Social Network Analysis and Mining (2022) 12:42
https://doi.org/10.1007/s13278-021-00854-9

ORIGINAL ARTICLE

Quantitative analysis of fanfictions' popularity

Zhivar Sourati Hassan Zadeh[1] · **Nazanin Sabri**[1] · **Houmaan Chamani**[1] · **Behnam Bahrak**[1]

Received: 25 July 2021 / Revised: 25 December 2021 / Accepted: 27 December 2021 / Published online: 7 March 2022

Abstract
Considering the social origin of fanfictions and the fact that these pieces of literature are under the direct influence of society's line of thought, they can be a valuable resource for studying what the common audience finds attractive and what captures their attention. The aim of this study is to find properties of the fanfictions that may contribute to their popularity among community members, including properties that are in contrast to their original stories and those that exclusively belong to the fanfictions themselves. For this purpose, we investigated different features of fanfictions, including their emotional arc and character graphs representing the characters' interactions in the story. Our results show that the similarity between fanfictions and their associated original story in terms of emotional arcs and character graphs significantly relates to fanfictions' popularity.

Keywords Fanfiction · Character graph · Emotional arc · Fanfictions' popularity

1 Introduction

The advent of the internet has allowed many devoted fans to write and publish continuations or adaptations of their favorite books online. These stories, referred to as fanfiction, have been portrayed in many different lights by the media. Additionally, fanfiction as a certain kind of written narrative is a "root metaphor" for psychology described by Sarbin (1986), and there is a significant space for extending the psychological studies on it. With some claiming they have great potential Burt (2017), while others say it is something to be feared Grady (2016). Fanfictions can be a place where the typical stereotypes are shaded. Fans can represent themselves in their stories, and fanfiction gives a voice to the people who are normally marginalized by society and positively affect their lives. The fanfiction community also gives amateur writers a space to practice writing and get immediate feedback Campbell et al. (2016). But how different are these stories from the original books that inspired them in the first place, and what might affect their popularity? At first, this question might not be worth pursuing as people may believe that there are no real-life impacts related to the topic. However, if we shift toward counterfactuals and appreciate that the fanfictions are presenting us what would've happened if the story was different, the importance of this topic would be more conspicuous. By analyzing fanfictions and their various properties, we are actually having access to different derivations of the main story. That said, investigating a story would be much easier as we can look at the other available versions of the story and see what their fans want to change about the original story. This could be even beneficial for future stories or personalized variations of a story. With the distributed mentoring that Campbell et al. (2016) attends to which the fanfiction community offers, the importance of this topic is intensified even more.

Two of the main approaches used in this study are emotional arcs and the graph of interactions between characters. The emotional arcs can be obtained by natural language processing (NLP) techniques and represent the reader-perceived emotional content of written stories as the reader progresses through the book. The emotional arc of a story does not give us information about the intended

✉ Zhivar Sourati Hassan Zadeh
zh.sourati@ut.ac.ir

Nazanin Sabri
nazanin.sabri@ut.ac.ir

Houmaan Chamani
houmaan.chamani@ut.ac.ir

Behnam Bahrak
bahrak@ut.ac.ir

[1] School of Electrical and Computer Engineering, College of Engineering, University of Tehran, Tehran, Iran

meaning of its content or plot, but it provides information about the overall trends of the story.

Another tool for delving deeper into the information that can be extracted from fanfictions is the graph of interactions between characters which can help us develop a better understanding of the level to which characters in the story are in contact with each other. Like the emotional arcs, having only the graph of characters does not reveal much information about the story's content but gives us a quantifiable basis for comparing stories.

As Reagan et al. (2016) puts forward, a cardinal part of a narrative in terms of its psychological aspects is the emotional experience that is evoked when people are reading it. They even help us get a broader understanding of the stories Hipson and Mohammad (2021). Plus, the characters' relations are one of the components of a narrative that make it memorable. We can mention thousands of examples, such as the friendship between Harry, Hermione, and Ron, or Sherlock Holmes and Dr. Watson. But if we go a bit further, we can conclude that the experience which is evoked is based on a story resonating with ourselves or our previous experiences as Kurt (2010) describes. Apart from that, one of the main reasons that fanfictions are prevailed being people want to change or replace certain parts of the original stories Thomas (2011); there are so many aspects to stories that people may target for this alternation. We opted to use two of the main components of the stories, namely the emotional flow and the characters' relations, to see if a deviation from the original stories in terms of these two aspects contributes to the popularity of fanfictions. Additionally, focusing on certain properties of fanfictions apart from their original stories, the length of the fanfictions and their genres were considered auxiliary factors affecting fanfictions' popularity. The contribution of this study to the existing literature would be to answer the following research questions:

- Does the extent to which the emotional arc of the fanfictions is similar to their associated original story relate to the level of popularity of fanfictions?
- Does the similarity of a fanfiction's characters graph to the original story's graph relate to the popularity of the fanfiction?
- Does the length of fanfictions play a role in determining their popularity?
- Are the genres that fanfictions fall under them affects their popularity?

The rest of this paper is structured as follows: Sect. 2 provides a brief overview of the related work. In Sect. 3 the data used in this dataset are described, and its characteristics are explained. Our methods and results are discussed in Sects. 4 and 5, respectively. Finally, we conclude our study in Sect. 6.

2 Related work

Milli and Bamman (2016) is one of the first computational studies conducted on fanfictions. The authors investigate how the main protagonists are deprioritized in these literary texts, and instead, more attention is placed on the secondary characters of the original stories. Building on the idea of this study, Büchler et al. (2018) explores how much of the original style of the Harry Potter novels is retained in both the movies and fanfictions. Frens et al. (2018) addresses the importance of distributed mentoring (in the form of reviews) in online writing communities (fanfiction.net) for the growth of young authors and finds that the number of reviews an author receives predicts increased lexical diversity in the chapters they have written. Additionally, changes in reading habits have been noticed. For instance, Pianzola et al. (2020) studies the stories and comments on Wattpad and points out the rise of a global reading culture in youth which is different from national preferences in certain topics and genres.

With consideration of popularity and success, a dataset of Italian fanfiction is introduced in Mattei et al. (2020). Using this dataset, the authors also discuss what writing styles are more successful (defined by the number of reviews they receive). They report a variety of linguistic features and their distributions that could lead to a successful fanfiction. Carvallo and Parra (2020) tries to discover which factors explain the popularity of the authors, and they find that the size of the authors' biography has a negative effect on the popularity of the author. In Zhao (2016), Zhao investigates the influence of short summaries on readership and popularity of fan-written stories by analyzing the future attractiveness of short summarizations with classification methods. Moreover, the effect of lexical richness, average sentence and paragraph lengths on the popularity of fanfictions are examined in Rubin and Girouard (2014).

In contrast to these articles, we investigate the popularity of fanfictions in the light of novel computational tools such as emotional arcs and character interaction graphs. Both the emotional arcs and the graph of characters' relations have been used in different settings to derive various conclusions regardless of the media. Different data structures such as audio and text or even video are being used to apply these tools and extract information. A number of studies such as Chu and Roy (2017) have attended to the subject that the engagement of the consumers or even the degree to which people favor one story over the other can be predicted by the emotional arcs of that story. With a slightly different approach, Berger (2021) investigates the effect of sentiment volatility (period-to-period shifts in sentiment) on the success of the stories. Even further, Del

Vecchio et al. (2021) uses the emotional arcs alongside an econometric model to predict the overall success parameters of the movies, including box office revenues. As we can see, there is much attention on the study of the popularity and evoked experience, focusing on the emotional arc of the stories. Wang et al. (2020) being the closest one to ours studying the popularity of fictions in terms of their emotional arcs, none of the already available studies, to the extent of our knowledge, have attended the similarity of fictions to their original ones in terms of the emotional flow of the stories.

Apart from the studies of emotional arc, the characters' relations as a social network can give us the ability to study the fanfictions from a more social perspective as Labatut and Bost (2019) provide us with different approaches and different interpretations. The literature also contains several studies on the modeling of these networks to dive deeper into their latent structures. One example is Aires et al. (2017) that investigates the similar structures that are observed in popular movies in terms of their characters' relations networks. Moreover, trying to predict the popularity of the movies based on their ratings, Kounelis et al. (2021) leverages character graph embeddings with a slightly different approach. Similar to emotional arcs, we can see that there is a rich foundation for modeling and studying the characters' relations network, especially in movies; however, the study of fanfictions and their popularity considering their similarity to the original stories in terms of the characters' relations network is not yet fully discovered.

3 Dataset

As previously mentioned, in this study, we aim to analyze the relationship between fanfictions and their original story. To do so, we use the website `fanfiction.net` as our source for fan-written stories. Due to the enormous collection the platform holds, we chose five of the most popular original series as of January 2021, namely, (1) Harry Potter, (2) Twilight, (3) Percy Jackson and the Olympians, (4) Lord of the rings and (5) Hunger Games. These series, in the same order they were mentioned, are the top five series in terms of number of fanfictions available online and account for 78% of all of them. For each story, the text and the popularity/interaction metrics (number of likes, number of followers, number of reviews, number of words and their genres) are gathered.

The number of collected stories from each book series is shown in Table 1. We can see that several thousand stories are collected for each series; however, after collecting the fanfictions, we placed a limit of ten thousand words on the minimum length of stories (referred to as *Long Stories* in the table), focusing only on stories that are long enough for making meaningful emotional arcs and characters graphs.

Table 1 Number of fanfiction stories in the dataset

Book series	# Stories	# Long stories
Harry Potter	6295	2344
Twilight	5378	2009
Percy Jackson and the Olympians	2511	822
Lord of the rings	5378	1504
Hunger games	9263	1687

Table 2 Number of fanfiction stories belonging to different languages in the dataset

Language	Number of Fanfictions
English	7178
French	551
German	42
Spanish	595

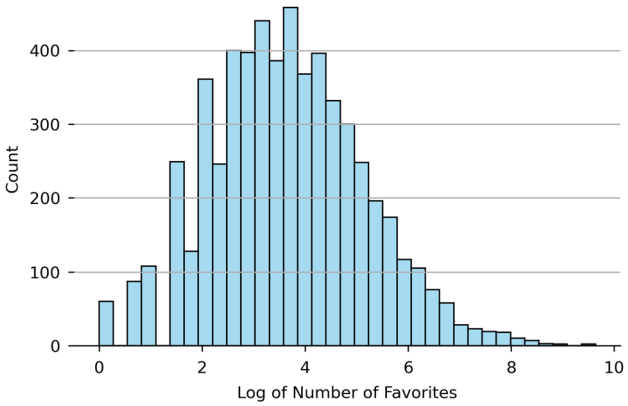

Fig. 1 The dstribution of number of favorites

Fanfiction stories in our dataset are available in different languages. Table 2 displays how many stories are available in each language after the elimination phase in terms of their length (languages are detected using the Python Langdetect library[1]). We can see that 85% of the collected stories are written in English. Emotional arcs can be studied in all the stories, as the dataset we are using to create emotional arcs allows us to do so. The emotional arc creation process is described in Sect. 4.

As previously mentioned, popularity/interaction metrics are collected for each story. These metrics include the number of people who reviewed the story, the number of

[1] https://pypi.org/project/langdetect/.

Table 3 Number of fanfictions for each series analyzed with at least 100 favorites and reviews

Series	Number of fanfictions
Harry Potter	80
The Hunger Games	73
The Lord of the Rings	70
Percy Jackson and the Olympians	72
Twilight	81

Table 4 Number of reviews, favorites and follows of fanfictions in the dataset

Property	Min	Mean	Max	Median
Reviews	1	54	27,205	9
Favorites	1	57	15,365	12
Follows	1	63	17,157	10

Table 5 Number of fanfictions falling under each genre

Genre	Number of fanfictions
Adventure	1848
Crime	25
Drama	1408
Family	786
Fantasy	469
Horror	68
Mystery	198
Romance	3475
Western	4
Without Genre	85

people who liked it (favorites), and the number of people who followed the story. The statistics of these metrics are presented in Table 4. We also extract the genres for each story. The numbers of fanfictions in the first five genres are shown in Table 5. Moreover, the distribution of the number of favorites for every fanfiction is demonstrated in Fig. 1.

As the process of analyzing the characters graphs (see Sect. 4.2) is computationally intensive, we reduced the number of fanfictions to be analyzed in Sect. 4.2 by applying a filter on the number of favorites and reviews a story has (only including those with at least 100 favorites and 100 reviews). In total, 376 of the available stories matched this criteria Table 3.

Lastly, for the purpose of comparing results with the original books, the text of all the books included in each of the series is also gathered.

4 Methods

In this section, we explain the methods used in this study in detail.

4.1 Emotional arcs

Stories all have specific shapes in terms of their flow of the emotional senses. The flow of stories can help us categorize them into distinguishable categories, which gives us a better understanding of their structure. The idea of the shapes of stories was first introduced by Kurt Vonnegut's rejected master's thesis[2]. He categorized stories based on the flow of their happiness into eight groups. Reagan et al. (2016) introduced a computational method for the same task, grouping stories into six main categories. Some stories start happily and end sadly, while others may start with a huge loss and then reach the climax of happiness in several moments. Other ones are somehow a mixture of these two classes. We followed the computational method of Reagan et al. (2016) for extracting emotional arcs, but to compare the extracted arcs from original stories and their fan-written counterparts. In what follows, we explain the process of extracting emotional arcs from fanfictions and clarify the process of calculating the similarity between fanfictions and original stories in terms of their emotional arcs.

4.1.1 Emotional arc extraction

To extract emotional arcs from fanfictions, we used Hedonometer words dataset[3]Dodds et al. (2014), which provides us with each word's happiness score in several different languages. To compute the happiness scores and their flow, we used a window-based method and extracted the words covered by a window of a specific size on the text and calculated the happiness score of that part of the text. In each step, the window slides forward, and the process is repeated till it reaches the end of the text. We chose to go with ten thousand words as the window size which is suggested by Dodds et al. (2011) for extracting emotional arcs. Note that choosing a small window size would make the emotional arcs sensitive to all the fluctuations, and a larger window size would not be able to capture essential nuances.

Two approaches for computing each section's happiness score were considered. The first method works as follows: a set containing all the words in that section is generated. The score of that specific section is calculated as the average happiness score of the words of that section, with words not having a happiness score excluded. The second method is

[2] https://www.youtube.com/watch?v=oP3c1h8v2ZQ.

[3] https://Hedonometer.org/timeseries/en_all/.

similar to the first, with the difference that the frequency of words was taken into account. Utilizing the first approach, we assume that a happy paragraph should contain more unique words with higher happiness levels instead of having repetitive happy words. Note that this assumption works perfectly for well-written stories but not for the chats and dialogues used in real-life scenarios.

To be able to create the windows and perform the aforementioned calculations, the following steps were taken. At first, books in PDF format were converted to text files by using the Fitz library[4]. The first and last parts of the books, which are mostly acknowledgments and unrelated parts to the story itself, were excluded.

Some pre-processing was done to make the text ready to use the Hedonometer dataset and get the best output. Natural Language Toolkit (NLTK) Bird et al. (2009) was used as the tokenizer. All the numbers and non-alphabetic characters (including punctuations) were dropped as they are neutral words and are not associated with any emotions. Similarly, words with a size less than two were removed as they are either propositions or leftovers after tokenizing (e.g., 's' after the apostrophe) and hold no valuable information during the extraction of the emotional arcs. Finally, all tokenized words were lower-cased, which was in response to all words being lower-cased in the Hedonometer dataset.

4.1.2 Emotional arcs similarity detection

Our method for analyzing fanfictions' similarities to the original books in terms of their emotional arcs is discussed in this section. For each pair of original books and fanfictions associated with the original series, a similarity score was computed using Dynamic Time Warping (DTW) algorithm Bellman and Kalaba (1959). DTW is an algorithm for measuring similarity between two temporal sequences, which may have different lengths.

We took advantage of the Tslearn library Tavenard et al. (2020) that has implemented a version of DTW. Because of this algorithm's computational burdens, a sampling reduction in arc lengths was carried out. To drop all the fluctuations and the disturbance embedded in the emotional arcs, using time series analysis, we decomposed the arcs to their trends, seasonalities and noises exploiting the statsmodels library[5]. Having the components of the arc separated, we focused on the trend component as the overall trend and shape of the stories has the highest importance. All the similarities were computed for each pair of fanfictions and original books since writers may be influenced by all the originals when writing their story and not only one of the books in a series.

4.2 Characters' graph

To further investigate the relation between fanfictions and their associated originals, we generated the graphs of the characters for each story. Each node in the graph represents a character. Two nodes are connected if there are interactions between them in the context of the story. We assume the existence of interactions if both characters' names are observed within a 15-word window in the text, which is suggested by Beveridge and Shan (2016). In this section, our method for character graph creation is discussed.

4.2.1 Extracting the graphs of interactions

To generate character graphs, we first need to detect the names of people within the story. To do so, we used BookNLP Bamman et al. (2014) which enables us to extract characters in stories alongside their names and nicknames, exploiting the *Character name clustering* ability in the BookNLP pipeline. This method is advantageous to other named entity recognition methods as it can match and cluster all the ways a character has been mentioned throughout a story (e.g., "Tom," "Tom Sawyer," "Mr. Sawyer," "Thomas Sawyer" → TOM SAWYER).

To reduce the errors in clusters of names, we carried out several post-processing techniques on these clusters. The first issue with the clusters was that the results generated for each volume in a series were slightly different from other volumes, although they used the same characters. In other words, in the case of each book, clusters were comprised of names and nicknames that were more frequent in that specific book and different from other books. For example, in the case of the Harry Potter series, it was possible to have a cluster with names Harry and Harry Potter in the same cluster, but not Mr. Harry Potter, although this might not be the case in other volumes and clusters. To alleviate this issue, we stacked up all the original books and ten fanfictions and fed them to the BookNLP. In this way, clusters became more robust, and they were comprised of approximately all the names and nicknames alluding to the same characters in all the stories. After the post-processings were done on the clusters, we fed each individual fanfiction to the BookNLP pipeline and obtained the results. We then matched the names extracted against the overall characters' cluster computed from the stacked version of the books. As a result, we were able to create more accurate clusters by taking the similarities between every two clusters to merge any pair showing extreme similarities. The above technique was used in every stage to remove any redundant clusters and make results more precise. The outcome of these actions

[4] https://pypi.org/project/fitz/#description.

[5] https://www.statsmodels.org/dev/generated/statsmodels.tsa.seasonal.seasonal_decompose.html.

was character sets that are mostly identical across all the books in the same series and contributed positively to the task of comparing them.

Having the data ready for extracting the interactions between characters, we adopted an approach akin to Beveridge and Shan (2016). Each occurrence of pairs of characters with a distance of 15 words was interpreted as one interaction. Through this method, we created an array of which we used to generate the graphs of interactions between characters.

4.2.2 Character graph similarity between fanfictions and original stories

Moreover, to investigate the degree of structural similarity between fanfictions and their originals in terms of their characters' interactions' graphs, we computed the graphs' similarity scores as follows. For every two stories that we are interested in their graphs' similarity, we first matched their characters in terms of their similarity of names to extract the characters two stories had in common. Then, we computed the Hadamard product (a.k.a. elementwise product) of the adjacency matrices and counted the number of nonzero elements. This way, we count the number of common edges between the two graphs we are looking at. After normalizing this metric by the count of all edges in the graph of the fanfiction, we can interpret the result as the percentage of common edges in the fanfiction graph and the original one. A nuance that should be covered is that we matched the characters based on their names' similarities and different strictness levels. The algorithm we used was a modified version of an algorithm published in the late 1980s by Ratcliff and Obershelp under the hyperbolic name "gestalt pattern matching" provided by difflib[6]. Although this prunes some nodes we have in the characters graphs, the computed results would be more reliable to perform a regression analysis in Sect. 5.4.

5 Results

In this section, the results of our analysis and our interpretations have been thoroughly explained.

5.1 Relation between emotional arc similarity and fanfictions' popularity

Emotional arcs of the stories can provide us with the intuition that to what degree fanfictions share the same emotional flow in the story with their associated original books. Note that we use the term *similarity score* throughout this section to refer to the similarity scores between fanfictions and their associated original stories in terms of their emotional arcs.

The control variables used in this study can be divided into two categories. The first group contains all the information available on the `fanfiction.net` that we gathered. Number of reviews, number of follows, number of people who favored the fanfiction, the genres of each fanfiction, and the length of the fanfiction lay in this category. The other category consists of the properties we extracted from the text of each fanfiction. The average happiness score of the story was computed with a similar approach to Sect. 4.1. Plus, different properties of the characters' relations graph were extracted to account for their specific impact on the popularity of the fanfictions. These properties were: assortativity, average degree, density, diameter and the similarity of the fanfictions to their original stories in terms of communities. For extracting the communities and also to calculate their similarities, the Louvain method Blondel et al. (2008) and Jaccard similarity Jaccard (1912) were utilized. Note that each one of these properties can associate with the popularity of a fanfiction in a distinct way. For instance, the assortativity of the graph can represent the degree to which the characters featured in the story prefer having connections with people with whom they share similar interests. In other words, this way, we are controlling for different behavioral aspects of people involved in the line of the story.

We define the popularity of a fanfiction as its number of favorites. However, the number of reviews a fanfiction gets, as well, can be an indicator of its popularity as these two metrics are highly correlated with each other with more than 95%. By applying negative binomial regression Johnson (2011) to regress the popularity of fanfictions on the degree of similarity that fanfictions have to their original series in terms of their emotional arc, we can infer that there is a significant correlation between these two metrics. Our results in Table 7 show that the more similar fanfictions' emotional arcs are to their original series, the less they are popular among the fanfiction community. This confirms the hypothesis that since the authors of fanfictions are fans who prefer an alternate turn of events, common audiences of this community would prefer to read a story that suggests a different flow of emotions. The significance of the results is presented in Table 7 separately analyzed for each series.

5.2 Relation between genres and popularity of fanfictions

With genres extracted from `fanfiction.net`, we were able to analyze how falling under some genre would affect how popular a fanfiction will be. The genres associated with fanfictions are presented in Table 6. We found that having certain kinds of genres would influence the overall

[6] https://docs.python.org/3/library/difflib.html#module-difflib.

Table 6 Results of Regressing Popularity of Fanfictions on Genres

Series	Variable	Coefficient	*p*-value	Significance
Harry Potter	Family	0.362	0.015	*
	Fantasy	−0.393	0.014	*
	Sci-fi	2.976	0.042	*
	Romance	−0.084	0.523	
	Mystery	−0.496	0.008	**
Twilight	Family	0.021	0.582	
	Fantasy	−0.505	0.004	**
	Sci-fi	−1.591	0.006	**
	Romance	0.177	0.071	.
	Mystery	−0.557	0.012	*
Percy Jackson & Olympians	Family	0.751	0.005	**
	Fantasy	0.090	0.734	
	Sci-fi	−1.222	0.342	
	Romance	0.722	0.001	***
	Mystery	−0.094	0.781	
The Lord of the Rings	Family	0.356	0.026	*
	Fantasy	−0.144	0.351	
	Sci-fi	0.002	0.369	
	Romance	0.686	0.072	.
	Mystery	−0.496	0.037	*
The Hunger Games	Family	0.075	0.672	
	Fantasy	0.300	0.246	
	Sci-fi	−0.402	0.095	.
	Romance	1.029	0.000	***
	Mystery	−0.218	0.542	

$p < 0.1^{\cdot}$, $p < 0.05^{*}$, $p < 0.01^{**}$, $p < 0.001^{***}$

popularity of fanfictions. For example, we see that fanfictions categorized as romantic are generally more accepted by the common audience Rawlings (2002) except for Harry Potter. However, considering that the background of each original series may be different from the others, various combinations of genres can be influential in each case. For instance, writing more in sci-fi domain is not well championed in the Hunger Games or Twilight. Also one of the interesting findings is that authors are in favor of having a family aspect in their stories. But in general, we can conclude that the genres authors choose to write in, although different for each series, do relate to the popularity of fanfictions.

5.3 Relation between stories' lengths and fanfictions' popularity

By extracting the number of words in each fanfiction, we found a significant correlation (a zero p-value) between the length of the story and its popularity for each series. Also, the corresponding coefficient with this variable in our regression model was positive; in other words, the longer stories are more likely to be popular as Blog (2015) posits too. One possible explanation for this observation is that the longer the fanfiction, the author is more capable of making the content immersive, which is not really the case for short stories. Besides, with the fanfictions released on a digital platform, the size of the printed versions of bigger books does not bother the readers anymore. Plus, the presumption that people have less time for anything in the modern era would not be the case for people who love reading and are being active in these kinds of communities. The adjusted R^2 reported for the results in Tables 7 and 6 are presented in Table 8.

5.4 Relation between graph similarity and popularity

In this section, we present the results obtained from the analysis of the networks of characters and their interactions within the fanfictions and their associated originals.

To see how the graph similarity is related to the popularity of fanfictions, in the regression performed, we incorporated the effect of characters' interactions network by including it as a variable in the regression. The results are shown in Table 7. For the series examined in this study, except for the Harry Potter, we can infer that the degree to which a fanfiction is similar to its original story in terms of character graph is related to the popularity of that fanfiction. To exemplify, Fig. 2 illustrates this inference. In Twilight series, we can observe that the fanfictions having more similarity to their original book are not as popular as the ones who have chosen an alternate network of characters' interactions. Having that said, we can say that in terms of the popularity of fanfiction, paying attention to the relationship between characters in the story and their deviation from how the original story used to plot them do matter, but based on what stories we are talking about. In some stories, we observe that the way characters are in contact with each other is not something that people want to change about the original story. As Jamison and Grossman (2013) raises, fanfictions are a way to get back to the older way of thinking. Certain parts of the stories are just there to be in common and untouched among all people who narrate them. In fact, this is how people kept the stories in the past. In the case of Hunger Games, Twilight, and Percy Jackson & the Olympians, results suggest that more successful fanfiction authors were the ones who changed the way characters were in contact with each other. On the other hand, the results are different for the rest of the series, and keeping the characters' relations similar to the original story acts in favor of the popularity of the fanfictions.

6 Conclusion

As a huge source of fan-written stories inspired by original series, fanfiction help us develop intuition about what people like or do not like about the original series and how

Table 7 Results of regressing popularity of fanfictions on features including the extracted and metadata

Series	Variable	Coefficient	*p*-value	Significance
Harry Potter	Similarity of Emotional Arcs	−0.005	0.000	***
	Happy Ending	−0.260	0.074	.
	Average Happiness	0.138	0.054	.
	Average Degree	−0.140	0.088	.
	Community Similarity (Jaccard)	−0.275	0.185	
	Graph Similarity	0.183	0.174	
	Graph Diameter	0.124	0.077	.
Twilight	Similarity of Emotional Arcs	−0.006	0.001	***
	Happy Ending	−0.145	0.206	
	Average Happiness	0.309	0.079	.
	Average Degree	0.013	0.014	*
	Community Similarity (Jaccard)	−0.009	0.008	**
	Graph Similarity	−0.016	0.019	*
	Graph Diameter	−0.739	0.156	
Percy Jackson & the Olympians	Similarity of Emotional Arcs	−0.010	0.001	***
	Happy Ending	−0.052	0.735	
	Average Happiness	−0.066	0.005	**
	Average Degree	−0.211	0.077	.
	Community Similarity (Jaccard)	0.042	0.014	*
	Graph Similarity	−0.006	0.004	**
	Graph Diameter	−0.117	0.041	*
The Lord of the Rings	Similarity of Emotional Arcs	−0.008	0.000	***
	Happy Ending	−0.029	0.776	
	Average Happiness	0.476	0.174	
	Average Degree	−0.070	0.038	*
	Community Similarity (Jaccard)	0.054	0.029	*
	Graph Similarity	0.058	0.036	*
	Graph Diameter	0.193	0.109	
The Hunger Games	Similarity of Emotional Arcs	−0.002	0.031	*
	Happy Ending	0.004	0.072	.
	Average Happiness	−0.405	0.062	.
	Average Degree	−0.251	0.103	
	Community Similarity (Jaccard)	−0.429	0.105	
	Graph Similarity	−0.102	0.035	*
	Graph Diameter	−0.010	0.009	**

$p < 0.1^{\cdot}$, $p < 0.05^{*}$, $p < 0.01^{**}$, $p < 0.001^{***}$

Table 8 The Adjusted R^2 for Regressions Done with the Popularity of the Fanfictions as Response Variable

Series	Adjusted R^2
Twilight	0.14
The Lord of the Rings	0.26
Harry Potter	0.18
The Hunger Games	0.28
Percy Jackson & the Olympians	0.23

they prefer the story to be. In this study, we focused on two of the factors influencing the popularity level of fanfictions. The original stories and how fanfictions are similar to them in terms of the emotional arcs and characters graph found to be significantly related to the popularity of fanfictions. Also, the genres of the fanfictions and their length were indicators of the popularity of fanfictions. In addition to emotional arc, similarity, genres, length of stories, character graph similarity, we also analyzed the standard metrics in social network analysis on the characters graphs such as degree of assortativity, average degree and diameter of the character graphs. We explored how the communities extracted from character graphs can be similar or different from the original stories.

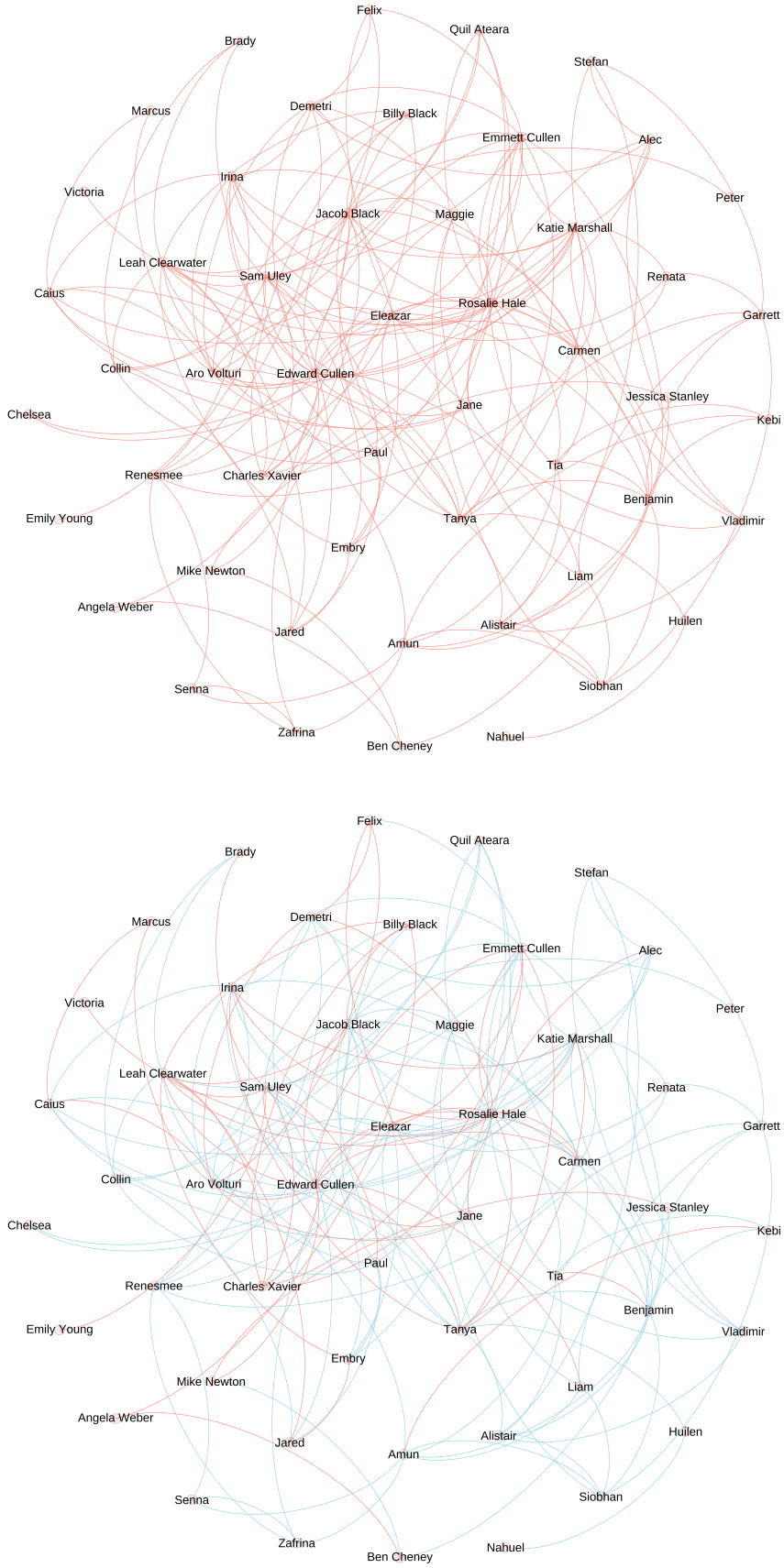

Fig. 2 The Character Graphs of three stories associated with the Twilight series: (**a**) Twilight Breaking Dawn, (**b**) a fanfiction being similar to the original story in terms of characters' interactions, (**c**) a fanfiction being dissimilar to the original story in terms of the characters' interactions. The edges in blue in figures (**b**) and (**c**) represent the interactions in the original story that are not present in the fanfiction, and the edges in red represent the interactions that are present in both the original story and the fanfiction

Fig. 2 (continued)

We found none of the mentioned variables relevant to the popularity of fanfictions. Although attending to the importance of characters' interactions graph and emotional arcs have been previously explored to a certain degree, as fanfictions might still have particular connections to their originals, paying attention to the similarity between fanfictions and their originals is an area that should be explored more. Also, as stated in Sect. 2, examining graph of characters interactions has received scant attention in fanfictions compared to movies. As a result, this study contributes to laying the ground for increased explorations similar to the work has being done for films in written stories. Furthermore, regarding the results we reported in this research article, several important questions can be brought up. Delving deeper into the particular features that every series possess, one can try to find the underlying factors which affect the results for different series. For instance, whether sticking to the original setting would be beneficial or not would necessitate further investigation over the unknown elements which are most probably related to the series' social theme.

Funding No funding was received for conducting this study. The authors have no relevant financial or non-financial interests to disclose.

Data availability The datasets generated during and/or analyzed during the current study are not publicly available due to the policy of fanfiction.net that the data should not be publicly disclosed.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Code availability (software application or custom code) The codes used during the current study are not publicly available.

References

Aires V, Nakamura F, Nakamura E (2017) A social network analysis of successful movies with complex networks. Anais Estendidos Do XXIII Simpósio Brasileiro De Sistemas Multimídia E Web. pp. 99-102, https://sol.sbc.org.br/index.php/webmedia_estendido/article/view/4842

Bamman D, Underwood T, Smith NA (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), Association for Computational Linguistics, Baltimore, Maryland, pp 370–379, https://doi.org/10.3115/v1/P14-1035, https://www.aclweb.org/anthology/P14-1035

Bellman R, Kalaba R (1959) On adaptive control processes. IRE Trans Autom Control 4(2):1–9. https://doi.org/10.1109/TAC.1959.1104847

Beveridge A, Shan J (2016) Network of thrones. Math Horizons 23(4):18–22

Bird S, Klein E, Loper E (2009) Natural language processing with Python: analyzing text with the natural language toolkit. O'Reilly Media, Inc

Blondel V, Guillaume J, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. (IOP Publishing, 10), https://doi.org/10.1088%252F1742-5468%252F2008%252F10%252Fp10008

Bronwen Thomas (2011) What is fanfiction and why are people saying such nice things about it?? Storyworlds A J Narrat Stud. https://doi.org/10.5250/storyworlds.3.2011.0001

Burt S (2017) The promise and potential of fan fiction. https://www.newyorker.com/books/page-turner/the-promise-and-potential-of-fan-fiction, [Online accessed 25-December-2020]

Büchler M, Franzini G, Kestemont M, Manjavacas E (2018) The re-creation of harry potter: Tracing style and content across novels, movie scripts and fanfiction. In: DH, pp 715–716

Campbell J, Aragon C, Davis K, Evans S, Evans A, Randall D (2016) Thousands of positive reviews: distributed mentoring in online fan communities. Proceedings Of The 19th ACM conference on computer-supported cooperative work & social computing. pp. 691-704, https://doi.org/10.1145/2818048.2819934

Carvallo A, Parra D (2020) Analyzing network effects on a fanfiction community. 1909.02886

Chu E, Roy D (2017) Audio-visual sentiment analysis for learning emotional arcs in movies. In: IEEE international conference on data mining (ICDM), pp. 829–834. https://doi.org/10.1109/ICDM.2017.100

Del Vecchio M, Kharlamov A, Parry G, Pogrebna G (2021) Improving productivity in Hollywood with data science: using emotional arcs of movies to drive product and service innovation in entertainment industries. J Op Res Soc 72(5):1110–1137. https://doi.org/10.1080/01605682.2019.1705194

Dodds PS, Harris KD, Kloumann IM, Bliss CA, Danforth CM (2011) Temporal patterns of happiness and information in a global social network: hedonometrics and twitter. PLoS ONE 6(12):e26752. https://doi.org/10.1371/journal.pone.0026752

Dodds PS, Clark EM, Desu S, Frank MR, Reagan AJ, Williams JR, Mitchell L, Harris KD, Kloumann IM, Bagrow JP, Megerdoomian K, McMahon MT, Tivnan BF, Danforth CM (2014) Human language reveals a universal positivity bias. Preprint available at http://arxiv.org/abs/1406.3855

Flipsnack B (2015) Are books getting bigger and better?. [online] https://blog.flipsnack.com/flipsnack-study-are-books-getting-bigger-and-better/ Accessed 1 October 2021

Frens J, Davis R, Lee J, Zhang D, Aragon C (2018) Reviews matter: how distributed mentoring predicts lexical diversity on fanfiction.net. arXiv preprint arXiv:180910268

Grady C (2016) Why we're terrified of fanfiction. https://www.vox.com/2016/6/2/11531406/why-were-terrified-fanfiction-teen-girls, [Online accessed 25-December-2020]

Hilbe JM (2011) Negative Binomial Regression, second edition.. (Springer Science,2012,3), https://doi.org/10.1007%252Fs11336-012-9263-7

Hipson WE, Mohammad SM (2021) Emotion dynamics in movie dialogues. arXiv preprint arXiv:2103.01345

Jaccard P (1912) The distribution of the flora in the alpine zone.1. (Wiley,1912,2), https://doi.org/10.1111%252Fj.1469-8137.1912.tb05611

Jamison AE, Grossman L (2013) Fic: Why fanfiction is taking over the world. Smart Pop, an imprint of BenBella Books, Inc

Jonah B, Kim YD, Meyer R (2021) What makes content engaging? How emotional dynamics shape success. J Cons Res 48(2):235–250. https://doi.org/10.1093/jcr/ucab010

Kounelis A, Vikatos P, Makris C (2021) Movie Recommendation System Based on Character Graph Embeddings. Artificial Intelligence Applications And Innovations. AIAI 2021 IFIP WG 12.5 International Workshops. pp. 418-430

Labatut Vincent, Bost Xavier (2019) Extraction and analysis of fictional character networks: a survey. ACM Comput Surv 52(5):89. https://doi.org/10.1145/3344548

Mattei Andrea, Brunato Dominique, Dell'Orletta Felice (2020) The Style of a Successful Story: a Computational Study on the Fanfiction Genre, vol 2769. CEUR, Bologna, Italy

Milli S, Bamman D (2016) Beyond canonical texts: a computational analysis of fanfiction. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pp 2048–2053

Pianzola F, Rebora S, Lauer G (2020) Wattpad as a resource for literary studies. quantitative and qualitative examples of the importance of digital social reading and readers' comments in the margins. PLoS ONE 15(1):e0226708

Rawlings P (Ed.) (2002) Americans on Fiction, 1776-1900 (1st ed.). Routledge. https://doi.org/10.4324/9781351223461

Reagan AJ, Mitchell L, Kiley D, Danforth CM, Dodds PS (2016) The emotional arcs of stories are dominated by six basic shapes. EPJ Data Sci 5(1):1–12

Rubin VL, Girouard V (2014) Comparative stylistic fanfiction analysis: popular and unpopular fics across eleven fandoms

Sarbin TR (Ed.) (1986) Narrative psychology: the storied nature of human conduct. Praeger Publishers/Greenwood Publishing Group

Tavenard R, Faouzi J, Vandewiele G, Divo F, Androz G, Holtz C, Payne M, Yurchak R, Rußwurm M, Kolar K, Woods E (2020) Tslearn, a machine learning toolkit for time series data. J Mach Learn Res 21(118):1–6 (http://jmlr.org/papers/v21/20-091.html)

Vonnegut K (2010) Kurt Vonnegut on the Shapes of Stories. [Video]. YouTube. https://www.youtube.com/watch?v=oP3c1h8v2ZQ. Accessed 1 October 2021

Wang X, Zhang S, Smetannikov I (2020) Fiction popularity prediction based on emotion analysis. 2020 International Conference On Control, Robotics And Intelligent System. pp. 169-175, https://doi.org/10.1145/3437802.3437831

Zhao A (2016) Predicting popularity of fanfiction stories based on title and summary