Supporting Information for

The World of Harry Potter: How well do the Films Represent the Books?

Helga Þórey Björnsdóttir (s213615), Rebekka Jóhannsdóttir (s212963), Sara Húnfjörð Jósepsdóttir (s212952)

This PDF file includes:

Supporting Information Text

Here we present our main result for the Harry Potter character network generated from character pages in the wiki hosting site *Fandom* (1), in addition to term-frequency inverse-document-frequency (TF-IDF) word clouds generate from the film scripts and books. Additionally, the results from the analysis of female and male character sentiment distribution are presented here. This appendix template was provided by *The Proceedings of the National Academy of Sciences* (PNAS).

Main result of the Harry Potter character network

The Harry Potter character network consisted of 462 character nodes and 7108 connections with an average degree $\langle k \rangle = 30.77$ (S1). The results confirmed that it was a small-world and scale-free network with power-law degree distribution. The two sub-networks, one for characters appearing in the films and another one of characters appearing in the books, also possessed these properties. We also present a visualization of Bellatrix Lestrange's connections in Figure S3.

Word clouds generated for each book and film

Word clouds were generated by calculating the term frequency-inverse document frequency for each book and film to see which are the most important words in the series. Larger words indicate higher importance, see Figures S4 and S5.

Sentiment Distribution for each gender

For the purpose of illustrating the sentimental differences between male and female characters in films and books, a sentiment distribution was generated for each gender. We thought it would be interesting to see if female characters are portrayed as happier than male characters, as they are in so many other films. See Figures S6 and S7

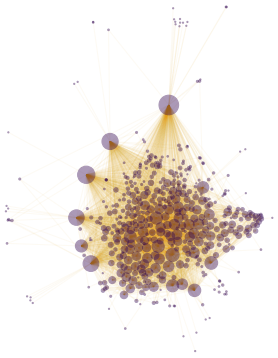

Fig. S1. A visualisation of the Harry Potter character network. The yellow lines represent the links between the characters (purple nodes). The network forms a large hub, a consequence of the scale-free property of the network (2).

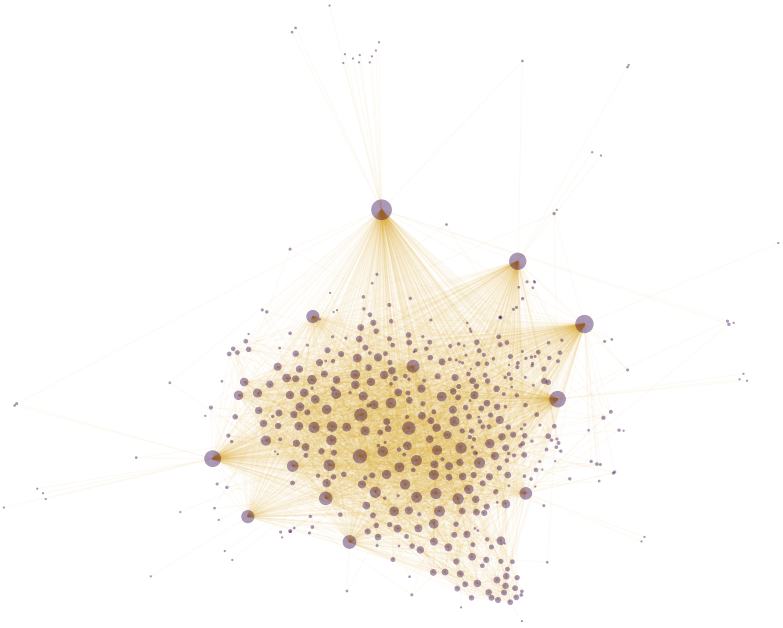

Fig. S2. The giant connected component of the Harry Potter Network. The component is shown for better visualisation of the network. The average shortest-path length being very small is clearly visible in this graph.

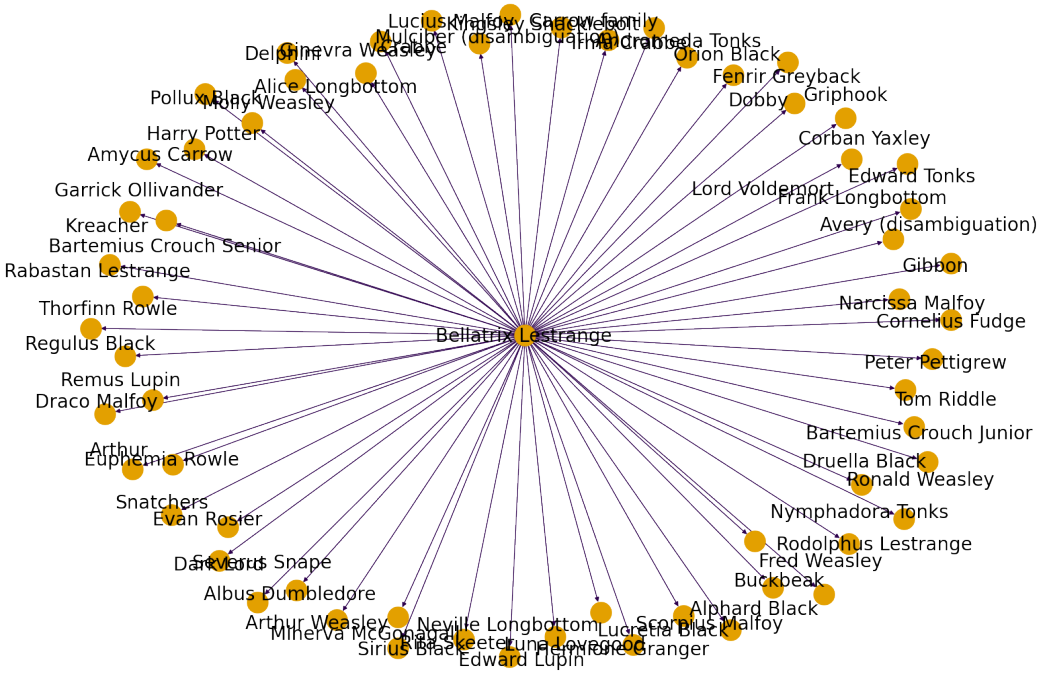

Fig. S3. Bellatrix Lestrange's connections in a network. She is connected to major characters such as Harry Potter, Lord Voldemort and Albus Dumbledore. An unexpected connection is, for example Fred Weasley as there is not a direct connection between them. However, Fred was killed by the Death Eaters and Bellatrix's last word before her death was his first name, "Freddie" as she was taunting his mother after his death (1).

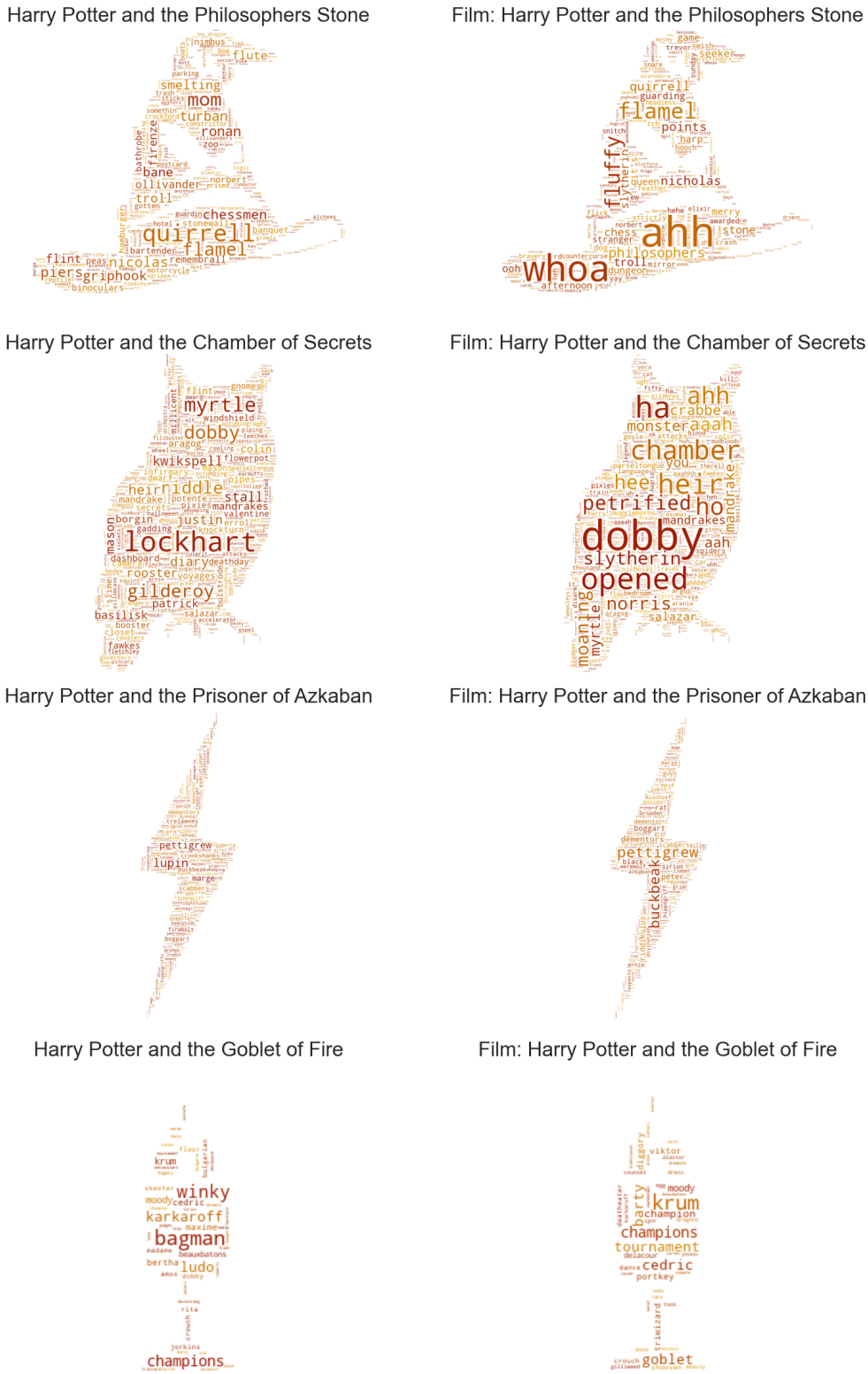

Fig. S4. Word clouds for book/film 1 to book/film 4 calculated with TF-IDF. Larger words on the word cloud indicate higher TF-IDF values and are therefore more important. We observe major differences in some of the word cloud pairs, for example, we see a lot of exclamation words such as "ahh" and "woah" in the first film. The first book/film word clouds both contain Flamel, the maker of the philosopher's stone and Quirrel, who was one of the villains of the first book. In the second film, we see a large importance on the word "Dobby" which is not very large in the book. The same goes for Buckbeak in film number 2.

Harry Potter and the Order of the Phoenix

Film: Harry Potter and the Order of the Phoenix

Harry Potter and the Half Blood Prince

Film: Harry Potter and the Half Blood Prince

Harry Potter and the Deathly Hallows

Film: Harry Potter and the Deathly Hallows Part 1

Film: Harry Potter and the Deathly Hallows Part 2

Fig. S5. Word clouds for books 6 and 7 and films 6 to 8. Larger words on the word cloud indicate higher TF-IDF values and therefore more are important. Book and film number 6 share similar vocabularies. We observe a lot of dissimilarities in the final films and books. The films seem to put large emphasis on Dobby and Bathilda which are very small in the book word cloud.

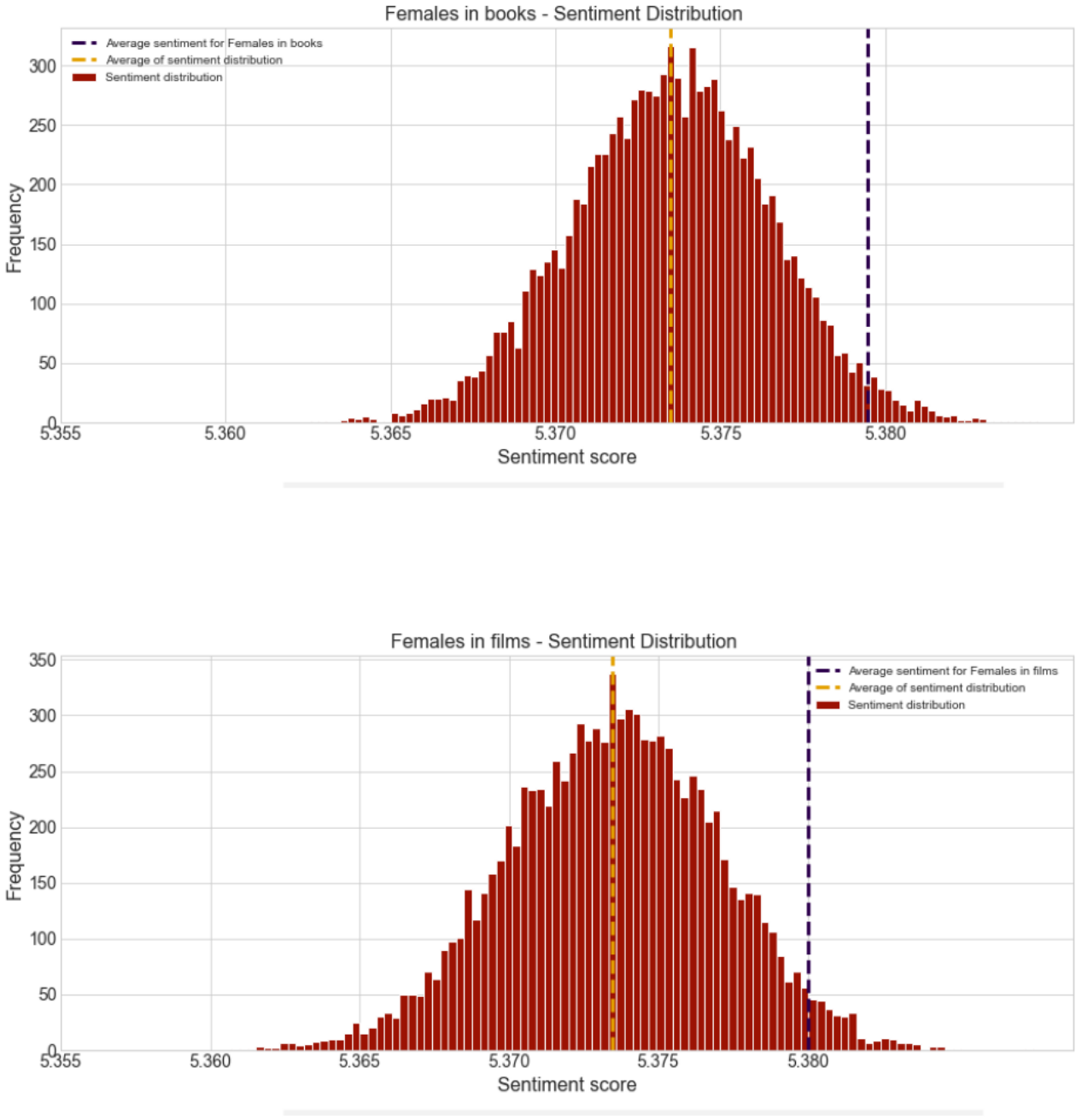

Fig. S6. Sentiment Distribution for female characters in the books and films separately. The figure shows the average sentiment for all female characters for films and books in purple dashed line, the distribution of same sized set of randomly selected characters and the distribution average. This shows that the overall sentiment remains nearly unchanged: It is, however, slightly lower in films, indicating that a few minor characters from the books were left out in the films.

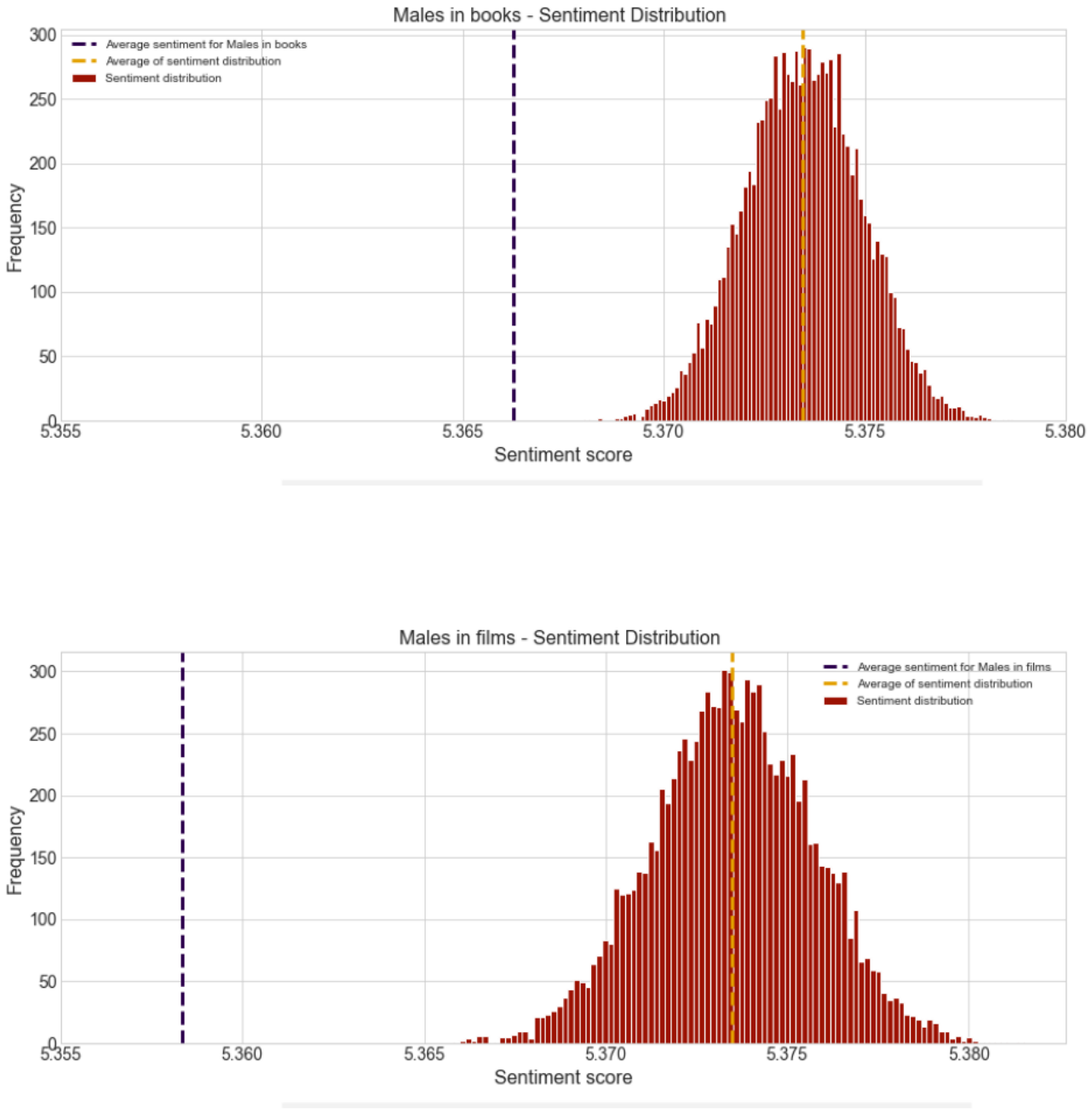

Fig. S7. Sentiment Distribution for male characters in the books and films separately. The figure shows the average sentiment for all male characters for films and books in a purple dashed line, the distribution of a same sized set of randomly selected characters, and the distribution average. The overall sentiment changes quite a bit between the films. In the studies it was discovered that the major characters tend to have lower sentiment than minor ones. This shows that many minor male characters were left out of the films

References

1. Harry potter wiki (2022).
2. AL Barabási, M Pósfai, *Network science*. (Cambridge University Press, Cambridge), (2016).