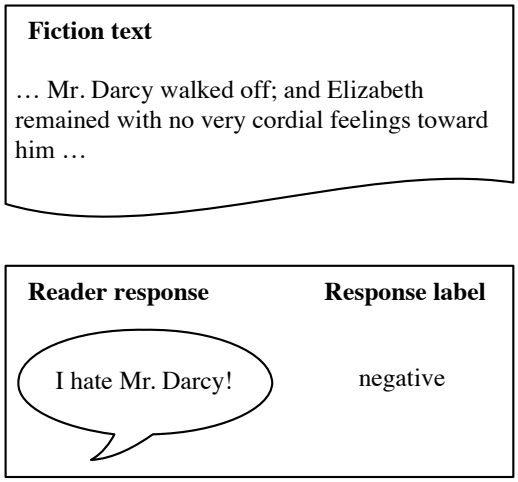

Figure 2: Illustration of data for predicting reader responses. Here we are using features derived only from FICTION TEXT to predict the RESPONSE LABEL.

how readers will react?

To test the feasibility of this task, we focus on reader responses to character. A RESPONSE to a character is operationalized as a sentence from a reader review mentioning that character and no other characters. We create an annotated dataset by randomly sampling a single character with at least 10 reader responses from each of the 500 stories described in §3.2. From this set, we randomly select exactly 10 responses for each character, to yield a total of 5,000 reader responses.

We then present these 5,000 responses to annotators on Amazon Mechanical Turk, and ask them to judge the sentiment *toward the character* expressed in the response as either positive, negative, neutral, or not applicable (in the case of character recognition errors). The overall agreement rate among annotators for this classification task is moderate (Fleiss' κ = 0.438), in part due to the difficulty of assessing the responders' attitudes from short text; while some responses wear their sentiment on their sleeve (*I knew I hated Brandon!*), others require more contextual knowledge to judge (*Ah Link* or *Akira appears!*).

In order to create a higher-precision subset we select responses with only unanimous positive or negative votes from 3 different annotators, yielding a total dataset of 1,069 response labels. We divide the dataset into 80% training/development and 20% for a held-out test (with no overlap in stories between training and test).

We also bootstrap additional semi-supervised data by training a sentiment classifier on the unigrams of the reader responses in the training data (with a 3-class accuracy of 75%; compared to majority baseline of 49.7%), predicting the sentiment label for all responses in the dataset, and selecting examples that a.) have 95% prediction confidence and b.) whose stories do not appear in the training or test data. We sample selected examples to respect the label distribution in the training data, yielding an additional 25,000 data points to supplement learning.

Our core task is to use *only* the text of the story (and not the reader response) to predict the corresponding response sentiment label in order to understand what aspects of the story (and a character's role within it) readers are reacting to. We experiment with several features to represent the characters:

- AGENT, PATIENT, PREDICATIVE, POSSESSIVE relations for each character (as output by BookNLP), both in the specific chapter and in the book overall (under the rationale that readers are responding to the *actions* that characters take).
- Unigrams spoken by the character, both in the chapter and in the book overall.
- Character gender.
- Character's frequency in the book (binary indicators of the decile of their frequency of mention).
- Skip-gram representations trained on 4.1B words of fanfiction text (200-dimensional, grouped into 1000 clusters using k-means clustering); these cluster identities form additional binary features by replacing the lexical indicators for the text features above.

We perform model selection using tenfold cross-validation on the training data alone, training ℓ_2-regularized logistic regression on one partition of the data, tuning the ℓ_2 regularization parameter on another, and assessing performance on a third (note none of the test data described above is seen during this stage).

A majority class (all-positive) baseline on the training data results in an accuracy rate of 75.6%;

only syntactic features yield a significant improvement, achieving an accuracy rate of 80.5%.

Table 3 lists the most informative features for predicting negatively-assessed characters. While these characteristic features have face validity and offer promise for understanding character in more depth, we do not see similar improvements in accuracy on the truly held-out test data (run once before submission); this same feature set achieves an accuracy rate of 70.4% (compared to a majority baseline on the test data of 71.4%). Part of this may be due to the sample size of the test data ($n = 199$); a bootstrap 95% confidence interval (Berg-Kirkpatrick et al., 2012) places the accuracy in the range [0.648, 0.754]. However, this more likely constitutes a negative result that reflects the inherent difficulty of the task; while syntactic features point to a real signal that readers are reacting to when writing their responses, literary character is of course far more complex, and more sophisticated representations of character—and of the readers who react to them—are likely warranted for real predictive performance on this task.

agent	patient	predicative	possessive
hissed	hate	pregnant	phone
sneered	done	human	state
shoved	see	afraid	tone
glared	hated	stubborn	face
paused	asked	person	spot
respond	face	boy	plan
caught	pissed	angry	wand
scowled	blame	stupid	pain
walked	shocked	free	emotions
had	used	mother	chakra

Table 3: Syntactic features most predictive of a negatively-assessed character.

4 Conclusion

In blending aspects of large-scale literary data and social network structure, fanfiction publication constitutes a vibrant ecosystem with ample textual evidence for the production and consumption of literary texts. In this work, we have briefly illustrated three aspects of this data that have the potential to yield interesting insight—the relationship between fanfiction stories and their original source material, the social network structure of authors and their respondents, and the possibility of predicting reader responses from serially published text. Many questions remain; in providing a quantitative description of this dataset, we hope to highlight its potential for analysis, and encourage other work in this domain.

Acknowledgments

We thank Cecilia Aragon and our anonymous reviewers for their helpful comments. This work is made possible by the use of computing resources from Berkeley Research Computing.

References

David Bamman, Ted Underwood, and Noah A. Smith. 2014. A Bayesian mixed effects model of literary character. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 370–379, Baltimore, Maryland, June. Association for Computational Linguistics.

Jennifer L Barnes. 2015. Fanfiction as imaginary play: What fan-written stories can tell us about the cognitive science of fiction. *Poetics*, 48:69–82.

Taylor Berg-Kirkpatrick, David Burkett, and Dan Klein. 2012. An empirical investigation of statistical significance in NLP. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, EMNLP-CoNLL '12, pages 995–1005, Stroudsburg, PA, USA. Association for Computational Linguistics.

RW Black. 2006. Not just the OMG standard: reader feedback and language, literacy, and culture in online fanfiction. In *Annual Meeting of The American Educational Research Association, San Francisco*, volume 10.

David M. Blei, Andrew Ng, and Michael Jordan. 2003. Latent dirichlet allocation. *JMLR*, 3:993–1022.

Kristina Busse. 2009. In focus: Fandom and feminism: gender and the politics of fan production. *Cinema Journal*, 48(4):104–108.

Julie Campbell, Cecilia Aragon, Katie Davis, Sarah Evans, Abigail Evans, and David Randall. 2016. Thousands of positive reviews: Distributed mentoring in online fan communities. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, CSCW '16, pages 691–704, New York, NY, USA. ACM.

Snigdha Chaturvedi, Shashank Srivastava, Hal Daume, and Chris Dyer. 2015. Modeling dynamic relationships between characters in literary novels.

Micha Elsner. 2012. Character-based kernels for novelistic plot structure. In *EACL*.

Christine Handley. 2012. Distressing damsels: narrative critique and reinterpretation in star wars fanfiction. *Fan Culture: Theory/Practice, Newcastle upon Tyne: Cambridge Scholars Publishing*, pages 97–118.

Mohit Iyyer, Anupam Guha, Snigdha Chaturvedi, Jordan Boyd-Graber, and Hal Daumé III. 2016. Feuding families and former friends: Unsupervised learning for dynamic fictional relationships. In *North American Association for Computational Linguistics*.

Matthew Jockers. 2015. Revealing sentiment and plot arcs with the syuzhet package. `http://www.matthewjockers.net/2015/02/02/syuzhet/`.

Soomin Jwa. 2012. Modeling L2 writer voice: Discoursal positioning in fanfiction writing. *Computers and Composition*, 29(4):323–340.

Jayne C Lammers and Valerie L Marsh. 2015. Going public: An adolescent's networked writing on fanfiction.net. *Journal of Adolescent & Adult Literacy*, 59(3):277–285.

Hui Min Annabeth Leow. 2011. Subverting the canon in feminist fan fiction. *Transformative Works and Cultures*, 7.

Alecia Marie Magnifico, Jen Scott Curwood, and Jayne C Lammers. 2015. Words on the screen: broadening analyses of interactions among fanfiction writers and reviewers. *Literacy*, 49(3):158–166.

Saif Mohammad. 2011. From once upon a time to happily ever after: Tracking emotions in novels and fairy tales. *CoRR*, abs/1309.5909.

Christine Scodari and Jenna L Felder. 2000. Creating a pocket universe: Shippers, fan fiction, and the X-Files online. *Communication Studies*, 51(3):238–257.

Bronwen Thomas. 2011. What is fanfiction and why are people saying such nice things about it? *Storyworlds: A Journal of Narrative Studies*, 3(1):1–24.

Hardik Vala, David Jurgens, Andrew Piper, and Derek Ruths. 2015. Mr. Bennet, his coachman, and the Archbishop walk into a bar but only one of them gets recognized: On the difficulty of detecting characters in literary texts. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 769–774, Lisbon, Portugal, September. Association for Computational Linguistics.

Veerle Van Steenhuyse. 2011. The writing and reading of fan fiction and transformation theory. *CLCWeb: Comparative Literature and Culture*, 13(4):4.

Jun Xu. 2011. Austen's fans and fans' Austen. *Journal of Literary Semantics*, 40(1):81–97.

Yukun Zhu, Ryan Kiros, Richard S. Zemel, Ruslan Salakhutdinov, Raquel Urtasun, Antonio Torralba, and Sanja Fidler. 2015. Aligning books and movies: Towards story-like visual explanations by watching movies and reading books. *CoRR*, abs/1506.06724.