

Across the Pages: A Comparative Study of Reader Response to Web Novels in Chinese and English on Qidian and WebNovel

Ze Yu*, Federico Pianzola

Center for Language and Cognition, University of Groningen, The Netherlands

Abstract

The evolution of online reading platforms has transformed engagement with fiction, with platforms like WebNovel bridging cultural boundaries through translated Chinese web novels. This study employs topic modeling to compare reader responses to the same stories published in Chinese on Qidian and in English on WebNovel, focusing on English and Chinese language comments. We identify shared and unique themes, revealing that while both communities emphasize characterization and story development, cultural-specific expressions and platform dynamics shape readers' interactions. Our findings underscore the nuanced interplay between language, culture, and the affordances of digital platforms in shaping global literary consumption and community engagement.

Keywords

Digital social reading, reader response, cross-cultural studies, topic modeling

1. Introduction

The evolution of online reading platforms, from early content-centered libraries like Project Gutenberg and the Internet Archive to user-centered platforms such as Fanfiction.net, Archive of Our Own, and Wattpad, highlights the growing importance of online reading. These platforms not only provide a space for collective reading but also foster social interaction and community building, offering writers interactive opportunities [7] and readers direct engagement with authors [20]. Online reading communities are crucial for young readers and writers, fulfilling emotional and social needs [25], and providing a space for marginalized voices [5, 3]. Many researches [10, 21] have explored the various aspects of Digital Social Reading (DSR), and one of the many focuses is the reader response [11, 19]. However, current research on storytelling and reader response has overlooked cross-cultural comparisons, with only a few exceptions [15]. This research gap has been identified not only in the field of DSR but also in comparative literary studies, which primarily focus on understanding the cultural influences behind literature by examining authors as transcultural readers rather than investigating the perspectives of readers. Consequently, this focus on literary production leaves a gap in our understanding of how readers from different cultural backgrounds interpret and engage with

CHR 2024: Computational Humanities Research Conference, December 4–6, 2024, Aarhus, Denmark

*Corresponding author.

✉ z.yu@rug.nl (Z. Yu); f.pianzola@rug.nl (F. Pianzola)

🆔 0009-0005-5648-6470 (Z. Yu); 0000-0001-6634-121X (F. Pianzola)

© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



literature. Questions such as whether cultural settings influence the understanding of topics, characters, and plots, or if culture shapes reading in certain dimensions but not others, have not been extensively investigated [29]. At the same time, the most frequently studied books and readerships remain those with distinguished popularity, commercial success, social impacts, and scholarly prestige in the Anglophone world, which are subject to historical and social-cultural biases such as classism, sexism, racism, colonialism [2, 27, 15], and might not be inclusive enough to understand reader response more broadly.

Hu et al. [15] conducted a comparative analysis on the book list and tags across Goodreads and Douban –an international and Chinese platform, respectively –and noticed divergences in readers’ collective understanding of classics. Following up on that study, we are interested in exploring what differences there might be in the readers’ response to the same stories when they are read in different languages. To this end, we conducted a comparative study of different aspects of reader response, using topic modelling to focus on the aspects that are mentioned in the comments left on two reading platforms publishing the same stories in Chinese and English translation. The research question that we address is: do readers of different languages and cultural backgrounds have a different focus when commenting on a story, such as the plot, the characters, or the setting? Do such reader responses have recognisable features that are specific to one language and culture? The dataset we worked on are from two DSR platforms: Qidian.com (original Chinese web novels) and Webnovel.com (English translation).

1.1. Digital Social Reading Platforms for Chinese Stories

Since its birth in the 1990s, Chinese webnovels have grown rapidly to become a new form of Chinese literature. In recent years, Chinese webnovels have become increasingly popular worldwide, emerging as a significant form of participatory transcultural storytelling. They not only have a large number of loyal fan-readers in China, but have also become increasingly popular among international readers by being translated into many languages and circulating in different countries [16]. As defined by Michel Hockx, Chinese online literature is “Chinese-language writing, either in established literary genres or in innovative literary forms, written especially for publication in an interactive online context and meant to be read on-screen” [14]. The spread of Chinese literature thanks to digital technology provides opportunities for the cultural influence of this literature abroad, but at the same time, his reach dilutes the national attributes of Chinese online literature, loosening the boundaries between different cultural spheres [16].

Some scholars [26, 16, 22] have pointed out that Chinese online literature has reached such a wide audience due to its foundation in a rich literary tradition spanning classical, modern, and contemporary China, and has inherited and integrated Western fantasy elements and Hollywood narrative techniques while maintaining connectivity with other popular literature and then localizing and recreating them in the Chinese context [16]. The imaginative worlds constructed in online fiction encompass history, military, war, romance, recent actual events, and the future, all expressed in a traditional and secular Chinese style, while at the same time allowing readers from all over the world to find familiar elements in them. This has fostered its connectivity with other successful popular cultures, enabling it to attract readers from other cultural backgrounds while giving them a sense of familiarity with the narrated world. An-

other explanation for the success of Chinese novels is that these online novels were created for hedonistic reading, or “popcorn literature” (*Shuang*). The process of reading such novels is supposed to be light, pleasurable, and exciting. For example, many of these Chinese novels may cover elements of Taoism, the Three Thousand Worlds, etc., but they are not conveyed in a very orthodox or obscure way, but rather in a simple and clear way, without requiring relevant knowledge or a specific cultural background to understand them. Readers from any cultural background can easily get into the story. Following Hollywood movies, Japanese animation, and Korean dramas, Chinese web novels have become the fourth largest cultural phenomenon in the present world [18, 16].

Qidian is one of the earliest online reading platforms in mainland China, founded in 2002, with an innovative pay-to-read model and works covering urban, fantasy, romance, science fiction, mystery, sports, games etc., extending to more than 200 genres. Nowadays Qidian is one of the largest online reading platforms with more than 30 million registered readers and more than one million stories.

WebNovel, a platform owned by the same corporation as Qidian, was officially launched in 2017, and is the first official platform for the overseas dissemination of Chinese online literature. It started by providing English translations of stories that were originally published on Qidian, and later also expanded to encourage authors to freely create and upload their own stories on this platform. According to the 2023 China Online Literature Overseas Trend Report [1], WebNovel has launched translations of about 3,600 Chinese stories, with 238 translated works that have been read by more than 10 million people, and 9 that have exceeded 100 million readers. The gender distribution of WebNovel consists of 66% male readers and 34% female readers. The top ten countries in terms of percentage of visits by country/region are the United States, India, the United Kingdom, Indonesia, Thailand, Venezuela, Ghana, Vietnam, and Egypt, with the United States dominating the percentage of visits to the website, accounting for 20.7% by October 2020.

2. Parallel Corpus

The corpus creation process involved a manual search for translated novels available on Webnovel.com within the NOVEL category, only targeting completed works.¹ Subsequently, each identified translated novel was mapped with its original counterpart on Qidian.com. This process yielded a total of 120 novels for inclusion in the corpus. However, the copyright for 10 of these novels on Qidian.com had expired and it was difficult to locate comments of these novels on the Chinese platform, rendering them ineligible for inclusion. The final corpus consists of 110 stories [28]. According to WebNovel’s categorization visible on the website interface, these 110 stories consist of 103 Male Lead and 7 Female Lead. We scraped the metadata from each platform and the respective comments and replies for each story, including the user profile of the readers who left the comments. Table 1 has shown the metadata for comments and replies, based on the information provided, we could also map the interactions between comments and their replies.

¹The corpus metadata and the code for the analysis are available at <https://github.com/zeyu-acad/Qidian-Webnovel-Corpus>; the full dataset can be accessed at [28]

Table 1
Corpus Metadata for WebNovel (EN) and Qidian (CN)

Source	CommentId	Comment Content	ReplyId	Reply Content	BookId	UserId	User Level	Rating Score	Reply Amount	Like Amount	Create Time	Quote ReviewId	Quote Content	Quote UserId
WebNovel (EN)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Qidian (CN)	✓	✓	✓	✓	✓	✓	✓	-	✓	✓	✓	✓	✓	✓

The length of the novels exhibits a wide range, spanning from 288 to 3,588 chapters, with corresponding word counts varying between 1,027,000 and 8,448,900 (measured in Chinese characters). Notably, there is a slight difference in the genres and categories of stories on Qidian and WebNovel.

We also collected the user profiles of readers who has left comments or replies on the stories, as shown in Table 2.

Table 2
User Metadata for WebNovel(EN) and Qidian(EN)

Source	UserId	User Name	Gender	Level/levelInfo	Writing Days	Reading Hours	Num of Read Books	Description	Date Joined	Location	Num Followers
WebNovel (EN)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-
Qidian (CN)	✓	✓	✓	✓	-	-	✓	✓	-	✓	✓

Unlike Qidian where the readership is mainly native Chinese speakers or overseas Chinese-using groups, readers on WebNovel do not necessarily come from the same region. We looked into the location distribution of readers in the dataset, using the location available on the users profiles. For the WebNovel dataset, we have the readers from 244 countries, and the top 10 locations with the most users are shown in Table 3.

Table 3
Location Distribution of users on WebNovel(EN)

Location	Number	Percentage (%)
Global	55100	42.80
United States	15891	12.30
Philippines	14425	11.20
India	9591	7.40
Indonesia	3231	2.50
Nigeria	2972	2.30
Malaysia	2277	1.70
Canada	2046	1.50
Australia	1602	1.20
United Kingdom	1584	1.20
Brazil	1478	1.10

We have also taken into account that book comments left on WebNovel are not necessarily in English, so we looked into the language distribution of comments and replies using automatic language detection.

The results (Table 4) show that English comments accounted for 72.7% of the total replies, and English replies accounted for 68.2% of the total replies. All other languages only account

Table 4
Metadata for stories on both platforms

Source	Genres	Categories/Tags	Total Comments	Replies	Primary Language	Comments Language Distribution	Replies Language Distribution
Qidian.com	14	27	2,791,837	855,577	Chinese	Chinese: 95.7%, English: 0.1%	Chinese: 97.2%, English: 0.05%
Webnovel.com	8	40	327,988	96,250	English	English: 72.7%, Others: 27.3%	English: 68.2%, Others: 31.8%

for no more than 2% of the comments/replies each. Given that our research is focused on comparing Chinese and English readerships, we only consider replies and comments written in English.

3. Methodology

To identify which aspects of the stories Chinese- and English-speaking readers focus on when commenting online, we employed topic modeling. We considered both Latent Dirichlet Allocation (LDA) and embeddings-based modeling. LDA has the advantage of being able to assign several different topics to one document, by generating models with multinomial distribution over topics. However, its efficacy in analyzing social media data has been highly criticized [9, 24]. Noisy and sparse datasets are unsuitable for LDA [6] due to a lack of enough textual features for statistical learning [4, 8].

BERTopic [13] employs a clustering embeddings approach and extends it by incorporating a class-based variant of TF-IDF for creating topic representations. It has been proved that its effectiveness to generate insights from short and unstructured text offers the most potential [8]. However, with BERTopic each document is only assigned to a single topic. Even though topic probabilities can be extracted, they are not equivalent to an actual topic distribution [8], meaning that we may lose the ability to analyze the complexity of each document, especially for longer comments where various aspects of a story might be commented on.

We conducted preliminary evaluation on the topic modeling of comments in both languages, evaluating the performance of LDA and BERTopic with various core configurations and pre-trained transformer models.

3.1. Data Preparation

Preprocessing is a critical step in ensuring data quality and consistency. Although BERTopic does not require preprocessing of the input text, a study in comparison of LDA and BERTopic [17] shows that topic diversity and coherence is higher in both cases with fully preprocessed text. So we preprocessed our data as well.

We randomly examined a sample of comments and found that there were comments with unusual expressions. For example, "XpXPPXPXPXPXPXPXPZP", where "xp" refers to the experience points a user can gain to level up and get rewards on both platforms). There are also some random typing (e.g. "F f f f f f. F f g. F t t f. T t. T t t. T t r. R r. R e e w") and sequences of emojis. We removed such comments with unconventional and inconsistent spelling, as well as punctuation, numbers, and stopwords. We also performed lemmatization and stemming to enhance the coherence of the analysis.

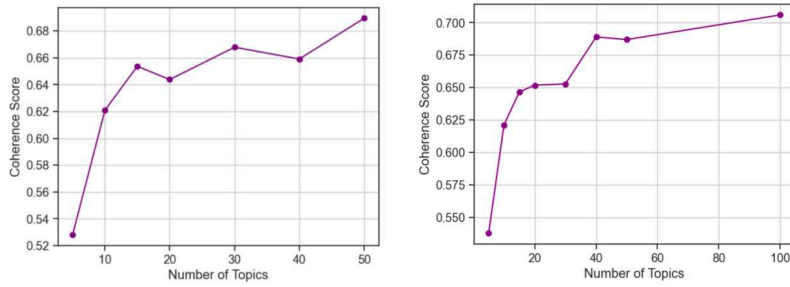


Figure 1: (left: LDA Topic Coherence on English comments of WebNovel; Right: LDA Topic Coherence on Chinese comments of Qidian)

3.2. Topic Modeling Evaluation

We first applied LDA to the comments of both platforms separately. Initially, we assessed the C_V [23] coherence of the topics by using a range of topic numbers for both language dataset. The results, illustrated in 1, suggest that 30 topics could provide a good balance between the number of topics and their coherence on both dataset. Then we conduct an evaluation including a close reading of the topic words, to better understand the underlying meanings of the topics [8]. The LDA generated topics were quite difficult to interpret, so we decided to try BERTopic², using the same number of Topics. BERTopic-generated topics were easier to interpret. Moreover, when integrated with the KeyBERT-inspired approach[12], the model performed better in generating overall coherent yet diverse topics, even though it did not achieve the highest C_V score (0.49).

Based on this evaluation, we chose to use BERTopic.

4. Results

The results of the topic modeling (Table 5) suggest that readers who leave comments in both languages pay attention to characterisation, story development, reading experience, but they also leave comments without any specific meaning, just with the intention to gain account experience (WebNovel Topic 5; Qidian Topic 0). Additionally, some readers of both platforms were able to recognise elements of the novels’ setting borrowed from other popular cultures (WebNovel Topic 16 and 17; Qidian Topic 2).

However, beside these commonalities, the topics also reflect some features that are exclusive to comments in each language. For example, Chinese-speaking readers on Qidian use many formulaic sentences (Qidian Topic 11, 17, 18, 20, 22, 26), which are unique phrases or sentences that can only be understood in the same cultural context. The literal meaning of these expressions may seem irrelevant to the story text, but their derived meaning can be captured by other readers and invite a discussion. For example, “缝缝补补又三年。” means “After mending, the worn-out cloth can last for three more years.” and it is an expression used to refer to a

²To address the volume of Chinese comments and mitigate memory issues during topic modeling, we utilized the (`min_df`) parameter to indicate the minimum frequency of words.

recurring thing. In the comment, it refers to a similar event/ pattern that keeps repeating in the story. Chinese readers tend to leave comments (Qidian Topic 7) when implicit sexual depictions appear in the text, as a playful way of proving that they had noticed. This may be due to content censorship, as the platform is aimed at an all-ages audience and the terms of service forbid explicit sexual depictions to appear in the text. For Example, “我怀疑你在开车但我没有证据。” means “I suspect you are implying a sexual scene but I don’ t have any evidence.”

As for the reader response on WebNovel, we can identify some unique topics, for example, the push for updates of the story (WebNovel Topic 3), and comments left as bookmarks, like “Plan to read” (WebNovel Topic 26). We also observed readers express gratitude towards the translator when they approved the good translation quality of the work (WebNovel Topic 24). There are also comments related to sensitive themes (WebNovel Topic 27) that did not show up in the Qidian topics. These comments (Appendix A) suggest that the stories in question may include themes or narratives that are racially insensitive or sexist. Readers might find these themes offensive and respond strongly to them. Similarly, when the content of a story does not meet the readers’ expectations, readers who are deeply invested in it may have strong negative reactions.

When retrieving topics and corresponding comments, we performed simultaneous keyword searches on the dataset. By manually querying topic-related terms, we identified related comments and found that the actual number of comments belonging to a given topic exceeded those clustered by topic modeling. For instance, WebNovel topic 27 clustered 15 comments annotated as Sensitive/Violence. However, a keyword search for “racist,” which frequently appeared in these comments, yielded 564 entries. Additionally, we observed that topic modeling for both languages indicated -1 (unassigned) as the dominant cluster. This might suggest that many comments may belong to other topics within the clustering. The presence of such a large number of unassigned clusters warrants further investigation.

5. Conclusion

This study highlights the differences and commonalities in reader responses to Chinese web novels across two platforms: Qidian and WebNovel. The findings indicate that readers of both languages focus on characterization, story development, and the overall reading experience. Additionally, interactions extend beyond the narrative, as readers leave comments to gain account experience, a behavior prompted by the platforms’ affordances rather than by a direct engagement with the story. Distinctive patterns also emerged in the comments of both languages. Chinese-speaking readers on Qidian frequently use culture-specific formulaic sentences and comment on implicit sexual scenes. These kinds of comments not directly related to the story create a shared understanding among readers, fostering a collective awareness of the subtleties in the narrative. It builds a sense of community where readers acknowledge the same hidden layers of meaning, and might enhance the collective reading experience. Readers who leave comments in English on WebNovel emphasize pushing for story updates, recommending the story (or not), using comments as bookmarks, and expressing gratitude towards translators in case of high-quality translations. Notably, WebNovel also sees a higher occurrence of sensitive themes, with comments often criticizing perceived racism and sexism in the novels.

Table 5
Comparison of Topics/Keywords between WebNovel and Qidian

No.	WebNovel Topics/Keywords	Size	Annotation	Qidian Topics/Keywords	Size	Annotation
-1	read_reader_novel_write	137,404	Outliers	Outliers	1,286,482	主角_没有_知道
0	novel_good_book_story	92,267	Book, Chapter, Author	Comments for experience	1,427,477	哈哈_一楼_知道
1	handout_review_commentary_topic	1,596	Comments for experience	Location	6,251	中国_日本_中文_上海
2	owner_branding_customer_business	1,182	Character	Video Game	4,520	辐射_石油_化学_氧气
3	release_mass_update_updatecrazy	1,152	Push for updates	Story content related/Character	2,760	npc_地图_玩家_rbq
4	point_yyyyyyyyyyyyy_then_time	973	Recommendation (pos/neg)/ Comment for experience	Specific story content related	1,420	404_地震_停车场_龙卷风
5	experience_experie_exo_expp	570	Comments for experience/Personal reading experience	Specific story content related	1,191	cy_细胞_懒癌_癌细胞
6	sect_sectact_dragoon_exp	533	Story content related	Story content related/Character	998	gay_前列腺_gaygay_真的
7	daily_average_een_hga	372	Comments for experience/ Reading behavior	Implicit sexual scene	941	证据_开车_怀疑_没有
8	village_town_peasant_hillside	344	Specific story content related	Specific story content related	799	空调_冰箱_变暖_全球
9	entertainment_periodical_magazine_weekly	275	Recommendation (pos/neg)	Specific story content related	523	纳米_硅胶_望远镜_硅基
10	marry_marriage_wedding_bride	258	Specific story content related	Violence	485	暴力_serious_why_so
11	coin_img_currency_gold	174	Recommendation (pos/neg)	Formula	367	名单_枪毙_以下_目标
12	garlic_bacon_cheese_pancetta	135	Specific story content related	Specific story content related	297	scp_基金会_介入_调查
13	health_science_medicine_lose	72	Covid/ Character	Specific story content related	258	wifi_5g_ip_无线
14	chicken_farm_hate_feed	71	Specific story content related/Sensitive Theme	Reading behavior	244	下载_浏览器_打开_订阅
15	reserve_relay_share_responsibility	71	Specific story content related	Specific story content related	231	退休_老年痴呆_痴呆_心好
16	manga_anime_japanese_war	70	Japanese Manga	Specific story content related	220	python_php_语言_最好
17	flash_video_late_remote	68	Video Game/Specific story content related	Formula	134	衬衫_价格_便士_十五
18	death_dead_life_journey	64	Story content related/Character	Formula	133	fbi_open_door_the
19	below_left_middle_look	58	Recommendation (pos/neg)	Specific story content related	131	真空_压缩_密度_flash
20	odor_sound_door_hearse	32	Story content related/Recommendation (pos/neg)	Formula	119	are_you_how_old
21	fabric_complex_mobile_cloth	31	Story content related	Specific story content related	105	next_boy_door_like
22	contrast_color_eye_effect	31	Story content related/Character	Formula	68	惊人_相似_历史_总是
23	mind_philosophy_hand_martial	29	Story content related/Character	Specific story content related	62	java_天下第一_最好_语言
24	honey_chocolate_cereal_milk	22	Gratitude for translator	Specific story content related	40	阿司匹林_asmr_asuna_阿斯匹林
25	abstraction_object_define_whole	16	Comments for experience	Specific story content related	35	世界_版本_打卡_v1
26	plan_read_model_start	15	Bookmark	Formula	26	三年_三天_缝缝补补_我妈
27	trumpet_punish_murderer_sauce	15	Sensitive Theme/Violence	Specific story content related	21	type_new_newtype_阿姆罗
28	ranker_company_firm_rank	12	Advertisement	Specific story content related	13	pm2_超标_pm10_行业协会

Overall, these insights into readers' responses underscore the importance of cultural context and platform-specific dynamics in shaping reader interactions.

6. Limitations and future work

Online platforms offer valuable resources for comparative analysis of how individuals engage with fiction. This corpus provides an important opportunity to observe how readers from diverse cultural backgrounds perceive and interact with fiction in a highly interactive manner. While this study provides valuable insights into reader response in different language settings, several limitations need to be acknowledged. First, we only performed topic modeling on comments, as comments are more directly related to the story and less likely to deviate from the main topic compared to replies to comments, which are the next level of discussion. However, excluding replies means we may have overlooked a richer spectrum of reader response behaviors. Integrating replies into future analyses could provide a more comprehensive understanding of the dynamics and nuances in reader interactions. Second, when performing topic modeling, we set the same number of topics for both language datasets. This uniform approach does not account for the possibility that the inherent number of topics might differ between the two datasets. As a result, setting a fixed number of topics might lead to the aggregation of distinct topics, thereby limiting the completeness of our exploration of reader responses. For

example, great amount of comments was classified to the "Unassigned" the topic. Additionally, the scope of our corpus was largely limited to Male Lead stories –this being the category with more translated stories –which may not fully represent the diversity of online reading across different cultural and social contexts. Lastly, while our analysis aimed to compare reader response across different language settings, it did not explicitly account for cultural differences that might influence these responses.

Future studies will focus on refining topic modeling results for a deeper analysis of topics. We plan to incorporate replies into the model to examine the overall topics in reader responses and conduct close reading to the topic samples including the comments that are in unsigned topics. Additionally, we will add more metrics to the modeling, such as focusing on differences for specific stories when analyzing the topics of comments and replies on both platforms. Business reports [1] on Chinese online stories indicate that more Chinese stories are being translated into English. Consequently, we will continue to expand our parallel corpus to include more platforms and cover a broader range of genres and categories, thereby enhancing the generalizability of our findings. In our effort to compare across cultures rather than just languages, we plan to incorporate cultural context into the analysis. This will involve using a native language identifier for English comments and replies combining with user profile to investigate whether and how cultural backgrounds of non-native English speakers shape reader responses. At the same time, we will take into account the possible impact on the readers' reception of differences between the translated edition and the original novel that may be created in order to improve cultural adaptation.

References

- [1] 2023 *China Online Literature Overseas Trend Report*. https://www.cssn.cn/wx/wx_ttxw/202402/t20240226_5734785.shtml. 2023.
- [2] M. Antoniak and M. Walsh. *The Crowdsourced 'Classics' and the Revealing Limits of Goodreads Data*. 2020.
- [3] C. Aragon and K. Davis. *Learning in large-scale environments: Writers in the Secret Garden, fanfiction, youth, and new forms of mentoring*. With a forew. by C. Fiesler. The MIT Press, 2019.
- [4] G. Cai, F. Sun, and Y. Sha. *Interactive Visualization for Topic Model Curation*. 2018.
- [5] J. Campbell, C. Aragon, K. Davis, S. Evans, A. Evans, and D. Randall. "Thousands of positive reviews: Distributed mentoring in online fan communities". In: *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. Acm, 2016, pp. 691–704.
- [6] Y. Chen, H. Zhang, R. Liu, Z. Ye, and J. Lin. "Experimental explorations on short text topic mining between LDA and NMF based Schemes". In: *Knowledge-Based Systems* 163 (2019), pp. 1–13. URL: <https://doi.org/10.1016/j.knosys.2018.08.011>.
- [7] D. J. A. J. Contreras, H. G. N. Gonzaga, B. M. C. Trovela, and M. A. C. G. Kagaoan. "The "wattyfever": Constructs of Wattpad readers on Wattpad's role in their lives". In: *Laguna Journal of Arts and Sciences Communication Research* 2.1 (2015), pp. 308–327.

- [8] R. Egger and J. Yu. “A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts”. In: *Frontiers in Sociology* 7 (2022), Article 886498. URL: <https://doi.org/10.3389/fsoc.2022.886498>.
- [9] R. Egger and J. Yu. “Identifying hidden semantic structures in Instagram data: a topic modelling comparison”. In: *Tourism Review* 2021 (2021), p. 244. DOI: 10.1108/tr-05-2021-0244.
- [10] S. Evans, K. Davis, A. Evans, J. A. Campbell, D. P. Randall, K. Yin, and C. Aragon. “More than peer production: Fanfiction communities as sites of distributed mentoring”. In: *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 2017, pp. 259–272.
- [11] J. Frens, R. Davis, J. Lee, D. Zhang, and C. Aragon. “Reviews matter: How distributed mentoring predicts lexical diversity on Fanfiction.net”. In: *Proceedings of the 2018 Connected Learning Summit*. Carnegie Mellon University: ETC Press, 2018, pp. 87–97. URL: <https://doi.org/10.48550/arXiv.1809.10268v>.
- [12] M. Grootendorst. *KeyBERT: Minimal keyword extraction with BERT*. Version v0.3.0. 2020. DOI: 10.5281/zenodo.4461265. URL: <https://doi.org/10.5281/zenodo.4461265>.
- [13] M. Grootendorst. *BERTopic: Neural topic modeling with a class-based TF-IDF procedure*. 2022.
- [14] M. Hockx. *Internet Literature in China: I Eat Tomatoes, Coiling Dragon*. New York: Columbia University Press, 2015. URL: <https://www.wuxiaworld.com/novel/coiling-dragon-preview>.
- [15] Y. Hu, T. Underwood, G. Layne-Worthey, and J. S. Downie. *Cross-Cultural Classics: Preliminary Findings from Goodreads Based in the U.S. and Douban Based in China*. 2023. URL: <https://doi.org/10.5281/ZENODO.8107838>.
- [16] Z. Li. “Cross-culture, translation and post-aesthetics: Chinese online literature in/as world literature in the Internet era”. In: *World Literature Studies* 15 (2023), pp. 45–61. URL: <https://doi.org/10.31577/WLS.2023.15.3.5>.
- [17] D. Medvecki, B. Bašaragin, A. Ljajić, and N. Milošević. *Multilingual transformer and BERTopic for short text topic modeling: The case of Serbian*. 2024. URL: <https://doi.org/10.48550/ARXIV.2402.03067>.
- [18] Y. Ouyang and Y. He. “Wangluo wenxue yanjiu de jige xueshu redian [On issues of Internet literature as a research topic]”. In: *Theoretical Studies in Literature and Art* 39.3 (2019), pp. 174–183.
- [19] F. Pianzola. *Digital Social Reading: Sharing Fiction in the 21st Century*. MIT Press, 2021. DOI: 10.1162/ba67f642.a0d97dee.
- [20] F. Pianzola, M. Toccu, and M. Viviani. “Readers’ engagement through digital social reading on Twitter: the TwLetteratura case study”. In: *Library Hi Tech* 40.5 (2022), pp. 1305–1321. DOI: 10.1108/lht-12-2020-0317.

- [21] S. Rebora, P. Boot, F. Pianzola, B. Gasser, J. B. Herrmann, M. Kraxenberger, M. M. Kuijpers, G. Lauer, P. Lendvai, T. C. Messerli, and P. Sorrentino. “Digital humanities and digital social reading”. In: *Digital Scholarship in the Humanities* 36.Supplement_2 (2021), pp. ii230–ii250. DOI: 10.1093/llc/fqab020. URL: <https://doi.org/10.1093/llc/fqab020>.
- [22] X. Ren. “Mapping globalised Chinese webnovels: Genre blending, cultural hybridity, and the complexity of transcultural storytelling”. In: *International Journal of Cultural Studies* 27.3 (2024), pp. 368–386. URL: <https://doi.org/10.1177/13678779231211918>.
- [23] M. Röder, A. Both, and A. Hinneburg. “Exploring the space of topic coherence measures”. In: *Proceedings of the eighth ACM international conference on Web search and data mining*. 2015, pp. 399–408. URL: <http://svn.aksw.org/papers/2015/WSDM%5C%5FTopic%5C%5FEvaluation/public.pdf>.
- [24] M. J. Sánchez-Franco, M. H. G. S. González Serrano, M. A. d. S. dos Santos, and F. C. Moreno. “Modelling the structure of the sports management research field using the BERTopic approach”. In: *Retos: Nuevas tendencias en educación física, deporte y recreación*. 2023, pp. 648–663.
- [25] R. Shang, Z. Xiao, J. Frens, and C. Aragon. “Giving and receiving: Reciprocal review exchange in online fanfiction communities”. In: *Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing*. 2021. URL: <https://doi.org/10.1145/3462204.3481758>.
- [26] Y. Shao, Y. Ji, and Y. Xiao. “The Overseas Circulation of Chinese Online Literature in the Perspective of Media Revolution”. In: *Theory and Criticism of Literature and Art* 38.2 (2018), pp. 119–129.
- [27] R. J. So and G. Wezerek. “Opinion | Just How White Is the Book Industry?” In: *The New York Times* (2020). URL: <https://www.nytimes.com/interactive/2020/12/11/opinion/culture/diversity-publishing-industry.html>.
- [28] Z. Yu, F. Pianzola, and E. Tatar. *Qidian-Webnovel Corpus 110*. 2024. DOI: 10.34894/gqxx3k. URL: <https://doi.org/10.34894/GQXX3K>.
- [29] Y. Zhang and G. Lauer. “Introduction: Cross-Cultural Reading”. In: *Comparative Literature Studies* 54.4 (2017), pp. 693–701. URL: <https://www.muse.jhu.edu/article/680885>.

A. Appendix

a. “It’ s well written for the most part, but it still falls into the ever common trap of most modern setting Chinese novels; it’ s racist, sexist, homophobic and transphobic. The premise is interesting at the beginning, but it just gets boring after a while, following the same plot line with small tweaks of *gain new knowledge* —> *come across wacky situations than can somehow be solved with it* —> *solve problem, but not without spending 10 chapters talking about how everyone underestimates him* and rinse and repeat. I read up to nearly chapter 500, and the quality just decreases. Even if all those things still don’ t put you off, you still don’ t need to read this, I’ m sure you can find some other face slapping system novel to read, god

knows there' s so f***ing many out there, some of which should probably be at least a bit less racist.”

b. “If you are a big racist then you may like this novel. If not then, as 90% of Chinese novel, you will understand that the author is a ***** racist and that he doesn't understand what people think of China.I mean: who doesn't know that China is the country with the highest number of people who ”disappeared” mysteriously and the number 1 in terms of extermination of minorities?”