Using networks to navigate a corpus of reader reviews

An explorative study of absorption expressions in different genres

Master-Thesis
zur Erlangung des Akademischen Grades

Master of Arts

vorgelegt dem Fachbereich 07 – Geschichts- und Kulturwissenschaften
der Johannes Gutenberg-Universität Mainz

von

Tina Ternes

geboren am 30.12.1994 in Wiesbaden

2023

Fach:	Digitale Methodik in den Geistes- und Kulturwissenschaften
Erstgutachter:	Prof. Dr. Gerhard Lauer
Zweitgutachterin:	Prof. Dr. Moniek M. Kuijpers

Abstract

This research demonstrates the value of network analysis as a powerful tool for identifying meaningful clusters within collections of reader reviews. The data basis for this study is the AbsORB corpus, where 1025 reviews from Goodreads were manually annotated for instances of absorption using the Story World Absorption Scale (SWAS).

The investigation focuses on a subcorpus consisting of 199 reading reviews for 49 books across the genres of Fantasy, Romance, Horror/Thriller, Mystery, and Science Fiction. The data is incorporated into a network based on text-similarity as measured by TF-IDF and clustered using the Louvain algorithm. The subsequent qualitative analysis is guided by the examination of keywords and concordances.

Initial attempts at network construction revealed that Fantasy reviews introduced noise into the overall classification of genres. Consequently, these reviews were excluded from the corpus. Although genre is not the primary discriminating factor, Mystery and Romance reviews exhibit greater distinctiveness compared to the other genres. Additionally, the language employed in absorption statements proves to be a significant discriminator during the clustering process, as they are frequently reflected in the keywords associated with each cluster. Furthermore, reviews that deviated from the absorption patterns of the genre in general tend to form separate clusters.

Zusammenfassung

Die vorliegende Arbeit zeigt die Anwendbarkeit von Netzwerkanalysen als leistungsfähiges Werkzeug zur Identifikation aussagekräftiger Cluster in Sammlungen von Leserrezensionen. Die Datenbasis für diese Studie bildet das AbsORB-Korpus, in dem 1025 Rezensionen von Goodreads manuell auf Vorkommnis von Absorption anhand der Story World Absorption Scale (SWAS) annotiert wurden.

Die Untersuchung konzentriert sich auf ein Subkorpus bestehend aus 199 Leserezensionen für 49 Bücher aus den Genres Fantasy, Romance, Horror/Thriller, Mystery und Science Fiction. Die Daten werden in ein Netzwerk integriert, das auf Textähnlichkeit basiert und mithilfe des TF-IDF-Verfahrens berechnet wird. Anschließend werden sie mit dem Louvain-Algorithmus zu Clustern zusammengefasst. Die nachfolgende qualitative Analyse erfolgt anhand der Untersuchung von Keywords und Konkordanzen.

Erste Versuche der Netzwerkkonstruktion zeigten, dass Fantasy-Rezensionen die Gesamtklassifizierung der Genres verfälschten. Daher wurden diese Rezensionen aus dem Korpus ausgeschlossen. Obwohl Genre nicht der primäre Klassifizierungsfaktor ist, differenzieren sich Mystery- und Romance-Rezensionen im Vergleich zu den anderen Genres deutlicher. Darüber hinaus stellen die verwendeten Formulierungen in Absorption Aussagen ein wichtiges Unterscheidungsmerkmal des Clustering-Prozesses dar, da sie sich häufig in den Keywords der jeweiligen Cluster wiederfinden. Darüber hinaus bilden Rezensionen, die von dem Verteilungsmuster von Absorption in ihrem Genre im Allgemeinen abweichen, tendenziell separate Cluster.

Table of contents

1 Introduction

> Yes, a story needs to captivate and be interesting of course. And a book needs to be well written no doubt there. And of course it's not just one component done well that makes a book great. However, if a reader can't loose [sic] themselves in the lives of the characters in a story, if they can't connect with these characters and FEEL along with them; their joy, their fears, their love, their hurt, their emotions, then the read mostly settles into the good category. But it takes GREAT characters to catapult a good book into GREATNESS. (PushingTheLimi00778153)

One of the most devastating moments in my research career was seeing a friend, whom I had asked to take part in an online reading study I designed, closing the tab mid-experiment and returning to his studies. His response was the first of many I would receive that did not at all enjoy the short story that I had carefully selected and analyzed to make sure it contained all the fascinating aspects that would elicit the reader response I wished to study. I could not understand how people would not appreciate the subtle design of the text and the complex emotions it conveyed. I wonder whether the person quoted above would consider the story I chose to be great.

This, however, is an illustrative example of the fundamental influence readers' preferences have on their reading experience. Therefore, rather than expecting one's own perception to be – at least in part – universally applicable, it is essential to take real readers into account if one wants to understand the intricacies of literary reading. Online reader reviews provide a wealth of information giving in depth insight into why readers did – or did not – enjoy a particular text, which parts of the text were considered most important, and how it made them feel. While the diversity in form and content does pose a challenge in systematically investigating reader reviews, this study is based on a corpus that already incorporates a classification of reading experiences. The AbsORB corpus (Kuijpers, Rebora, et al., 2023) provides reviews of novels of various genres from the social reading site Goodreads. This dataset includes manually annotated metadata on absorbing reading experiences in the reviews.

In an endeavor to find out whether texts of different genres elicit a specific response in readers (or readers with specific preferences seek out texts that incorporate these dimensions) this study evaluates the distribution of absorption annotations in a subset of

five genres. Assuming reading experience plays a prominent role in the reviews, they are incorporated into a network visualization based on their similarity in vocabulary. This is to find out whether reviews of the same genre are more closely connected than reviews of different genres. Furthermore, it is hypothesized that a clustering algorithm applied to the network results in meaningful collections of reviews that can be analyzed in terms of genre preferences, absorbing reading experience and style. This is investigated by a qualitative analysis of the resulting clusters.

The thesis is structured as follows: Chapter 2 gives an overview of methods applied in empirical literary studies, specifically reporting on studies investigating reader reviews. Furthermore, the concept of absorption and its adaptation for a corpus of reader reviews is introduced. Chapter 3 describes the dataset investigated in this study as well as the methods that are used to analyze it. Chapter 4 presents the results of the network and qualitative analysis. Finally, chapter 5 concludes the thesis by summarizing the key findings, discussing their implications, and suggesting avenues for future research.

2 Theoretical Background

The recognition of the reader as a subject of literary studies is attributed to the Konstanz school of Reader Response Studies (*Rezeptionsästhetik*) driven by scholars such as Wolfgang Iser, Hans-Robert Jauß, but also to international scholars such as Umberto Eco and Stanley Fish (Salgaro, 2021, p. 532). Building on Hans-Georg Gadamer's theory of the process of comprehension in reading as a dynamic encounter and fusion of the reader's and the text's horizon (Gadamer, 2010), Jauß developed the horizon of expectation (*Erfahrungshorizont*) as an objectifiable reference system modeling the reader (Jauß, 1996, p. 46). In a phenomenology of reading including psycholinguistic insight Iser (1976) coined the "implicit reader". This concept is to function as a structure that is anchored within the text and meant to guide the recipient as wandering viewpoint between the past and future horizons of the text (Iser, 1976, p. 193). With the advancement of linguistics, psychology and cognitive science, the development of reading models is still ongoing in the field of cognitive literary studies[1]. However, in empirical literary studies

[1] One example is the model of enculturated predictive processing, in which a text is understood as a probability design. This has to be navigated by the human brain between top-down predictions and bottom-up

the focus has shifted from abstract model-readers to the study of actual readers or 'extra-textual' datasets of literary reading (Swann & Allington, 2009, p. 247; Whiteley & Canning, 2017, p. 72).

However, this endeavor "tends to take place in two separate and oppositional fields: experimental and naturalistic" (Whiteley & Canning, 2017, p. 78). On the experimental side, reading is observed in controlled settings; evidence of reading experience is often captured explicitly post-hoc by the use of questionnaires or directly using thinking aloud protocols (Hoffstaedter, 1987) and implicitly through neurocognitive and behavioral correlates using eye tracking, fMRI or EEG. Moreover, experimental studies often manipulate the texts used as stimulus material to isolate factors that are hypothesized to contribute to a certain experience (see Jacobs, 2016 for a review of experimental studies and discussion of methods). Within the 'naturalistic study of reading' (Peplow & Carter, 2023, p. 472), evidence of reading experience is gathered outside of the laboratory in contexts such as reading groups, classrooms, social media platforms, and mass reading events. This approach also collects evidence of historical perspectives through sources like diaries, letters and publication records (Whiteley & Canning, 2017, p. 76).[2]

Both approaches contain methodological weaknesses. In naturalistic studies, "the researcher must take reading as it comes" (Swann & Allington, 2009, p. 249), making it impossible to predict what kind of experiences will be found in the data. In contrast, in experimental studies questionnaires and manipulation of the stimulus help isolating the features of interest and guiding the participants in delivering the data desired by the researchers (Whiteley & Canning, 2017, p. 77). "The strength of experimentalist research is that it can give us clear answers to clear questions. The downside is that this clarity is sometimes achieved at the cost of simplifying the phenomena under investigation so far that doubts arise as to the utility of the findings." (Hall, 2008, p. 22) The doubts expressed by Hall regard the topic of ecological validity, questioning whether the findings of experimental studies actually give insight into literary reading as it naturally occurs. Experi-

signals using hierarchical generative models based on bayesian statistics (Fabry & Kukkonen, 2019; Kukkonen, 2020).

[2] See Peplow and Carter (2023) for an extensive discussion of experimental and naturalistic studies and the methods of the naturalistic approach.

mental methods such as applying atypical reading conditions (sentence by sentence reading interrupted by questions or tasks, using decontextualized text-excerpts or even artificial texts constructed specifically for the experiment) (Hall, 2008, p. 31) do not result in insights into literary reading but rather "literary reading *related behaviors*" (Hall, 2008, p. 32). Furthermore, he claims that there is a vast gap between what literary scholars formulating the hypothesis in question perceive in texts and what their participants – "Naive readers" (Hall, 2008, p. 32) – pick up.

On the other hand, naturalistic data, apart from the risk of not containing what the researchers are looking for, often take on unpredictable forms, making them harder to use for quantitative analysis (Swann & Allington, 2009, p. 249). Moreover, the direct influence of the text has to be inferred by the interpretation of the analyst (Whiteley & Canning, 2017, p. 78). However, this type of data has the advantage of not being "restricted to a stereotyped set predetermined by the researcher (e.g. a series of Likert items)" (Swann & Allington, 2009, p. 249) and therefore not posing the problem of ecological validity (Pianzola et al., 2020, p. 2). In summary, "[a] complex field of activity (literary reading) inevitably requires a multiplicity of approaches to deepen our understandings [sic]" (Hall, 2008, p. 22). Furthermore, insights from naturalistic data can be used for theory building that can be incorporated into experimental studies; instruments tested in lab environments can be validated by applying them to data from actual reading environments (Kuijpers, Lusetti, Lendvai, et al., 2023, p. 18).

2.1 Reader Reviews

An illustrative example of naturalistic data sets is to be found in reader reviews, which can be described as "private criticism in the public sphere" (Steiner, 2008, p. 1). Reviews give 'ordinary' readers the opportunity to publicly describe their experience with what they read. This phenomenon has a long tradition since „[p]ersonally written texts about literature or reading recommendations could be found before the internet in letters to the editor and were particularly prominent in subcultural genre magazines and fanzines" (Steiner, 2008, p. 3). Nonetheless, the emergence of online social reading platforms such as Goodreads, LibraryThing and Wattpad or e-commerce platforms (like Amazon) allowing users to publicly review books has made it possible to easily collect vast amounts of data on "non-elite [reading] reception" (Whiteley & Canning, 2017, p. 78). Furthermore,

it enables the investigation of the phenomenon of "digital social reading" (Rebora et al., 2021, p. 231). These sources serve to gain valuable insight into readers and reading communities (Antoniak et al., 2021, p. 294; Pianzola et al., 2020, p. 2).[3]

Obviously, this data has to be studied in context. As Zhang et al. (2019) found in an investigation of Chinese platforms, reviews on social reading platforms tend to be longer and more focused on content-related aspects, while reviews on e-commerce websites are shorter and more likely to comment on logistics and packaging (p. 1170). It is also important to note that social reading sites, like all social media platforms, are a place of self-presentation. We can only see the reading persona people are willing to share with the world (Boot, 2023, p. 4; Driscoll & Rehberg Sedo, 2019, p. 257; Pianzola et al., 2020, p. 34), while at the same time the possibility of staying anonymous creates an environment where people feel comfortable sharing strong emotions and opinions (Steiner, 2008, p. 5). Lastly, one has to keep in mind that even though online book reviews are a widespread phenomenon, readers that write online reviews are only a subset of readers in general. There is no evidence on what distinguishes them from readers who keep their reading experience private or at least out of the internet (Boot, 2023, p. 6).

Studies on reader reviews employ a huge variety of methods. This can be understood as a result of the complexity of the underlying data. The reviews themselves show a wide range of topics mentioned and ways these are addressed, which often necessitates qualitative work. On top of that, the websites from which the data is retrieved offer countless types of metadata on the books discussed as well as the users discussing them.

2.1.1 Close Reading

An extreme case of qualitative work can be found in Kuijpers (2022), where close reading and text world theory are used to analyze a single review from Goodreads, exemplifying the many dimensions (or worlds) reviewers can navigate in discussing their engagement with a novel. Another example of a study focusing purely on close reading is Steiner's (2008) analysis of Amazon reviews on a 'literary' novel as it is characterized in the study and a novel by an established 'chick-lit' writer. The latter sparked special interest because

[3] See Rebora et al., 2021 for a systematic review of directions in studies on digital social reading.

while thematically remaining in the typical area for chick-lit, the style of writing appealed more to the 'literary' audience. The study identifies tensions between professional and non-professional reviewers in the genre of chick-lit, with professionals often dismissing this genre (p. 6) and amateurs calling this out as "elitist and ignorant about the importance of personal reading experiences" (p. 8). The notion of reading experience is a common point of discussion in the reviews of the chick-lit novel. Reviewers of the literary novel on the other hand did not distance themselves from professionals, but rather assimilated their reviews to academic discourse in language and in intertextual remarks, often referring to prestigious classics (p. 9). Thus, the study could show to what extent the book at hand and its context in terms of author and genre has an influence on how reviews are written.

2.1.2 Thematic Coding

Another tactic in tackling the diverse subjects of reviews is thematic coding, which aims at developing annotation schemes to the corpus at hand. These schemes either originate from a bottom-up approach, where common themes are identified in a first analysis of the data and the coding scheme is often refined in several iterations (Allington, 2016; Doche & Ross, 2022; Driscoll & Rehberg Sedo, 2019; Tselenti et al., 2023), or a top-down approach that applies existing schemes or theories to it (Nuttall, 2017; Savolainen, 2019). Of course, mixed approaches are also possible (Nuttall & Harrison, 2020). After applying the annotation schemes, the coded text can be analyzed as it is or alternatively be used further for quantitative analysis. Allington (2016) aims to explore differences between professional reviews in British periodicals and amateur reviews from Amazon on the same novel. By visualizing similarities and differences in a network it was found that a minority of Amazon reviews were similar to those of professional reviewers in periodicals according to the coding scheme (p. 272). This included that amateur reviews used fewer citations and did not praise the humor or the narrative as much, instead being more likely to lament the characters (p. 272). This finding could, on the one hand, be a challenge to the results produced by Steiner (2008), who found that readers of literary novels adopted the language of professionals, as discussed in 2.1.1. On the other hand, the thematic analysis could have captured overarching structures within the review that were not visible during close reading and vice versa. However, Allington's (2016) findings show that even within the discussion of the same book, there are differences in the way

readers express their opinions. This is further explored in a study by Doche & Ross (2022). They not only identified three types of readers in analyzing reviews on a poetry collection, but also found a striking difference between positive and negative reviews on this more demanding type of text: "[T]he most significant difference between positive and negative reviewers resides in their ability to treat difficulty as an affordance rather than a hindrance. Positive reviewers achieve this while negative reviewers do not" (p. 18). Furthermore, they could observe that across all types of readers, there was a strong focus on experience, often combined with personal anecdotes (ibid.).

One of the most exhaustive attempts on thematic coding is found in Driscoll & Rehberg Sedo (2019), who focused on almost 700 reviews from Goodreads on seven titles. They found several features in almost all of the reviews, like evaluation (present in 70 % of the reviews, p. 251), descriptions of temporal, intellectual, emotional and physical reading experience (86 %) (p. 252) – which corresponds to the aforementioned findings of Doche & Ross (2022) – and discussion of characters (46 %, Driscoll & Rehberg Sedo, 2019, p. 256). A less frequent finding was the mentioning of the author (20 %), wherein they found evidence for a greater feeling of familiarity with female authors (p. 255). However, they did not only discuss these broader categories, but also the sub-categories. In the discussion of characters, most readers expressed affection towards them, while only 2 % directly identified with characters and 6 % contained comparisons between characters and the readers themselves or someone they knew (pp. 254-256). In the discussion of reading experience, 68 % of all reviews specifically reported on emotional experience, 25 % of which used love-based language (p. 252); 8.8 % used embodied metaphors to describe their reading experience, while only 4.9 % did mention direct physical experience. Notably, all mentions of physical experience were found in reviews of one novel, namely *The Rosie Project*, for which, additionally, the usage of embodied metaphors was higher than the average with 11.6 % (p. 254). This is particularly noteworthy since Goodreads users ascribe this novel to the romance as well as the chick-lit genre. While genre was not the main concern of this study, Nuttall & Harrison (2020) found a wealth of body metaphors and physical reactions in their analysis of 200 Goodreads reviews on the romance novel *Twilight*. This challenges the claim of emotion on Goodreads tending "to be discussed abstractly, without reference to the body, physicality, or materiality" (Driscoll &

Rehberg Sedo, 2019, p. 257) and leads to the assumption that genre might be an important factor here.

2.1.3 Computational Methods

As discussed before, emotions play an important role in reader reviews, which is why sentiment analysis is frequently used in this context. Sentiment analysis is a term for methods used to evaluate emotive language in a text. This can be achieved either by the use of machine learning or the usage of sentiment lexica. It has been applied to detect the average range of negative and positive sentiments for reviews of certain works (Driscoll & Rehberg Sedo, 2019, p. 253) or to map the sentiment of comments to the sentiment of the text at hand (Pianzola et al., 2020)[4]. Still, there are several problems that have been identified in this regard. Firstly, users of social reading platforms frequently use slang, which is not accounted for by sentiment lexica or might be misinterpreted (ibid, p. 37). Secondly, it has been noted that phrases that in the context of reader reviews are quite low in intensity, like "really enjoyed it," have been given high sentiment values (Driscoll, 2016, p. 253). Therefore, it is recommended to be combined with human interpretation (ibid., p. 257).

Another method which is frequently applied in analyzing reviews is topic modeling, a technique where unsupervised machine learning algorithms are used to extract important themes from texts. These themes appear in a preset number of word-collections, the topics, which then have to be categorized by the researcher. It has been used to gain insight into the collective reading experience to describe the facets that are important to readers of specific genres, like Children's Books (Choi & Joo, 2020) or Classics (Walsh & Antoniak, 2021), but also to investigate differences between genres (Antoniak et al., 2021). While investigating the signature topics of one genre can lead to valuable insights, cross-comparison between genres can help put these findings into context. Analyzing 319,850 reviews on 17,440 books of 20 genres from Librarything, Antoniak et al. (2021) found topics likely to appear across genres. These topics included describing the circumstances of surrounding the acquisition of the book and reading (Topic 0) or numbers

[4] It should be noted that the cited study analyzes Wattpad, where readers can comment on specific passages of the book.

(Topic 2). Others were more genre-specific like romantic relationships (Topic 19) and emotional reaction to characters (Topic 23) for Romance, politics/sociology/religion/psychology (Topic 25) for Science Fiction, series and plot summaries (Topic 22) for Fantasy, movie adaptations and short stories (Topic 6) for Horror or Mystery (Topic 26) for Crime and Mystery (see Figure 1: Distinctiveness of topics for different genres (Antoniak et al., 2021, p. 20)Figure 1).

Figure 1: Distinctiveness of topics for different genres (Antoniak et al., 2021, p. 20)

Another approach to cross genre comparison applied by Antoniak et al. (2021) is training a classifier, a supervised or unsupervised statistical model that is to distinguish genres based on review data. They opted for a supervised approach using a logistic regression model on the review texts, from which they coded the last 100 words of each review to TF-IDF weighted unigram features using the tags that were most frequently assigned to

the books as labels (p. 15). Rather than aiming for maximal predictive performance they were interested in exploring ambiguities and overlaps between genres as well as reviews that seem to differ from typical reviews for the same genre, assigning a surprisal score to each review to indicate the probability of a mismatch. They found that the average surprisal score was highest for the genres Young Adult, Family, Classics, Children, and Fantasy, attributing this to either a high similarity between those genres or to the genres encapsulating a broad range of themes and discourses (p. 16-17). Another possible explanation is the community homogeneity within a genre. To measure similarity in reading habits they analyzed the genre tags found on the individual user profiles, which represent the genres that were frequently assigned to books by the respective reader. Doing this, they found that some genres like Romance, Vampires and Horror are more likely to attract similar reviewers than others. This leads to the assumption "that these tight-knit communities have higher shared standards for their ratings." (p. 22). The trained classifiers approach was also applied by Chang et al. (2020) who used historic reviews from the British Periodicals Collection dating from 1800 to 1950. Instead of focusing on review communities, they found a correlation between their classifier that was trained for distinguishing books of different genres and the classification of reviews, concluding that "[c]losely-knit genres also produce clusters of closely-related reviews" (p. 7).

Based on the previous observations, it is obvious that gaining insight into the machinations of reading experience as seen in reader reviews often requires the combination of a set of approaches. This entails working with several corpus-related methods such as word frequencies, keyword analysis, collocations and concordance analysis. For example, in a corpus of suspense and 'literary' novels in Dutch and English, focusing on the collocations of 'plot' and 'beautiful', it could be shown that the reviews demonstrated elaborate discussions on the narration and other aspects of literary narration (Boot, 2023, p. 19). This is particularly noteworthy taking into concern the criticism of online reviews to focus primarily on personal reading experience. Upon further investigation, the same study could also identify differences between suspense and 'literary' readers. Collocations of 'beautiful' indicated a preference for characters, plot and the book's materiality (e.g. the cover) within suspense readers, while the collocations in the 'literary' subcorpus consisted mainly of the story and stylistic aspects (ibid., p. 11, 19). Furthermore, a concordance analysis, where words are displayed in context, revealed that body parts play

an important role in the reviews, especially in the description of positive reading experiences (ibid., p. 13-14). They conducted a keyness analysis to investigate statically distinctive words or part of speech-tags that are used in positive reviews as compared to negative reviews and vice versa. This allowed them to show that positive reviews mostly use present tense and third person verbs to report on events in the story and praise the author (ibid., p. 16), while negative reviews are more likely to use first person and past tense, explaining their negative reaction to the book from their perspective (ibid., p. 20). Another study examined 472,810 reviews in Dutch from Goodreads as well as several Dutch book reviewing platforms, investigating review length and the occurrence of subjectivity in different genres including Fantasy, Historical Fiction, Literature, Literary Thriller, Regional, Romance, Science Fiction, Suspense and Youth Literature. It was found out that Science Fiction reviews in their corpus were significantly longer than those on other genres (Koolen et al., 2020, p. 155). Furthermore, subjectivity as indicated by the frequency of first and second person singular pronouns was most prominent in readers of Fantasy and Youth Literature (ibid., p. 156). However, taking into account the individual book, reviewer and platform, they identified that these factors heavily influence length and subjectivity, finding that frequent reviewers adopt or even create genre conventions since their reviews are more consistent with the average (ibid., pp. 156-157, 165-166).

Lastly, network analysis has proven to be a versatile tool in the investigation of online social reading. It has been used to map the social interactions of users or the interaction of users with different books and chapters on Wattpad (Pianzola et al., 2020) to investigate the differences between professional and non-professional reviewers (Allington, 2016; see 2.1.2) and to model the plot of novels on the basis of reader reviews (Holur et al., 2021). Networks are a mode of data visualization, which in core contain two elements: nodes and edges. Nodes can represent any kind of entity (e.g. persons, organizations, books, words etc.), while edges display the relationships between them. Edges can be directed (e.g. "Franz sends a letter to Hermann") or undirected (e.g. "*The Metamorphosis* and *A Hunger Artist* are both works by Franz Kafka"). Additionally, edges can be weighted (e.g. "*The Metamorphosis* and *A Hunger Artist* have 40 % shared vocabulary"; "*The Metamorphosis* and *The Earthquake in Chile* have 4% shared vocabulary"). Holur et al. (2021)

constructed a directed network in which nodes represented characters, which were mentioned in the reviews, as well as other metadiscursive or extradiegetic actants, and the edges were verb phrases connecting these actants. In doing that, Holur et al. did not only gain insight into how stories are retold by a collective readership, but also which comparisons are drawn across novels (p. 25).

2.2 Absorption in Reader Reviews

„Narrative absorption is an umbrella term for a specific mental state that, in general, can be described as altered embodied, cognitive, and emotional processes of being invested into the content of a fictional story“ (Kuijpers et al., 2021, p. 279). There are several approaches in empirical literary studies that aim to measure this phenomenon via self report scales concentrating on different dimensions (ibid., p. 284). Instruments developed in this field include the Narrative Engagement Scale (Busselle & Bilandzic, 2009), the Absorption-like States Questionnaire (ASQ) (Kuiken & Douglas, 2017), the Reading Flow Short Scale (RFSS) (Thissen et al., 2018) and the Story World Absorption Scale (SWAS) (Kuijpers et al., 2014). However, all of these overlap in capturing some forms of *Attention*, *Mental Imagery*, *Emotional Engagement* and *Transportation* (Kuijpers et al., 2021, p. 284).

The underlying concept of this study is the SWAS, which comprises all of the aforementioned dimensions. *Attention* is hereby conceptualized as the felt experience of effortlessly achieved deep concentration. *Mental Imagery* refers to the visual representations that emerge in one's mind while reading a story, encompassing the reader's visualizations of surroundings, characters, and situations. *Emotional engagement* entails feelings directed towards or shared with characters, such as sympathy, empathy, and identification. This dimension specifically focuses on the connection established with characters and does not account for overall emotional responses to the entire story, such as perceiving the story as emotional or eliciting a personal emotional response. Lastly, *Transportation* denotes the sensation of entering a story world while simultaneously maintaining a connection with the real world (Kuijpers et al., 2014, pp. 91–93).

In the course of the “Mining Goodreads” project (Rebora et al., 2020), the SWAS was reconceptualized to be used as a tagset to annotate reader reviews from Goodreads. This included simplifying the statements to adapt them to the language used on the platform.

For example, "the original A5 statement 'I was reading in such a concentrated manner that I had forgotten the world around me' [was changed] by splitting it into two statements, namely A2 'My attention was focused on the book' and A5 'While reading I forgot the world around me'" (Kuijpers, Lusetti, Lendvai, et al., 2023, p. 5). Furthermore, while the original SWAS was designed to capture absorption from reading a short story or excerpt of a novel, reviews are discussing entire books or even book series. Therefore, a new tag was introduced to accommodate for the anticipation of story events outside the novel reviewed. Since there were also other findings in the reviews indicating absorbing reading experiences, but were up to this point not covered by the SWAS, it was also adapted to include these. Examples for added tags are "'lingering story feelings' ('It is one of those stories that just sticks with you') and 'addiction' ('I could not get enough of their storyline')" (ibid., p.4) or portrayed an expansion to existing dimensions, "for example, negatively valenced emotional engagement such as anger and fear ('During her chase, I became afraid for her')" (ibid.). The former additions were first collected as SWAS-Related statements, later they were either incorporated into one of the existing dimensions or into the newly introduced dimension of Impact (ibid., 10). Oftentimes, these findings could be related to existing concepts in absorption research, such as the work of Bálint et al. (2016). Additionally, the higher order category of SWAS-Mention was introduced to account for statements that include absorption, but don't refer to the specific book reviewed, but were reports on general reading experience (ibid., p. 6).

Table 1 shows the expanded conceptual model of the SWAS resulting from the annotation process. The items highlighted in green are the modified versions of the original SWAS statements, orange highlighting indicates concepts taken from the absorption inventory of Bálint et al. (2016), and the blue items mark additions based on the findings during the annotation process (ibid., pp. 11-12).

This endeavor resulted in a corpus of 493 curated reviews (Kuijpers, Rebora, et al., 2023).

Absorption dimension	Absorption category	Number of total annotations	Absorption Present	Absorption Negated
Attention	A1 (Altered sense of time): While reading time moved differently	3	3	0
	A2 (Concentration): My attention was focused on the book	12	10	2
	A3 (General sense of absorption): *I was absorbed in the book*	194	186	8

	A4 (No distractions): I was not distracted while reading	5	3	2
	A5 (Forgetting surroundings): While reading I forgot the world around me	20	16	0
	A6 (Anticipation): I was on the edge of my seat / I wanted to know what would happen next	111	108	3
	A7 (Inability to stop reading): I did not want to put the book down/ I could not put the book down	150	145	5
Emotional Engagement	EE1 (Perspective taking): I could imagine what it must be like to be this character	35	35	0
	EE2 (Sympathy): I sympathized with this character	57	53	4
	EE3 (Emotional connection): I felt a connection to this character	79	69	10
	EE4 (Empathy): I felt how this character was feeling	73	72	1
	EE5 (Compassion for story events): I felt for what happened in the story	79	79	0
	EE6 (Anger): I felt angry at this character	20	19	1
	EE7 (Fear): I felt scared for this character	5	5	0
	EE8 (Emotional familiarity): I felt like I knew this character	11	11	0
	EE9 (Wishful identification): I wish I could be more like this character	8	8	0
	EE10 (Emotional understanding): I understood why this character did this	31	26	5
	EE11 (Parasocial response): I want to have some kind of relationship with this character	79	79	0
	EE12 (Participatory response): I wanted to involve myself in the story world events	42	42	0
Mental Imagery	MS1 (Imagery of character): I could imagine what the characters looked/smelled/felt/sounded like	18	15	3
	MS2 (Imagery of story events): I could see/hear/feel/smell the story events clearly in my mind	20	20	0
	MS3 (Imagery of story world): I could imagine what the story world looked/smelled/felt/sounded like	27	25	2
	MS4 (Realness): The character/story world felt real to me	73	73	0
Transportation	T1 (Presence): While reading this I was in the story world	13	13	0
	T2 (Merge of fiction in reality): Elements from the story world came into my world	14	14	0
	T3 (Proximity of story world): The story world felt close to me	4	4	0
	T4 (Deictic shift): I felt transported to the story world	19	19	0
	T5 (Part of the story world): *I felt part of the story world*	34	34	0
	T6 (Return deictic shift): I returned from a trip to the story world	3	3	0
	T7 (Travel in story world): I lost myself in the story world / I traveled with the characters through the story world	26	26	0
Impact	IM1 (Effortless engagement): It was an easy read / I devoured this book	108	68	40
	IM2 (Wish to reread): I will/have reread this book/parts of this book	116	112	4
	IM3 (Anticipation book series): I cannot wait to see how this unfolds in the next book	170	167	3

	IM4 (Addiction): I am addicted to this book / I cannot get enough of this book	91	89	2
	IM5 (Lingering story feelings): The book left me feeling … / This book stayed with me for a while	194	192	2

Table 1: Absorption dimensions for the annotation of reader reviews

3 Methods

3.1 Data

3.1.1 Corpus building

The annotation data obtained from INCEpTION (Klie et al., 2018) in XMI-Format was reordered by an updated version of the script from the Mining Goodreads Project (Rebora et al., 2020) to a text file, containing the review itself and a table, with the annotation data, as can be seen in Table 2. Tables like this were generated for every review, which was annotated. For this analysis, only the curated dataset was included, as well as just those reviews that actually contained annotations. The data was restructured in R (R Core Team, 2023; Wickham, 2022; Wickham et al., 2019; Wickham, François, et al., 2022; Wickham, Hester, et al., 2022; Wickham & Girlich, 2022), whereby the information contained in the file name and in the first column as seen in Table 2 was split into individual columns. The file name consists of a link snippet, the identifier of the book's title-page on Goodreads, the book title, and an unique identifier for the review (see Figure 2).

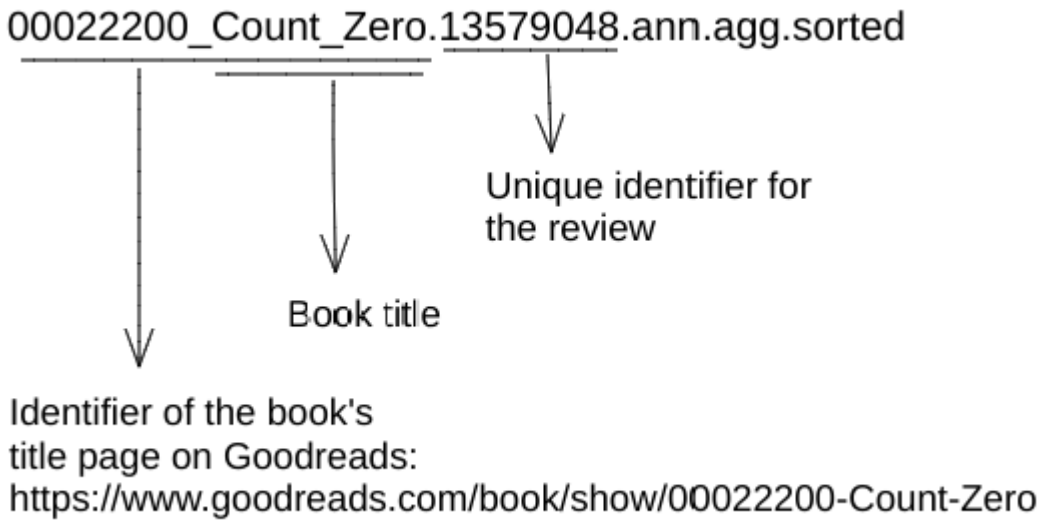

Figure 2: Splitting of the file name

The first column of the file contains information about the annotation. While the numbers represent the character index of the on- and offset of the annotated part inside the review as a whole, the rest of the information consists of three to four parts, as seen in the first row compared to the second and third row in Figure 1. The presence of the term “Mention” indicates “a non-review specific instance of absorption” (Kuijpers, Lusetti, Renner, et al., 2023, p. 7), where users report on their reading experiences outside of the work currently discussed. This is recorded in the data table in the column heading "mode" as either "Absorption" or "Absorption Mention" (e.g. “I just lose myself in the story” versus: “I love a story that I can lose myself in”). The next item is “SWASRelated” in rows one and two or “SWASSpecific” in row three, where “SWAS-Specific” describes items that were directly derived from the SWAS (Kuijpers et al., 2014). According to Kuijpers, Lusetti, Renner et al. (2023, p. 6) “[t]he "SWAS-Related" category includes tags for aspects that can conceptually be counted as part of a narrative absorption experience, but that were not reflected in the statements on the Story World Absorption Scale”. Thereupon follows the tag, which in “SWAS-specific” cases is represented by an alphanumeric code indicating the SWAS-dimension and number of items and in “SWAS-related” cases the concept of the tag.

Mention_SWASRelated_ParasocialResponse_Present 422 476	Now he's an old friend I just like spending time with.	@cura
SWASRelated_Realness_Present 1611 1650	the future it paints is almost upon us.	@cura
SWASSpecific_T2_Present 1611 1650	the future it paints is almost upon us.	@cura

Table 2: Example of a table containing annotation Data for a review

However, the aim of the preparation of this corpus was the validation of the SWAS and entailed a reconceptualization, which now integrates the “SWAS-Specific” items including a new dimension of the scale. For this purpose, the data table was structured to reflect this rather than the specific/related categorization (Kuijpers, Lusetti, Lendvai, et al., 2023, p. 11). Therefore, the data table utilizes the columns “SWAS_category”, in which the SWAS-dimension is specified, and “SWAS_tag”, which contains the alphanumeric code and the concept of the tag. Lastly, the word “Present” or, in other cases, “Absent” indicates

whether the statement is absorption positive or negative (e.g. “I couldn't get into it.” versus: “this one is a real page-turner”).

In addition to the annotation, further information was added from the text files; specifically the annotation round and the full review itself. The first is specified in the file name of the full review (e.g. round eleven in “11.00022200_Count_Zero.13579048”), which like the table also includes the identification number of the review, so that these could easily be put together.

Since the definition and demarcation of genres constitutes a highly controversial discussion of its own in literary studies, it was decided during the preparation of the corpus to base this on the assessment of the users of Goodreads. Within the platform, users have the option to assign a book to one or more genres. This book then appears on lists of these genres, ranked according to the number of votes. However, despite the decision to be guided by user votes, this results in several approaches to defining genres, some of these being:

1. A book ranks high on a list and is therefore particularly representative of this genre.[5]
2. A book has been assigned to a certain genre by most users and thus corresponds to it.
3. A book has been mostly assigned to genre X and secondly to genre Y, which is a specification of genre X, and should therefore be assigned to the more specific/generic genre (X=Fantasy, Y=High Fantasy).
4. A book received the most votes for genre X, closely followed by the thematically distinct genre Y (X=Thriller, Y=Fantasy) with a large distance to the following genres Z and Q (X=10,000, Y=9,000, Z=2,000, Q=1,500). Therefore, it should be assigned to genres X and Y.

[5] One example of this approach can be found in Antoniak et al. (2021, p. 10), where they act upon the following reasoning: “For example, a top book for the tag fantasy might also be a top book for the tag science-fiction. Even if the book is tagged science-fiction more often than fantasy, we will still add the book to the fantasy genre if its fantasy ranking is in the top 1,000. In other words, the top books are the most popular books for the tag, not the books most specific to the tag.” (Antoniak et al., 2021, p. 10)

The information about the genre already available in the data set was driven by lists sourced from Goodreads. In the available dataset, the genre information was presented in tabular form, as illustrated in Table 3. As can be seen, this presented a challenge to the intended investigation of genre distinctions, given the considerable overlaps and absence of a hierarchical arrangement.

title	fantasy	thriller	romance	classics	contem-porary	science fiction	horror	mystery	historical fiction
Eragon	x								
ameri-can_gods	x				x	x	x		
graceling	x		x						
Good_Omens	x				x	x			
New_Moon	x		x		x		x		
Eclipse	x		x		x		x		
Prince_Caspian	x			x					

Table 3: Genre information from the original dataset

Therefore, a new approach was chosen for this investigation, corresponding to 2 of the list above, which involved scraping the information directly from Goodreads. Since the URL paths to the presentation pages of the books were accessible in the available dataset, it allowed for a straightforward implementation of a web scraping program. The program was designed to extract the top three most-voted genres for each book, along with their respective vote counts to ensure that this information would be available for future investigations seeking to follow strategies 3 or 4. Additionally, this method was devised so as to allow for a more nuanced selection of books as it enabled avoiding overlaps between sub-corpora.

However, at the time of the investigation, the Goodreads website was in a transitional phase between two versions, resulting in occasional display of the old or new version of the site. As a consequence, when iterating over pages, it was often the case that data could not be extracted for several titles. Even modifying the code to accommodate the new site did not improve the situation. Due to time constraints, it was necessary to adopt a pragmatic solution, whereby the scraping code was reapplied to the URL paths that had yielded no results in the previous rounds until data had been collected for all titles. The

data of the three highest-rated genres as well as their respective vote count was added in corresponding columns to the data table.

3.1.2 Selection of Sub Corpora

Overview: All genres

Genre	Books (n)	Reviews (n)	Annotations (n)
Fantasy	24	109	535
Fiction	13	44	129
Young Adult	15	69	441
Thriller	2	7	27
Historical: Historical Fiction	3	14	74
Classics	2	9	35
Science Fiction	9	32	91
Childrens	1	4	12
Horror	7	30	111
Mystery	9	31	80
Romance	9	41	251
Romance: Paranormal Romance	1	3	8
Romance: Historical Romance	2	9	31
Contemporary	1	5	14
Fantasy: Paranormal	1	5	38

Table 4: Genre representation in the overall corpus (see Cuilla, 2022 & Lin, 2023 for table creation)

Table 4 shows an overview of the different genres and their representation in the corpus. As we can see, there are a few parameters to consider when choosing the genres to be included into the analysis. However, the data included in this analysis only contains those reviews that were tagged at least once for direct and positive absorption. Furthermore, only five reviews per book were included in the annotation process, thereby the range of the number of reviews per book is constituted as one to five. This means that the number of books in the genre does not necessarily guarantee the number of reviews, let alone annotations.

While Fantasy, Fiction and Young Adult are the genres best represented in the corpus, they are also quite ill-defined. As can be seen in Figure 3, the new releases on Goodreads tagged for these genres overlap substantially. Therefore, it was decided not to include Fiction and Young Adult in the analysis. Fantasy seems to be more narrowly definable, since it can be seen as a more specific category within Fiction as well as Young Adult.

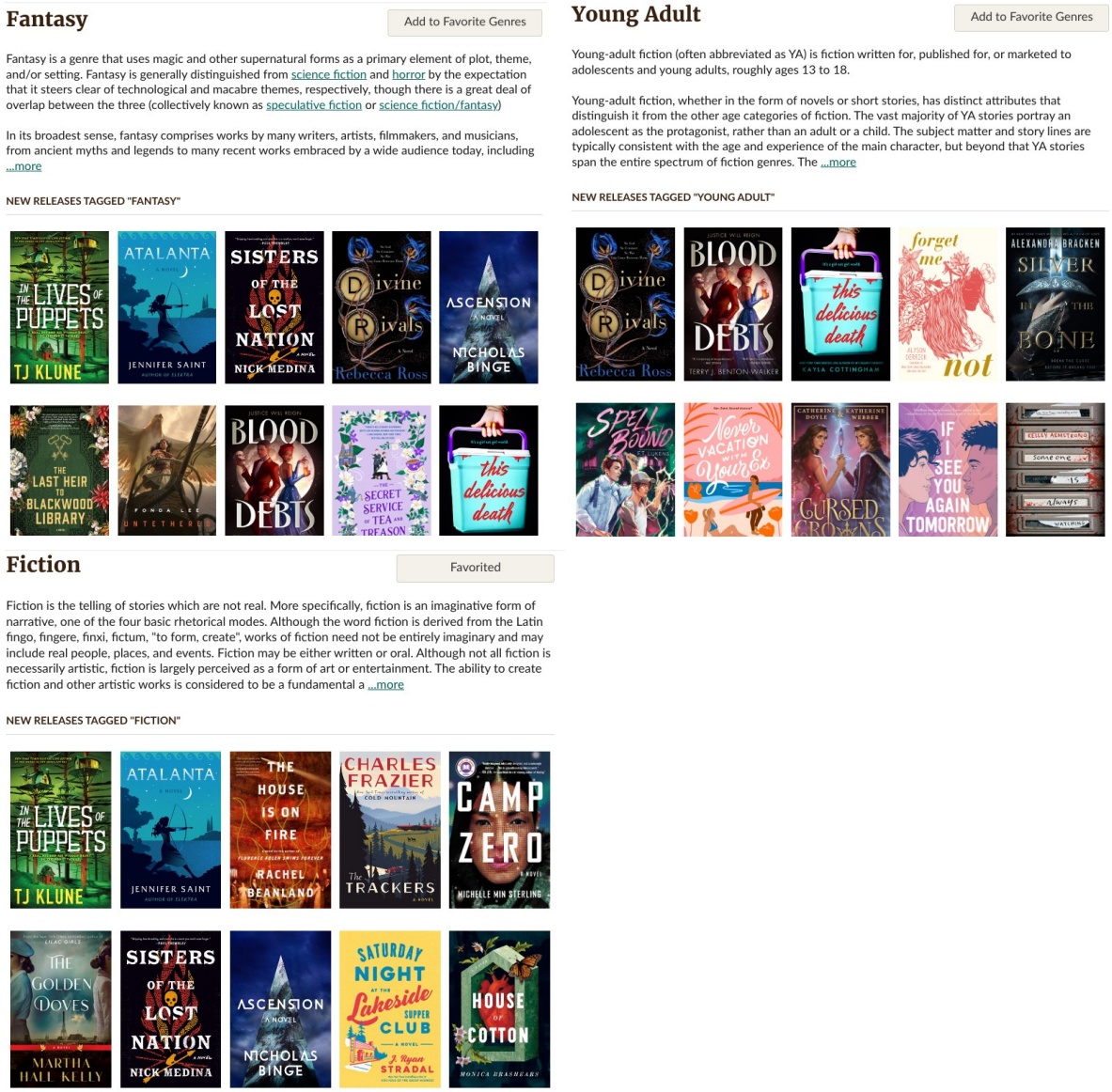

Figure 3: new release pages for genres on Goodreads (Fantasy Books | Goodreads, 2023; Fiction Books | Goodreads, 2023; Young Adult Books | Goodreads, 2023)

The second biggest genres in the corpus are Science Fiction, Horror, Mystery and Romance. Since those are clearly definable and distinguishable, they were chosen to be included in the analysis. Moreover, Horror is represented by fewer books than the other genres and is expected to have a similar effect on readers as the Thriller genre. Hence,

these genres were merged, thereby creating the same book count as in Science Fiction, Mystery and Romance.

Overview: Selected genres

Genre	Books (n)	Reviews (n)	Annotations (n)
Fantasy	24	109	535
Romance	9	41	251
Horror/ Thriller	9	37	138
Science Fiction	9	32	91
Mystery	9	31	80

Filtered by subgenres

Genre	Books (n)	Reviews (n)	Annotations (n)
Fantasy	13	58	263
Romance	9	41	251
Horror/ Thriller	9	37	138
Science Fiction	9	32	91
Mystery	9	31	80

Table 5: Genre representation in the subcorpus before (left) and after subgenre exclusion (right)

As can be seen in Table 5, the Fantasy subcorpus is much larger than the other genre corpora. To re-balance this, it was filtered to exclude books that are attributed to one of the other genres in the second or third genre ascription. Thereby, the number of books was decreased to 13 and the number of corresponding reviews to 58. While still being bigger than the other corpora, this decision was deemed sufficient to balance out the overall corpus. This reduction process was also performed for Romance, but there seemed to be no overlap with the other genres under investigation. What can be noted about the resultant corpus is the number of annotations. Romance seems to have the highest density of annotations, followed by Fantasy and Horror/Thriller, whereas Mystery and Science Fiction appear to have fewer instances of absorbed reading in them. Thus, in the following chapter, there will be a closer investigation of absorption distribution in different genres.

3.1.3 Preliminary Analysis

By analyzing the distribution of tokens in the analyzed subcorpora (Table 6) we can discern first clues as to differences in reviews of different genres. Firstly, Romance and Fantasy seem to include longer reviews than the other genres, with Fantasy having the greatest range of review-length and Mystery the smallest.

Tokeninfo Genre Corpora

Genre	Tokens (n)	Tokens (mean)	Tokens (median)	Tokens (min)	Tokens (max)	Reviews (n)
Fantasy	28602	493.14	505.5	109	896	58
Romance	24435	595.98	631	154	888	41
Horror/ Thriller	12196	329.62	203	108	852	37
Science Fiction	8969	280.28	164.5	115	927	32
Mystery	6083	196.23	173	113	624	31

Table 6: Tokeninformation on the genrecorpora; tokenization was conducted with quanteda (Benoit et al., 2018)

Figure 4 shows the frequency distribution of the SWAS-dimensions in the analyzed genres. Impact seems to encompass the highest frequency of annotations (at least 30% in every genre), while Mental Imagery and Transportation seem to be phenomena that are encountered more seldomly. Furthermore, the Romance genre has distinctively more annotations in the dimension of Emotional Engagement than any other genre and fewer annotations in the dimension of Attention. Another genre with a very distinct pattern in the Absorption-dimensions is Science Fiction with almost half of its annotations belonging to the Impact dimension, which is more than any of the other genres. This is followed by the Attention dimension with almost 30% of annotations and a lower score for the rest of the dimensions, compared to the other genres (particularly for Emotional Engagement and Mental Imagery).

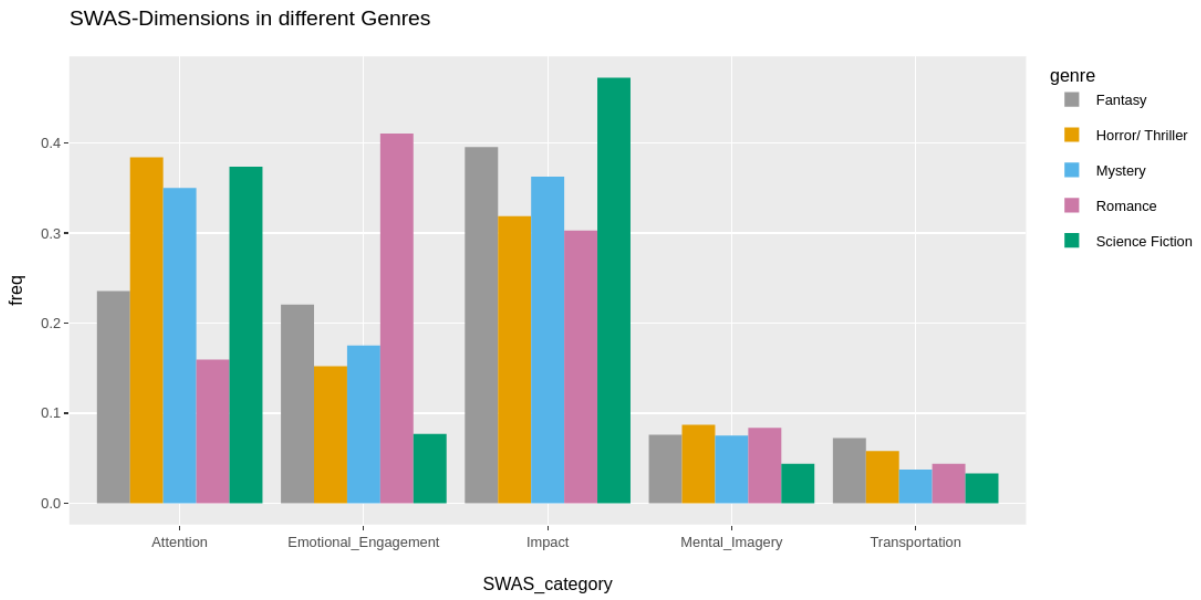

Figure 4: Percentage of tags inside the different genres, plots generated with ggplot2 and plotly (Sievert, 2020; Wickham, 2016)

Horror/Thriller and Mystery seem to follow a similar pattern. This might be due to overlaps in second or third genre, as well as being thematically closer to each other than the other genres as they both might include themes like following an investigation or crimes. This could trigger similar Absorption-responses in readers. They score notably high in the Attention-dimension and low in Emotional Engagement. Thereby, Horror/Thriller scores slightly higher in Attention, Mental Imagery – where it is marginally more frequent than the rest of the genres – and Transportation, while Mystery is more prevalent in Emotional Engagement and Impact. Fantasy scores relatively high in Impact and Emotional Engagement: it has the highest frequency of Transportation annotations and a comparably low frequency in Attention.

It should be noted that the data shown here encompasses all annotations for the respective dimensions including those tagged as Mention or Absent. This is justified by those instances being only a small fraction of the Annotations and by the assumption that recognizing a lack of a certain experience hints at the importance of said experience to the genre. Furthermore, one should keep in mind that the frequencies shown here are derived from data of varying quantity, such as the number of Fantasy annotations being threefold the number of Science Fiction annotations.

3.2 Network Construction

As discussed in 2.1.3, network analysis has been used to study several aspects of online social reading. In the present research, it is used to investigate underlying linguistic structures in reviews. For this purpose, nodes represent the individual reviews and edges show whether a pair of reviews includes similar features. This approach has been applied in stylometry for purposes such as authorship-attribution, author gender, time, genre and narrative perspective (Eder, 2017; Herrmann, 2018; Jockers, 2013; Päpcke et al., 2022) and specific measures, such as Burrow's Delta and Eder's Delta, have been developed for the purposes stated before (Jannidis et al., 2015). However, these measures are designed for the analysis of a completely different text-genre than reader reviews and rely on the usage of most frequent words (mfw). In these contexts they show peak performance at amounts of 1000-1500 mfw (ibid., p. 3), which is easily achieved in the analysis of novels, but more than the amount of tokens of the longest review in this corpus.

3.2.1 Text similarity

Text similarity in this study is operationalized as *term frequency-inverse document frequency* (TF-IDF), which is a combinational measure that incorporates *term frequency* and *inverse document frequency*. *Term frequency* is used to assign weights to each term in a document, based on the number of occurrences. Thus, each term *t* in document *d* is assigned a weight in accordance to the number of times *t* appears in *d*, resulting in the term frequency $tf_{t,d}$. Within this weighting scheme, a document is defined as a *bag of words*, meaning that the order in which the terms appear is ignored. However, since all words are considered equally relevant in raw frequency, the weight doesn't reveal much about which terms have discriminating power in a collection of documents. The *document frequency* (df_t) on the other hand, is defined as the number of documents in a collection that include a specific term *t*. To use document frequency as a weight, *inverse document frequency* is calculated by taking the logarithm of the ratio between the total number of documents in a collection *N* and the document frequency of term *t*:

$$\mathrm{idf}_t = \log \frac{N}{\mathrm{df}_t}$$ (Manning et al., 2008, p. 108).

This gives higher weight for rare terms, whereas more common terms receive a lower score. To combine term frequency and inverse document frequency, TF-IDF is applied: $\text{tf-idf}_{t,d} = \text{tf}_{t,d} \times \text{idf}_t$. Thus, the weight assigned by TF-IDF to term *t* in a document *d* is

1. highest, when tf_t is high and df_t is low, which means that *t* carries significant discriminative power for the subset of documents it appears in.
2. lower, when either tf_t is low or df_t is high, which suggests that *t* is less relevant, because it is either less prominent within a document or less distinctive across the entire collection.
3. lowest, when df_t is highest. In this case, the term provides little differentiating value as it is ubiquitous throughout the collection (ibid., p. 109)

In essence, TF-IDF captures the relative importance of a term within a document by considering both its frequency within the document and its occurrence across the entire collection. Having established that, each document can be seen as a vector of terms with their corresponding weights (ibid.).

There are several ways to quantify the similarity of two documents d_1 and d_2 or their vector representations $\vec{V}(d_1)$ and $\vec{V}(d_2)$. The standard way, which is applied in this study, is to compute the *cosine similarity*:

$$\text{sim}(d_1, d_2) = \frac{\vec{V}(d_1) \cdot \vec{V}(d_2)}{|\vec{V}(d_1)||\vec{V}(d_2)|}$$ (ibid., p. 111).

Thereby, the numerator represents the dot product of $\vec{V}(d_1)$ and $\vec{V}(d_2)$, while the denominator is calculated as the product of their *Euclidean lengths.* The latter length-normalizes the vectors, since two documents can be very similar in content, but differ significantly in length.

3.2.2 Data Cleaning

As previously discussed, the connections between nodes in networks rely heavily on the significance a term has for a review as compared to the other documents in the corpus. Accordingly, proper nouns, such as names of authors and characters, titles or other story-specific terms, create a huge bias for connections between reviews for the same book or books of the same author. Therefore, these were stripped from the data. Since doing this

by hand would have taken a huge amount of time, a semi-automatic approach was applied.

The first step consisted of applying *Named Entity Recognition* – a natural language processing technique that identifies and categorizes named entities, such as persons, locations, organizations, and dates, within a given text – using SpaCyr (Benoit & Matsuo, 2020) to the review data. After that, terms labeled as "WORK_OF_ART", "PERSON", "GPE" (countries, cities, states), "ORG" (companies, agencies, institutions, etc.), "PRODUCT" (objects, vehicles, foods, etc.), "LAW" (named documents made into laws), "FAC" (buildings, airports, bridges, etc.), "MONEY", and "LOC" (non-GPE locations, mountain ranges, bodies of water)[6] were investigated to see whether they matched terms that should be excluded from the analysis. Unfortunately, every label contained several mismatches that needed to be filtered by hand. Moreover, there were instances in which the recognition was inconsistent, for example correctly labeling a character in some reviews, but not all. For instance, although "Hermione" was correctly labeled as a person in certain instances, it was occasionally overlooked. To address this, an approach was implemented to ensure consistent labeling within the same group of reviews that discussed the same book. This involved assigning the same label to any term that had been correctly labeled at least once within the book-specific review group. Since there were frequent issues with the book titles, words that appeared in the title of each book were excluded from the corresponding reviews. Occasionally, during subsequent stages of the analysis, it became evident that certain words were missed during the initial cleaning process. These were then separately excluded.

Finally, this study did not incorporate a pre-existing stopword list, but certain words were deemed biasing or unhelpful for the analysis. This included third person pronouns to prevent connections that were solely based on the gender of either the protagonist or the author or the number of protagonists. Other terms excluded are "the," "a," "an," "of," "or," "and," "to," "ll," and "ve".

[6] A list of entity-labels and examples was obtained from (Dasaprakash & Shaikh, 2019)

3.2.3 Graph creation

The cleaned data was organized into a *document feature matrix* using the quanteda package (Benoit et al., 2018). In this matrix, each row corresponds to a document, in this case the reviews, and each lemma in the corpus has its own column. The values in the matrix represent the frequency of occurrence for each word within the respective document. Additionally, the package enables the inclusion of document variables, such as review, title, and genre.

Following the creation of the document feature matrix, TF-IDF was calculated and computed to a matrix of distances and similarities between documents based on cosine similarity. The resulting *similarity matrix* is in itself a complete weighted network, since it displays the connections between every document. However, in this form, the network is quite dense, which is why a filtering code is employed to extract the network backbone. The filtering code for this study was adopted from Bail (2016), which itself is based on the disparity filter by Bessi and Briatte (2016). It allows to preserve "an edge whenever its intensity is statistically not compatible with respect to a null hypothesis of uniform randomness for at least one of the two nodes the edge is incident to" (Serrano et al., 2009, p. 6487). The statistical threshold (alpha) is thereby defined by the user. In this case, alpha was set to 0.085, which does not entail the standard threshold of statistical significance, but was informed by the resulting network. Setting the threshold to 0.05 resulted in very small clusters, which quite consistently paired nodes corresponding to the same novel, but consequently resulted in a huge loss of data in an already small dataset. Due to the explorative nature of this study that does not aim at statistical representativeness, it was assumed that even these looser connections could be informative and give directions for future studies using a bigger dataset.

To gain comprehensive information from a network structure, it is useful to decompose it into sub-units, so called communities or clusters. In the case of information networks, such as the present one, these can be used to identify common topics (Blondel et al., 2008, p. 11). The community-detection method used in this study is the *louvain algorithm* as implemented in the igraph package (Csardi & Nepusz, 2006). Here, the core measure for community detection is *modularity*, which is based on the level of partitioning or division strength within a network. The algorithm tries to maximize modularity in two iterative phases:

First Phase:

1. Assign each node *i* in a network its own community
2. Calculate the modularity for each node *i* if it was removed from its community and moved into the community of each neighbor *j* of *i*
3. Place *i* into the community that resulted in the greatest increase in modularity, if no increase is possible *i* remains in its original community
4. Repeat 1. to 3. until no more increase in modularity occurs

Second Phase:

1. A new meta-network is built whose nodes are the communities of the previous phase
2. The edges between the former nodes of the same community are now self-loops and the edges between communities are now weighted by the sum of the edges of the former nodes in the community
3. Repeat phase 1

This leads to a decrease in communities in each iteration until modularity has reached its maximum and there are no more changes in the network (Blondel et al., 2008, pp. 3–4).

Apart from the communities themselves, "the resulting meta-network, whose nodes are the communities, may then be used to visualize the original network structure." (ibid., p. 2) In the present study this is achieved by creating a layout that represents the resulting structure by placing higher values on the edges inside the communities, without losing their actual weight. This is then visualized in a graph using ggraph (Pedersen, 2022).

3.3 Cluster Analysis

While a network in itself provides a wealth of information, to correctly interpret what is seen in the presented structure, an in-depth analysis is needed. The main hypothesis of this study is that the network will reveal meaningful structures within the dataset. To uncover these structures, several methods are applied.

First, the newly revealed information of community-membership is extracted from the network data and added to the original dataset. After that, general descriptive statistics about genre-dispersion are calculated to put the result into perspective. Furthermore, the

analysis of absorption patterns as described in 3.1.3 is repeated in the context of the clustering. The next step is to recalculate text similarity for the reviews adding the document variable of community membership. This allows for the application of corpus related methods that will then be used as guidance for a close reading analysis. The methods applied are *keyness analysis* and *concordance analysis.*

Keywords are "words that play a role in identifying important elements of the text" (Bondi & Scott, 2010, p. 1) and while keyness can equally apply to other descriptives of text as discussed in 2.1.3, the appliance of TF-IDF as measure for clustering endorses the use of lemmas in this context. In keyness analysis, a score is calculated that represents the relative frequency of a word occurring in a target corpus as compared to a reference corpus. In this analysis, the method of calculation is a likelihood ratio test (Dunning, 1993). The target corpus is constituted by the respective cluster in discussion and the reference corpus includes all other clusters that are considered in the analysis. Although the keywords themselves are definitely evidence that point to the role a set of texts play in a greater context, they don't constitute a complete analysis (Bondi & Scott, 2010, p. 3). And while close reading is a good method to contextualize these cues, for a more systematic approach, for example to investigate the different ways a specific term is used, concordance analysis is a valuable approach. In this, words are shown in context, displaying a preset window of words appearing before and after the target word.

4 Results and Discussion

After computing the network, there are 14 clusters ranging from two to 24 reviews per cluster. While overall the network does not exhibit a clear structure, the biggest clusters hint at a relationship between genre and text similarity. In this regard, the most obvious example is cluster 01, containing 85 percent Romance reviews, but also cluster 03, where half of the reviews belong to the Mystery genre. Although there seem to be clusters dominated by the Fantasy genre, one has to keep in mind that it is the genre that is best represented in the network. Furthermore, a closer examination of the clusters containing Fantasy gave the impression that the Fantasy reviews did not quite fit into the clusters in which they were included without showing a coherent structure in themselves. One explanation for this might be that after removing all of the Fantasy books that shared genre descriptions with the rest of the genres under investigation, the remainder of these books

are so versatile that they neither share a common review style nor fit with the other genres. Another reason might be the unique themes and plot points not included in the data cleaning that cause overfitting amongst reviews of the same title but not with similar reviews. This is why the Fantasy reviews were taken out and the network analysis was rerun.

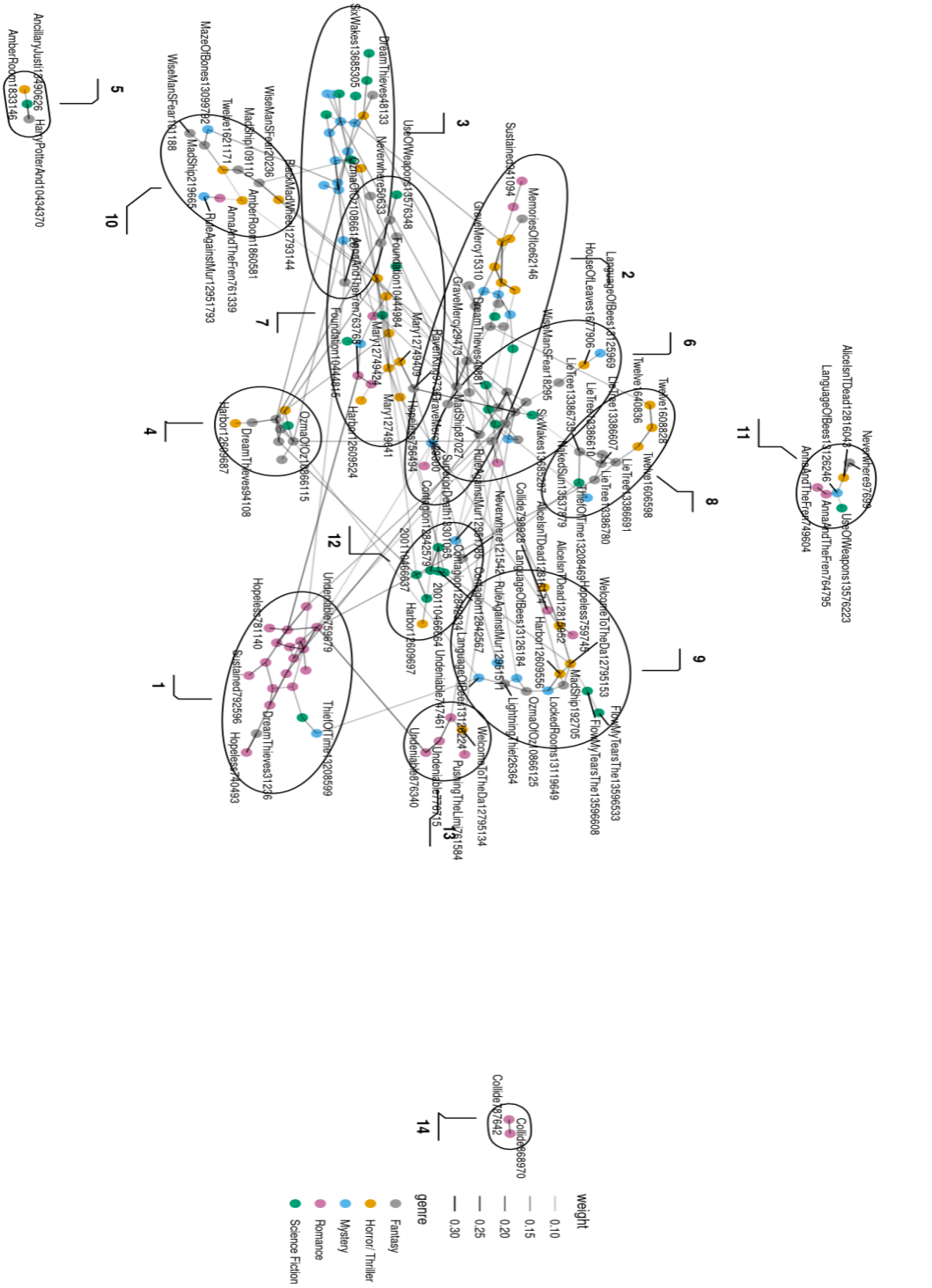

Figure 5: Network with Fantasy reviews, the circles highlighting the clusters were implemented with ggforce (Pedersen et al., 2020)

The elimination of Fantasy came with the loss of a lot of data, but this also makes the new network much more overseeable. Even though this network has one fewer cluster than the one before, the overall clusters, except for the Mystery cluster, are now much smaller. While the Mystery cluster stayed relatively stable, the Romance cluster that was observable in the last network was split up into several smaller clusters.

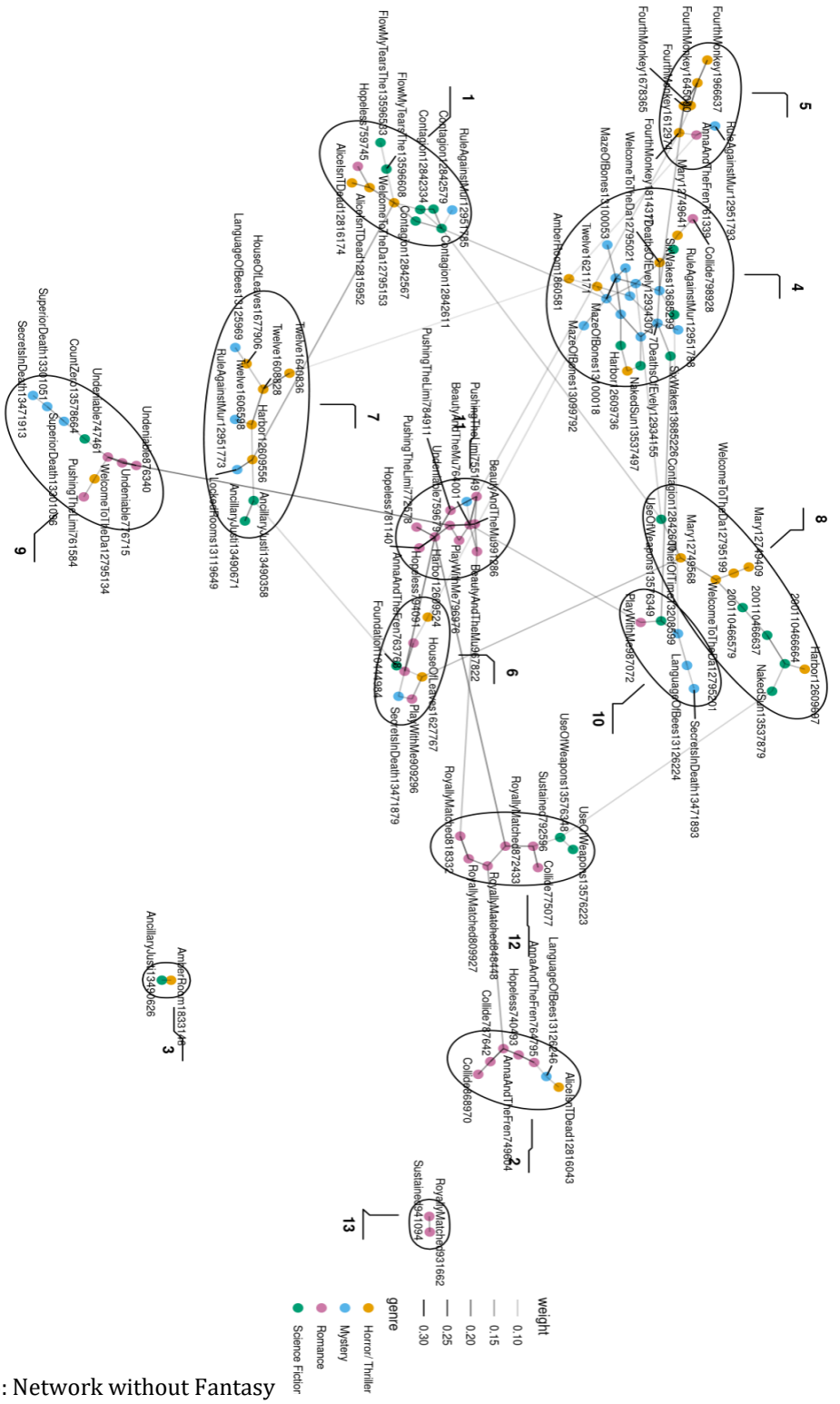

Figure 6: Network without Fantasy

For further analysis it was decided to further examine six of the clusters, namely the five biggest clusters (04, 11, 01, 08, 07), as well as cluster 12, which is the second biggest Romance cluster. The latter was inspired by the question as to why the Romance cluster that could have been observed before split up. Each of these six examined clusters show a dominance of at least 50 percent in one genre (except for cluster 08 which is split half-ways between Horror/Thriller and Science Fiction).

Genre distribution in Clusters

Cluster	Reviews (n)	Horror/ Thriller	Mystery	Romance	Science Fiction
4	23	0.217	0.522	0.043	0.217
11	11	0	0.091	0.909	0
1	11	0.273	0.091	0.091	0.545
8	10	0.5	0	0	0.5
7	10	0.5	0.3	0	0.2
9	9	0.111	0.333	0.444	0.111
12	8	0	0	0.75	0.25
6	7	0.286	0.143	0.429	0.143
5	7	0.714	0.143	0.143	0
2	7	0.143	0.143	0.714	0
10	5	0	0.6	0.2	0.2
13	2	0	0	1	0
3	2	0.5	0	0	0.5

Genre Representation in Clusters

Cluster	Reviews (n)	Horror/ Thriller	Mystery	Romance	Science Fiction
4	23	0.179	0.462	0.029	0.208
11	11	0	0.038	0.294	0
1	11	0.107	0.038	0.029	0.25
8	10	0.179	0	0	0.208
7	10	0.179	0.115	0	0.083
9	9	0.036	0.115	0.118	0.042
12	8	0	0	0.176	0.083
6	7	0.071	0.038	0.088	0.042
5	7	0.179	0.038	0.029	0
2	7	0.036	0.038	0.147	0
10	5	0	0.115	0.029	0.042
13	2	0	0	0.059	0
3	2	0.036	0	0	0.042

Table 7: Percentage of genres per cluster (left), percentage of genres in cluster as compared to the entire network

If one isolates the clusters that were chosen for this analysis, the network reveals a very interesting structure. It seems like the clusters predominantly consisting of Horror/Thriller, Science Fiction and Mystery (01, 04, 07 and 08) are more interconnected than the Romance clusters (12 and 11). Furthermore, the size of the nodes was modified to show degree centrality, which indicates how many nodes they are directly connected to. This gives an impression of their influence inside the clusters. We can see, for example, that the Mystery-nodes in cluster 04 seem to have more direct connections than the nodes

of other genres in the cluster, as is also observable for clusters 07 and 11 and their corresponding dominant genre.

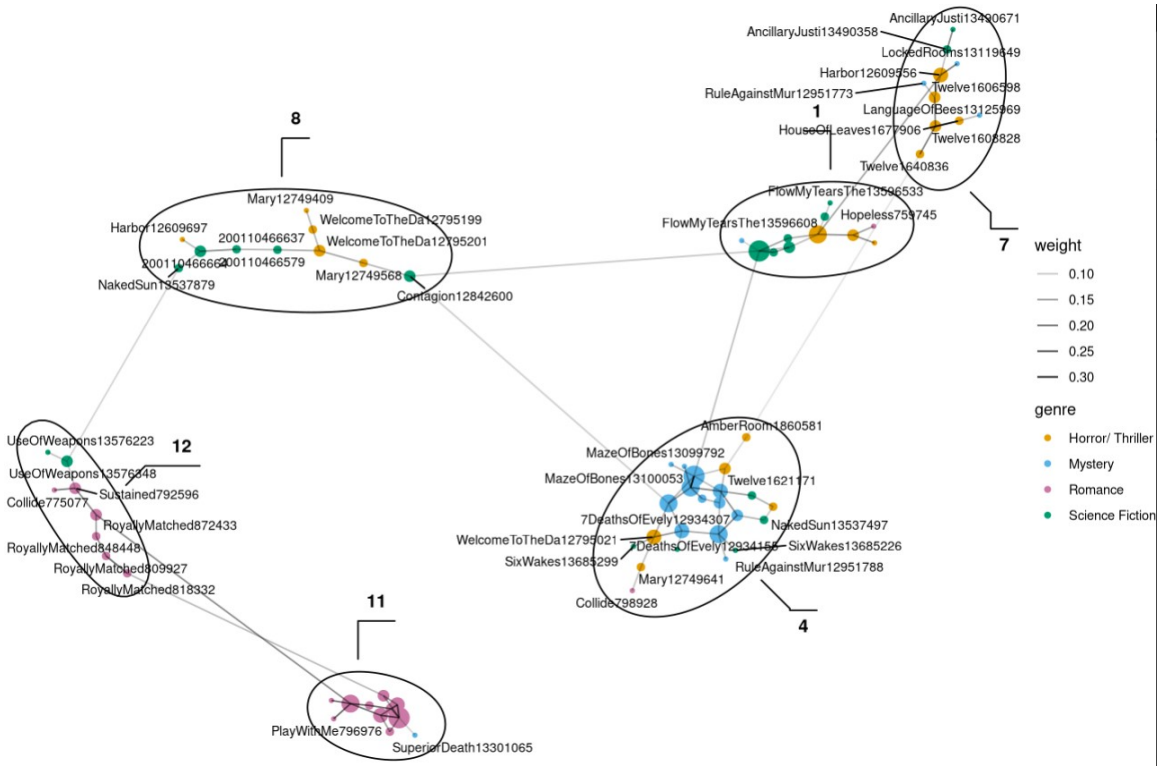

Figure 7: Isolated Network showing only the clusters under review

In the following, the clusters are discussed according to their absorption distribution, their position in the network and their content. As we can see in Figure 8, the distribution of absorption dimensions differs heavily between the clusters. As a guideline for the discussion of the content, keyness is used. To give structure and context to the keywords, they get pre-sorted into five categories:

- *Genre* includes any descriptors for the genre of the book
- *Experience* mostly encompasses verbs and adjectives but also nouns in metaphoric use ("it ripped out my heart") reviewers use to describe their reading experience (reactions, emotions)
- *General descriptors* are words readers use to describe the book and its content
- *Content* includes words that are uniquely identifiable as parts of the story of the books included (e.g. "vampires", "castle")

- *Additional information* contains terms that are not directly related to the reading process of the book, but give context to how readers found it or additional resources related to it (e.g. “podcast”, “movie”)

Of course, sometimes there are words that don’t fit into any of the categories; these will be collected in the additional category “other”, and words that fit into several of them. These terms will be highlighted in italics. To further investigate terms that are used in different contexts concordance analysis is used.

Lastly it should be noted that during the closer examination of the texts, it became apparent that some of the reviews were ones in which readers had received the book free of charge from the publisher or the author, but apart from mentioning that they did not stand out from those where readers had purchased the book and wrote the review on their own initiative.

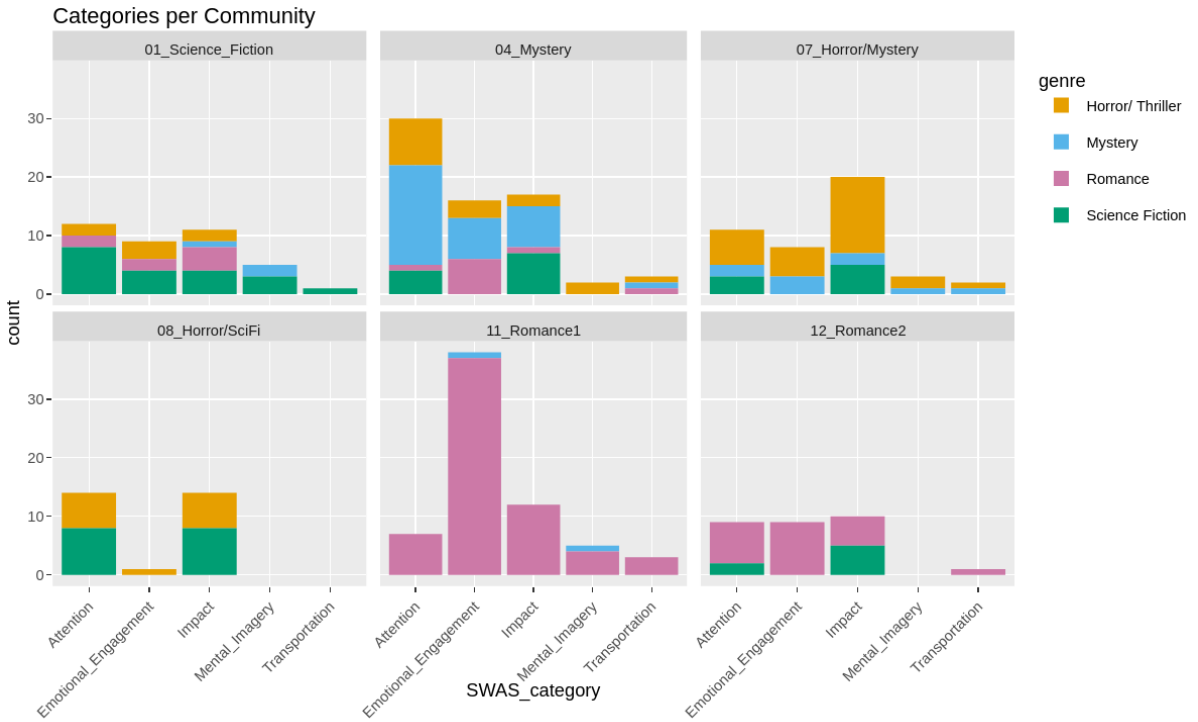

Figure 8: Distribution of annotations for absorption dimensions in the clusters

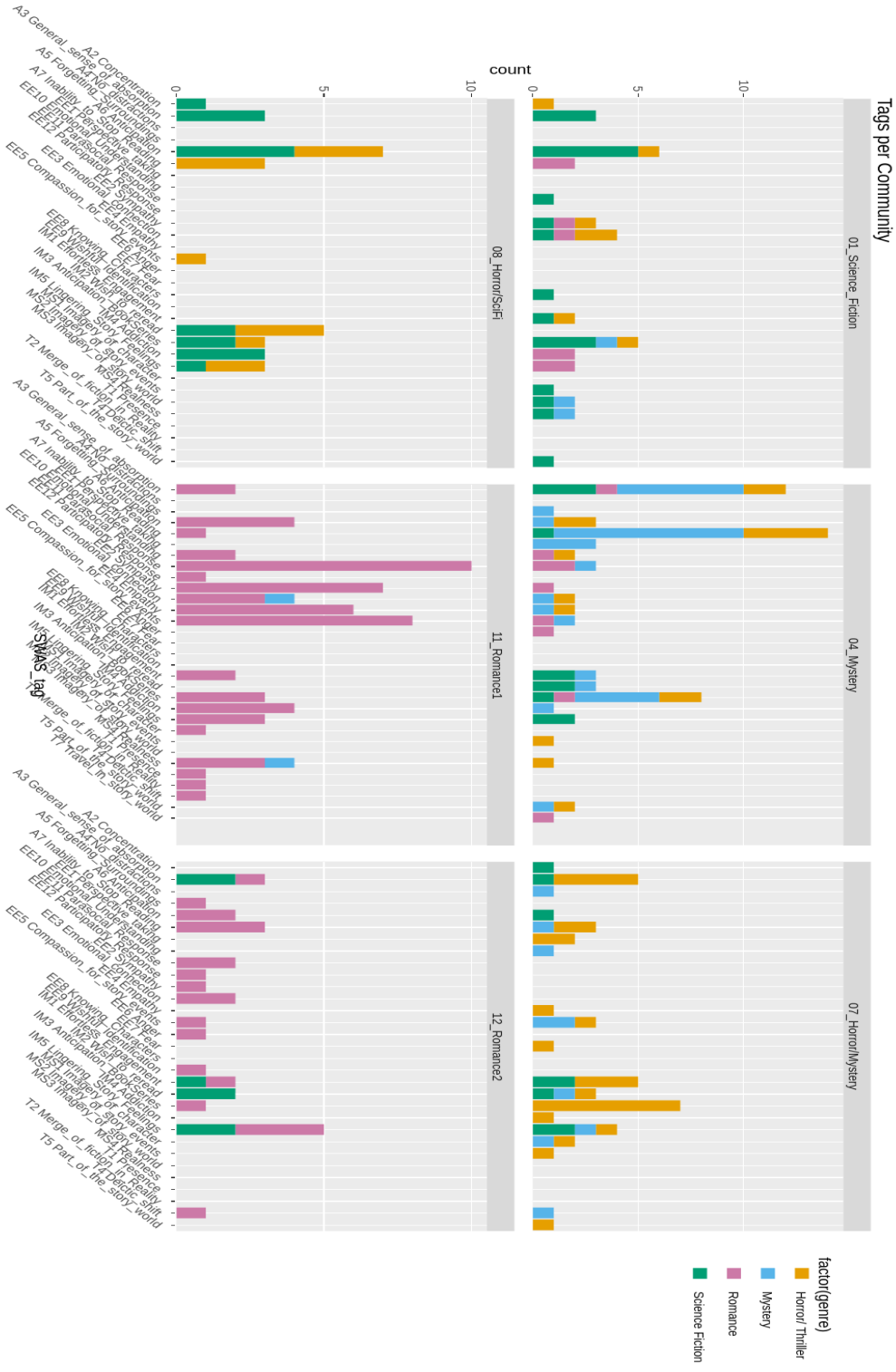

Figure 9: Distribution of fine-grained annotation per cluster

4.1 Cluster 01

Cluster 01, containing 11 reviews overall, is dubbed the Science Fiction cluster. This is because more than half of the reviews it contains are Science Fiction reviews (6), which make up 25 percent of all Science Fiction reviews in the network. This is followed by Horror/ Thriller (3) and one review of Mystery and Romance respectively. In the network, we see ties to the Mystery (04), Horror/Thriller-Science Fiction (08) and Horror/Thriller Clusters (07).

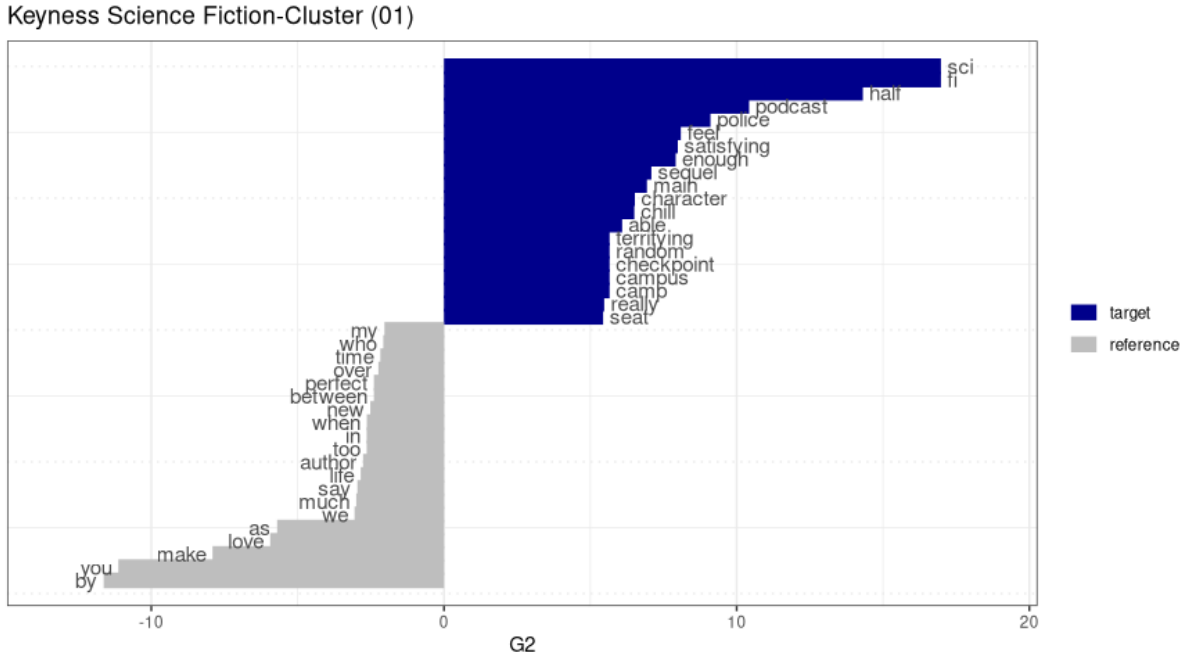

Figure 10: Keyness-plot Science Fiction cluster

If we look at the words that characterize this cluster, "sci fi" is at the very top and the only genre description in this collection. Noticeably, the reviews of other genres in this cluster do not use this description, while there are Science Fiction reviews describing books as Horror, which leads to the assumption that there are overlaps in the Science Fiction sample.

Genre:	sci, fi
Experience:	feel, satisfying, enough, chill, able, terrifying, seat
General descriptors:	half, random, really, main, character
Content:	police, checkpoint, campus, camp
Additional information:	podcast, sequel

Table 8: Categorized keywords for the Science Fiction cluster

While there are some unique content words inside this ("police", "checkpoint", "campus" and "camp"), we also see words that indicate reading experience ("feel", "satisfying", "enough", "chill", "able", "terrifying" and "seat"). Also, there are some general descriptors

("half", "random", "really", "main", and "character"); as well as additional information: ("podcast", "sequel").

Absorption-wise we see peaks in Anticipation, Anticipation Book Series, Sympathy and Emotional Connection. The most frequently used tag, Anticipation, is reflected in "seat", which is exclusively used in the expression "I was on the edge of my seat". The Emotional Engagement dimension is particularly surprising here, since it is quite underrepresented in Science Fiction reviews in general[7] and all Science Fiction reviews that contain Emotional Engagement seem to be accumulated in this cluster.

This could be explained by the notion that the terms "character" and especially "main character" seem to play an important role in this cluster, appearing 23 times, eleven of which in absorption-statements. While 10 out of 11 reviews mention them, only two of them state their actual names. This could be the reason why this term doesn't appear in the keywords of other clusters that show high Emotional Engagement such as cluster 11, as we will see further on. The reviews that mention the names of characters are the only Romance review, which presents the two main characters by name (Hopeless759745), and a Science Fiction review, which only argumentatively uses two of the side characters to criticize the author's work: "Ruth Rae is in the book and then she's not. Heather Hart is there and gone and there again. Every character has their own 20 page appearance before they are cast aside into the void again." (FlowMyTearsThe13596533)[8]. The main character on the other hand, who is referred to several times, remains nameless. It could be argued that for the reviewers in this cluster characters mostly serve as an essential part to drive the plot that one can sympathize with rather than objects of affection one can profoundly relate to: "they were engaging and dynamic enough to carry the plot and for me to care what happened to them" (Contagion12842567); "Hoover is able to allow the readers to understand the characters like the back of their own hand, and it's because of this full understanding that the reader is able to enjoy the book so much." (Hopeless759745)

---

[7] It should be noted that the comparison drawn here and the following is between a ratio-measure, where we see the degree of representation of Absorption dimensions in the different genres and the absolute number of absorption statements in the clusters.

[8] To protect the privacy of the reviewers, random checks were made to see whether the original review could be found on the internet via the citation text. This was not the case.

As we can see in the keywords "chill" and "terrifying" describing the experience readers are looking for, they don't attribute this to specific plot points or entities, but more to the atmosphere of the story-world and to narrative choices like point of view and pacing. Sometimes this is mentioned in a positive manner: "I personally could really feel the paranoia that this world he crafted was submerged in" (FlowMyTearsThe13596608); "The snippets from the pov of the 'boy' really added to the tension and the sense of danger that ran through the whole book." (Contagion12842567); "it was paced really well and kept the tension throughout the whole novel." (WelcomeToTheDa12795153) Other times it appears as criticism: "The bad parts were that he never uses much of this world!" (FlowMyTearsThe13596533); "it's just not my favorite POV to read because I feel like it takes me longer to connect to the books" (WelcomeToTheDa12795153); "I can see where some people would have thought the first half of the book being slow" (AliceIsnTDead12815952).

The verb "feel" occurs 17 times in the cluster (see Table 9), eight times in seven of the 38 absorption-statements. Even though one would instinctively associate that with the expression of emotion (especially given that this cluster contains a lot of references to the Horror genre) having a closer look at the usage of this term in context paints a different picture. About half of the times "feel" comes up, it is used to describe the atmosphere or in relation to the characters. In the other half, however, "feel" is often used as a hedging-device[9] to positively or negatively critique writing and plot while not claiming their stance to be read as objective truth, but as their subjective experience: "This whole book felt like it could be a short story that was bloated" (FlowMyTearsThe13596533).

Lastly, it is noticeable that according to the keyness-analysis the word "author" is used significantly less frequently in this cluster (zero times) than in the overall corpus. Nevertheless, approximately half of the reviews do mention the author, albeit by name, which due to the data-cleaning is not reflected in that observation.

[9] Hedges are linguistic devices that lower the epistemic status of a statement. They indicate that the statement is only a conjecture or a personal opinion (Hyland, 1998, p. 351).

docname	pre	key word	post
Atmosphere			
Welcome-ToTheDa12795153	From the very beginning I could	feel	the tension
Conta-gion12842611	The story was dark and intri-guing and I	felt	like I was there with the crew on Achlys
FlowMyTearsThe13596608	I really	felt	that this book captured the paranoid and mystical feel of
FlowMyTearsThe13596608	the paranoid and mystical	feel	of P.K.D . While I still have yet to read
FlowMyTearsThe13596608	I personally could really	feel	the paranoia that this world he crafted was submerged in
Writing			
Conta-gion12842579	I would call this book cine-matic . Reading it	felt	like watching a movie in the best way possible .
Welcome-ToTheDa12795153	liked the characters , for the most part , and	felt	that there was enough given about each character to feel
Al-iceIsnTDead12815952	I	felt	like the pacing of the story was good .
Al-iceIsnTDead12815952	the book being slow but I did-n't	feel	that way .
FlowMyTearsThe13596533	This whole book	felt	like it could be a short story that was bloated
FlowMyTearsThe13596533	they just	feel	like padding .
Welcome-ToTheDa12795153	it's just not my favorite POV to read because I	feel	like it takes me longer to connect to the books
Characters			
Welcome-ToTheDa12795153	there was enough given about each character to	feel	something for them.

Conta-gion12842611	Sometimes , while reading sci-fi , it	feels	difficult to connect with the characters and truly care
Conta-gion12842579	The characters all	felt	real , their motivations clear but sometimes complicated .
Plot			
FlowMyTearsThe13596533	WHY the main character doesn't exist . . . it	feels	. . . bizarre ? absurd ? not really satisfying
Welcome-ToTheDa12795153	it was setting up for the sequel , but I	felt	like I didn't really get any closure with this book

Table 9: Concordance of "feel" in cluster 01, blue highlighting shows statements that are tagged as Absorption

4.2 Cluster 08

Cluster 08 is precisely half Science Fiction (5) and half Horror/Thriller (5), containing 20 % of all Science fiction and 17 % of all Horror/Thriller reviews in the network. Looking at the network graph, the cluster connects to the Mystery (04) and the Science Fiction cluster (01) as well as the second Romance cluster (12). Another striking aspect of this cluster is its internal structure where we see a rather strict division between Horror/Thriller and Science Fiction reviews. However, two reviews constitute exceptions in that they are located on the opposite side of where they would have belonged in terms of genre. Upon closer investigation of these outliers, the Horror/Thriller-review in the Science Fiction part of the cluster itself is described as "suspense with just a touch of horror/science fiction" (Harbor12609697). Meanwhile, the review on the Science Fiction novel on the Horror/Thriller part of the cluster is not formulated as one would expect given the respective genre, but rather seems to describe a book one would attribute to the Horror genre: "It left me with nightmares that only made me want to read more. This book is thrilling, creepy, and addicting" (Contagion12842600).

The dispersion of absorption annotations aligns with the observations gained by examining the genres in the preliminary analysis which are highest in attention and impact, specifically Anticipation and Wish to Reread. Contrary to cluster 01, there is only a very

small amount of Emotional Engagement. This is also reflected in the keyness graph, where character is deemed as one of the words that are unlikely to appear in this cluster.

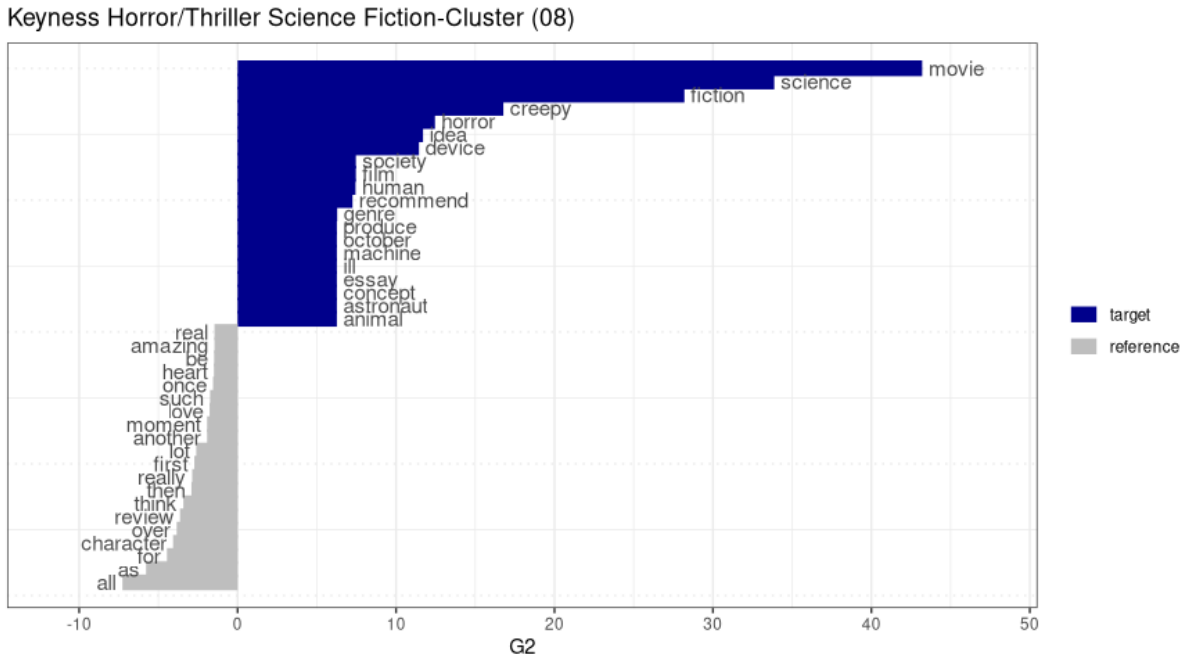

Table 10: Keyness plot Horror/Thriller Science Fiction cluster

Genre:	science, fiction, horror
Experience:	creepy
General descriptors:	idea, concept, essay, society, genre, ill
Content:	device, astronaut, machine, human, animal
Additional information:	movie, film, october
Other:	produce, recommend

Table 11: Categorized keywords for the Horror/Thriller Science Fiction cluster

The keywords reveal the genres included in the cluster: “science fiction” and “horror”. What can be observed here is that readers in this cluster prefer the word “science fiction” to “sci fi” as it is used in cluster 01, while the word “genre” itself is among the keywords that serve as general descriptors along with “ill”, “idea”, “concept”, “essay” and “society”. Furthermore, words that belong to the category additional information (“movie” and “film”) as well as the experience “creepy” are amongst them.

What is noticeable in this regard is that the keywords referring to content are all part of Science Fiction reviews, which can partly be explained by the longest review thoroughly describing the plot of *2001. A space Odyssey* (200110466579), elaborating on “animal”, “astronaut” and “device”. However, “human” seems to be a frequently used word in Science Fiction since the human characters are distinguished from androids or other species. Since this novel is intricately associated with the film version, it is no surprise that the

words “movie” and “film” play such an important role and that this cluster encompasses all reviews for it. But they are also reflected in Horror/Thriller reviews that don’t mention an existing film version by declaring their love for the cinematographic format of this genre or wishing for a film version of the book they read: “This is the perfect set up for one of those cheesy horror movies that you hate to love (or love to love in my case)” (WelcomeToTheDa12795201).

The keyword "genre" is not specific to Science-Fiction- or Horror/Thriller-reviews but found in both categories: “They created a new genre inside the science-fiction kind of movies” (200110466579); “It is near to impossible to produce a timeless piece of science fiction. Concepts and ideas date badly in the genre” (200110466664); “it doesn’t fall into the ‘torture porn' category of the horror genre” (Harbor12609697).

The examples for the discussion of the Science Fiction titles lead to the assumption that the books discussed here belong to the “classics” of the genre, which could be an explanation for reviewers preferring to spell out their genre classification, taking a more serious approach than reviewers in cluster 01, who preferred the shorthand “sci fi”.

Furthermore, most of the Science Fiction reviews in this cluster focused more on the book content, especially reflecting on the ideas discussed: “I loved HAL, and being a psychologist I could comprehend exactly what was going on in his electronic mind. Remember: we cannot create things from nothing, so if we create a mind, it must be in some parte [sic] human, with human responses.” (200110466579); “While reading I didn’t see the book as a story, but an essay on two ideas of robotics wrapped in literary style” (NakedSun13537879). They frame their reading as more of an intellectual exercise than reading for entertainment, which is also reflected in the descriptors “idea”, “concept”, “essay” and “society”, distinguishing them from the Science Fiction reviews in cluster 01 that are more drawn to the dystopian feel, focusing on the story. However, the Horror/Thriller reviews in this cluster are significantly less formal, as is reflected in the common denominator “creepy”. Unlike the Horror/Thriller reviews in the first cluster, there is a greater focus on the actual experience than on the textual devices that caused it, sometimes referring to circumstances that added to their fear or how it affected them afterwards: “It is haunting to say the very least, te [sic] kind of book that I [...] definitely should not have been reading right before bed!” (Harbor12609697); “I find myself thinking about this

story often, especially when I am alone at night walking past a mirror in my house, which is exactly what I love about reading creepy stories!" (Mary12749568)

4.3 Cluster 07

The Horror/Thriller Cluster (07) contains mostly Horror/Thriller reviews (5), which constitute half of the reviews in this cluster and 18 percent of all Horror reviews in the network, three Mystery reviews (12 % of all Mystery reviews) and two Science Fiction reviews (8 % of all Science Fiction reviews). It is connected to cluster 01 (Science Fiction) and 04 (Mystery). Notably, the direct connections are between nodes of the Horror/Thriller genre on both sides.

The absorption dimensions found most in this cluster are Attention and Impact, followed by General Sense of Absorption, Effortless Engagement, Anticipation Book Series and Lingering Story Feelings.

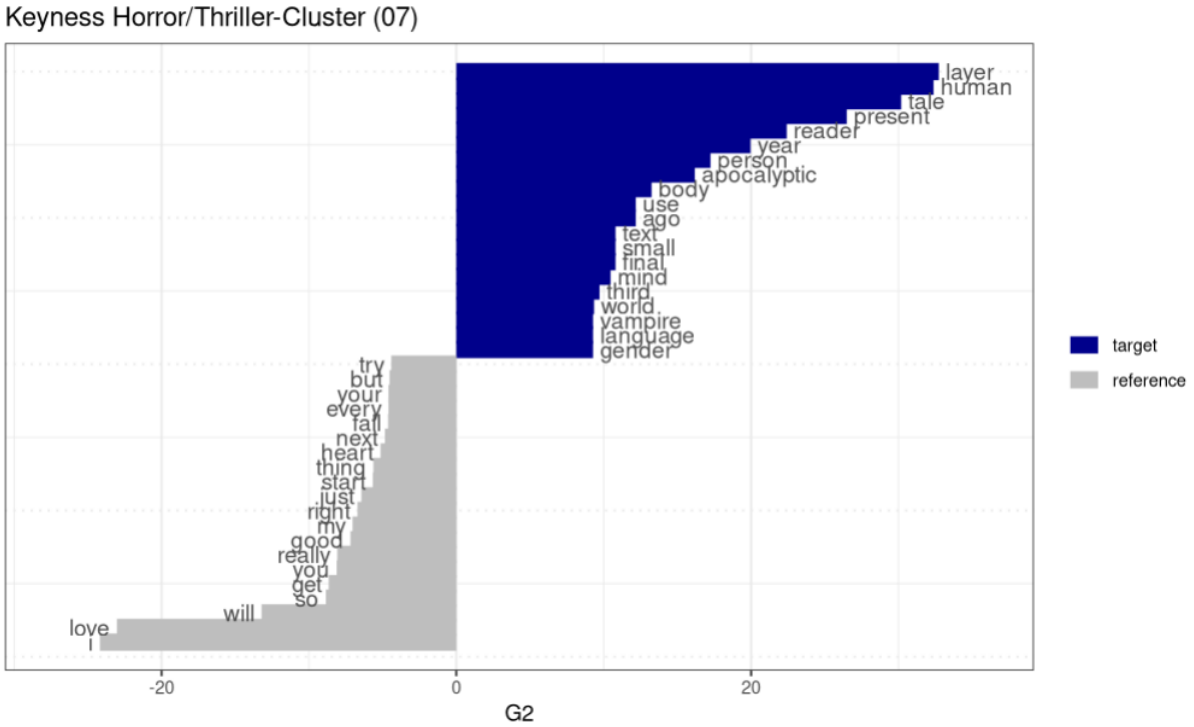

Figure 11: Keyness plot Horror/Thriller cluster

Most noticeably, there is no genre description in the keywords for this cluster. Secondly, the pronoun "I" is considered to be most unlikely to appear in this cluster. This is contrasted by the prominent role of the keyword "reader". While there are few cases where "reader" refers to actual readers, namely the reviewers themselves, the fan community

they feel being a part of, or potential readers that use their review for guidance: “I am not a fast reader” (Harbor12609556); “I join a rather large community of Cronin's readers who eagerly await book three” (Twelve1640836); “I would be correct to advise the shallow reader of either type to look elsewhere” (Twelve1640836). The most common usage of the term, however, is to elevate the report of their experiences to the more abstract subject of the implied reader: “Louise Penny has a way of opening a lid on even the tightest container of deep-felt emotion, stirring it about, and helping the reader discover something they, perhaps, didn't even know they'd been looking for.” (RuleAgainstMur12951773)

Genre:	none
Experience:	reader, mind, world
General descriptors:	layer, tale, third, person, text
Content:	present, human, apocalyptic, *person*, vampire, human, animal, year, use, gender, language, small, *mind*, ago, body, *world*
Additional information:	*year, ago,* final

Table 12: Categorized keywords for Horror/Thriller cluster

This authoritative stand reflected in the more generalized description of the reactions of a “reader” in contrast to the narrower scope of “I” also manifests itself in comparison to cluster 01, where the hedging term “feel” was often used to describe readers’ thoughts about the book, whereas readers of this cluster use a more assertive tone: “It's not that the reader is willing to suspend disbelief as much as it doesn't occur to the reader that what she's reading isn't believable. Cronin creates a viral embodiment of vampirism that connects so well to the vampire mythos that the reader finds his stories perfectly plausible” (Twelve1640836).

Overall, the general descriptors give the impression that the vocabulary used is mimicking professional language: “layer”, “third”, “person” and “text”, which points to the reviews in this cluster placing higher value on stylistic features and the poetic quality of the text than on an easy and entertaining read. This value system is explicitly mentioned by one reviewer: “Yes I’m shallow because normally I need a main character who is somewhat personable to interest me in spending time with a story” (AncillaryJusti13490358). Accordingly, the word “love”, which in other reviews often describes the readers’ relationship to the characters, is largely underrepresented in this cluster. Moreover, despite readers unanimously praising the books they are reviewing, the term is rarely used even

in this regard, indicating that reviewers in this cluster make an effort to state their admiration more eloquently: “House of Leaves is a unique achievement in literature.” (HouseOfLeaves1677906) – and criticizing themselves if they can’t maintain that standard: “I hate using this cliche, but I can't put it down.” (Harbor12609556)

The book under review is certainly a big factor in producing this style of writing, as a discussion of “Broken passages, inverted text, arcane footnotes, and four layers of narration” (HouseOfLeaves1677906) is only possible if it is offered by the text. Yet a comparison of reviews on the same book reveals that this cluster might be equally shaped by the style of the individual reviewer. In this cluster, *Locked Rooms* is reviewed as: “I don't care for narratives which shift from first to third person. It's disconcerting and distracts away from the story. I understand it's needed in this instance because Russell was not functioning properly throughout and Holmes became a bigger focal point, uncovering evidence the reader needs to see.“ (LockedRooms13119649) On the contrary, the same book clustered in cluster 04 with this reflection: “Also, I loved that we got to hear parts of the story from Holmes's point of view. Usually we only get to hear what he's thinking through Mary's narration. But here we actually got to see through Holmes's eyes, which was wonderful!” (LockedRooms13119495).

4.4 Cluster 04

The Mystery Cluster (04) is the biggest cluster containing 23 reviews, roughly half of them belonging to the Mystery genre (12), which equals 46 percent of the total number of Mystery reviews in the network. Furthermore, it consists of 22 % Horror/ Thriller (5) and Science Fiction (5) reviews respectively, which represent 18 and 21 percent of all reviews of those genres in the network, as well as one Romance review. It has connections to the Science Fiction (01), Horror/Thriller (07) and the Horror/SciFi clusters (08). Within the cluster itself, reviews on Mystery novels build the core of this cluster. This is shown by them being more strongly interconnected than the ones at the outer rim of the cluster, which belong to other genres.

The most dominant absorption dimension in the cluster is Attention, followed by Impact and Emotional Engagement, which corresponds to the overall distribution of Absorption in this genre. In the fine-grained annotation we see peaks in General Sense of Absorption,

Inability to stop Reading and Anticipation Book Series. The latter is reflected in the keyword "series", while attention itself is a theme in the reviews as reflected in the terms "happen" and "pay": "I can't wait to see what happens next." (NakedSun13537497); "My biggest piece of advice to anyone reading is: PAY ATTENTION." (7DeathsOfEvely12934387) The former is frequently used in statements annotated as Anticipation or Anticipation Book Series, whereas the latter is not covered by the annotation framework.

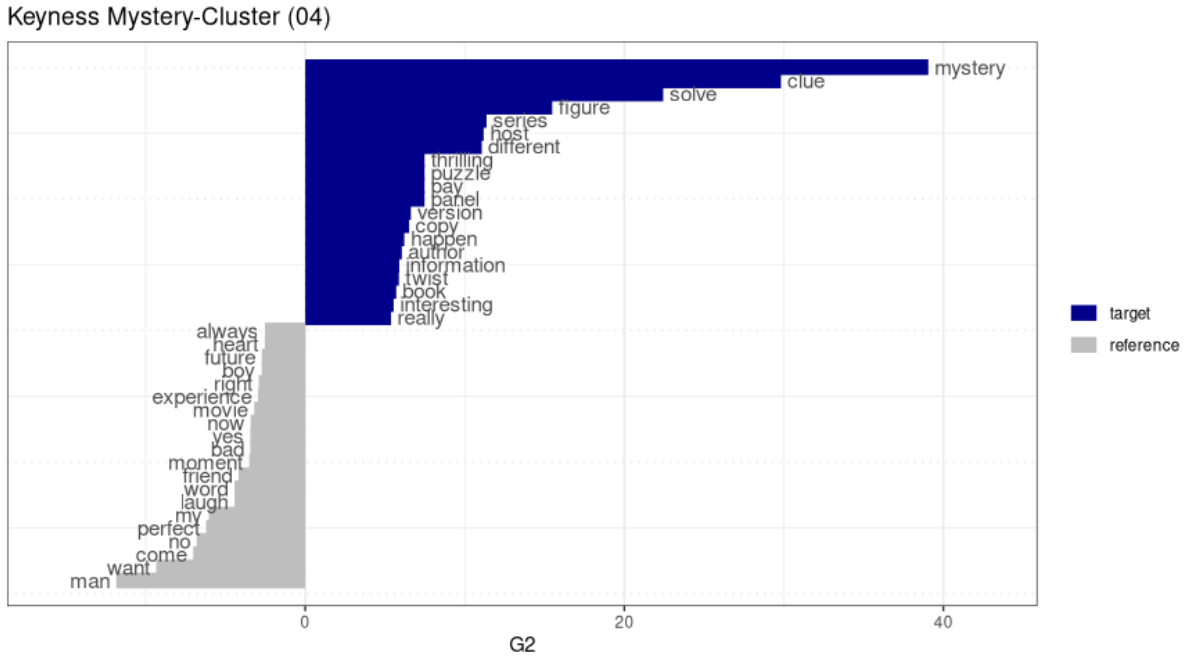

Figure 12: Keyness plot for Mystery cluster

Genre:	mystery
Experience:	thrilling, interesting, solve(1), figure, pay, happen, author
General descriptors:	book, different, twist, clue, puzzle, information, really
Content:	panel, host, solve, *mystery, happen*
Additional information:	series, version, copy, *happen, author*

Table 13: Categorized keywords for Mystery cluster

Regarding the keywords in general, it is significantly more difficult to distribute them into the categories that have previously been used. There are only two words that uniquely describe the content of two of the books in the cluster, namely "panel" and "host". The reason for this might be that stories in this cluster, unlike those discussed before, do not take place in alternate worlds or the future. This makes the vocabulary used for plot description less distinct. Arguably, "mystery" and "solve" are also descriptions of the content: "It was interesting to see Holmes in an entirely different environment still being himself when it comes to solving mysteries." (LockedRooms13119513). The word "mystery" encompasses its role as a genre descriptor as well as the fundamental concept of

most novels within this genre. This polyvalence could potentially contribute to the cohesive clustering observed in this genre.

Moreover, readers in this cluster share a lot of the same experience, as seen in the absorption-annotation, and thus tend to use similar language to express themselves. For example, eight out of 14 absorption statements annotated as Inability to Stop Reading used some form of not being able to "put down" the book, while two reviews that used the very same phrasing in Anticipation Book Series: "I cannot wait for the next book." (LockedRooms13119513 & MazeOfBones13100053). Furthermore, the reading experience of Mystery reviewers in this cluster significantly differs from that of readers of other clusters. One can observe evidence for active cognitive involvement, as expressed by the keywords "clue", "figure [out]", "puzzle" or "information": "Mind you, each scene, character, and information is important to the story; without each and every one of those, there would be a missing piece to the puzzle of this book" (7DeathsOfEvely12934155). This expresses that readers do not only enjoy the unraveling of the mystery by the characters, but consider their own efforts or success of solving the puzzle as equally important: "their [sic] are clues on every page of the book that will help you know what will happen next just like the character in the book" (MazeOfBones13099887); "If I could find something to dislike in this book, I'd say I was confused at certain times by the clues given. I find myself trying to figure out who did it, and never fully put the whole thing together. I never would have figured out the mystery on its own" (7DeathsOfEvely12934387). This leads to the point where they engage with the book even beyond the act of reading itself: "you have to do your research or know a lot of info to figure them out. [...] The fact that you can register online and get your own clues in the books is an added bonus as well. It has been fun figuring out our own clues and feeling like more of a part of the story." (MazeOfBones13100056)

This emphasis on active involvement in the process of understanding becomes particularly apparent when considering the reviews of other genres in this cluster. The overarching tenor of these reviews seems to be that the reviewed texts are, in different ways, rather intricate and that there is a focus on navigating these complexities. But unlike the Mystery reviews, the active cognitive involvement is not understood as a matter of course, but as an obstacle that has to be overcome: "I was neutral on Emily for about half the book, trying to figure her out, and then bam! I got it. I really liked her and felt for her."

(Collide798928); “It took me three years and three attempts to get into this book, but am I glad i [sic] finally did. For me the key was to figuring out of what i [sic] thought the starting anchore [sic] point and the timeframes of various chapters...I know its [sic] hard” (UseOfWeapons13576219).

4.5 Cluster 11

Cluster 11 is the biggest Romance-dominated cluster, containing only one review of a different genre (Mystery) and almost 30 percent of all Romance reviews in the network (10). Within the clusters discussed here, it is only connected to the second Romance cluster (see below).

Regarding the absorption dimensions of this cluster, there is a clear dominance of Emotional Engagement followed by Impact and Attention. This has also been observed in the preliminary analysis, although the discrepancy between Emotional Engagement and Impact is significantly larger in this cluster than in the entirety of the genre. Additionally, there is a relatively high amount of Mental imagery. In the fine-grained annotation, we see that Emotional Engagement is mostly expressed through Parasocial Response, Sympathy, Emotional Connection, Empathy and Compassion for story events. Impact is mostly observed as Addiction, followed by Anticipation Book Series and Lingering Story Feelings. The most frequently used tag for Attention is Anticipation and Realness for Mental Imagery.

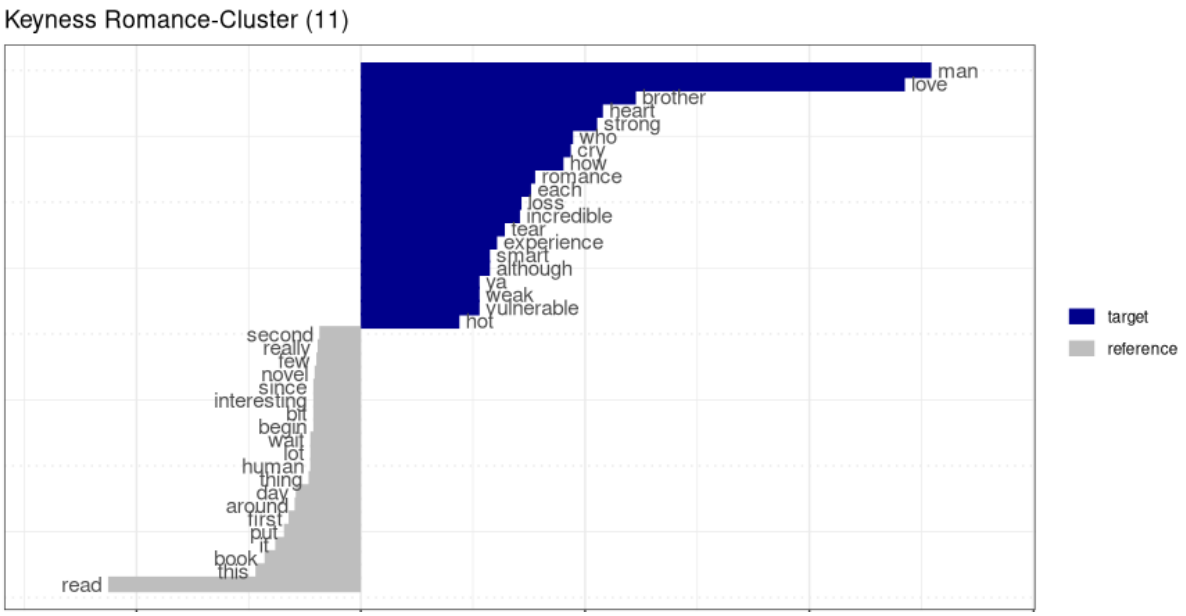

Figure 13: Keyness plot Romance cluster (11)

Genre:	romance, ya
Experience:	love, heart, cry, tear, experience
General descriptors:	incredible, smart, hot, strong, smart, vulnerable, weak
Content:	man, *love*, brother, *heart (1)*, loss, *experience*
Additional information:	*love*;
Other:	who, how, each, although

Table 14: Categorized keywords Romance cluster (11)

As genre descriptors we find "romance" and "ya" (Young Adult), which suggests overlaps of these genres in the cluster. Unlike the other clusters, the genre description "romance" is not as distinctive for this cluster as can be seen in the keyness graph. On the one hand, this seems surprising since just like "Mystery", "Romance" is not only a genre denominator but also the key concept underlying the storylines of this cluster. On the other hand, because the concept of romance is a common one in literature regardless of the genre and because this cluster doesn't encompass all of the Romance reviews in the network, the term might be quite common throughout the whole corpus. Furthermore, reviewers often describe the books as "love stories", which could be an expression that fits better into the informal style most of the reviews deploy. This can be seen in the frequent use of capital letters to emphasize words as well as descriptions of bodily reactions inside of asterisks: "I LOVE DEAN HOLDER. I WANT HIM, I NEED HIM...*exhales the breath I was holding the entire time and sighs*. And most of all, I JUST LOVE THIS BOOK...plain and simple." (Hopeless781140). Other evidence on the rather informal language are the comments on the reviewers' own words – "I wasn't especially fond of [...] her damn stubbornness [...] (right Becca, says the girl who is the epitome of stubborn)" (PlayWithMe796976) – direct addressment of the author often by their first name – "Oh, Penelope, where do I start?" (BeautyAndTheMu764001) – and the frequently used slang like "OMG", "LOL" or seemingly specific to the community: "insta-love".

The highest rated term for this cluster is "man"/"men". It appears 26 times in the cluster, with only ten of these occurrences describing the characters discussed (see Table 18 in appendix). The rest can be found in general discussions of men, comparing them to the men in their lives: "Men like my father and my uncles. Men I loved with all my heart. Men like Deuce." (Undeniable759679); "Once again, why haven't I met men like him? Oh right, they only exist in books as far as I am concerned." (PlayWithMe796976). This would suggest that Romance readers are mostly straight women. The female protagonists of the books are equally often discussed as their male counterparts. Nonetheless, they are more often seen as points of identification and adoration while men are rather discussed as

objects of desire: “I loved them both equally and individually (I loved Echo and was IN LOVE with Noah).” (PushingTheLimi772578). This is also reflected in the usage of the second strongest keyword “love”. It is more often directly used for the male main character rather than the female protagonist, while the focus lies less on the person themselves than their admirable qualities like strength. In total, the word “love” appears 81 times in this cluster, 24 times in relation to characters, twelve in relation to the novel or specific parts of the story, three regarding the author, 21 to describe the content or in citations, four to denote the genre and ten in relation to the reading experience (see Table 17 in appendix). Moreover, it is exclusively used in the Romance reviews of this cluster.

Regarding the keywords that represent the content of the stories reviewed, a similar observation can be made as the one regarding the discussion of the Mystery cluster. One could yet again argue that novels in this cluster do not touch on themes outside the readers’ reality and that therefore, only a few of the keywords can be tied to specific stories; however, it must be noted that only one of the reviews gives a proper plot description. The rest of the reviews only mention themes that are dealt with in the novel. For example, “brother” and “loss” point to the plots of Beauty and the Mustache and Pushing the Limits, but readers mostly seem to highlight that to share their own experiences: “As someone who just experienced a loss in my family, I can relate to all of this, which only made me cry more.” (BeautyAndTheMu764001); “the most prominent to his story is being separated from his two little brothers who are in foster care. Although nothing like this has happened to me personally I have a family member who has and so I know how messy, emotional, and complicated that whole situation is for everybody involved.” (PushingTheLimi755149)

In other instances, they share their reactions and further characterize the protagonists: “I felt so angry and frustrated at the cards he had been dealt over and over again. The unfairness and devastating losses he faced were so cruel I could feel the anger and inability to trust that he had come to accept as normal.” (PushingTheLimi772578).

Alternatively, they point to things they particularly enjoyed: “And the love he has for his brothers is the CUTEST damn thing I've read about in a while!” (PushingTheLimi784911) This is also showcased by the frequent use of quotes and the keyword “how”, which often combined with the phrase “I love”: “I love, oh how I love, how ‘no kissing’ can turn out to

be the most sensual, most intoxicating and the most sexiest scene you will ever read about." (Hopeless781140) Instead of modeling their texts on professional reviews, which give a plot descriptions and evaluations to recommend the book/novel, the reviewers, which are studied here, seem to assume that they address an audience who have also already read it, sharing their experience and favorite parts without any context: "I laughed, I cried (a lot) and I laughed until I cried many places but the raccoon was the best!" (BeautyAndTheMu967822) Rather than the plot, readers seem to put more emphasis on the characters: "strong", "smart" and "vulnerable" are frequently used to describe the characters and their bond, except for Beauty and the Mustache, where "smart" is often used because it apparently addresses quite complex themes: "I like that I always learn something educational in these books but I did not have to Google as much in this one." (BeautyAndTheMu967822) This is also expressed by the frequent usage of "love" with regard to characters, but also by the relative pronoun "who", whose prominence in the cluster points to the dominance of personal descriptions. The same goes for the determiner "each", whichmostly appears with "other", showing the emphasis on the relationships depicted. Therefore, while it does stand out in tone, it is no surprise that the only review not describing a book from the Romance genre completely confines itself to praising the female protagonist and the real-life connection between reviewer and book.

Apart from that, as can be derived from the rather strong keyword "incredible", for reviewers from this cluster reading is perceived as a very intense experience. This is further backed by the term "experience" itself, which functions as one of the keywords. "HOPELESS is a story you all need to live and experience yourself. You need to BREATHE it in and let it run through your veins until it's embedded in your soul." (Hopeless781140) This deeply embodied notion of reading is also reflected in the keywords "heart", "cry" and "tear" as well as in the usage of "love" to express the kind of reading experience they undergo: "Their story made me laugh, cry, squee and swoon, and I loved every single freaking minute of it!!!" (Hopeless794091). Notably, these experiences seem to lean more towards negative emotions, which, paradoxically, do promote a certain amount of pleasure: "This book will tear out your heart, shatter it into the tiniest little pieces and then slowly repair itself, one piece at a time. It is breathtakingly stunning and beautifully torturous." (Hopeless781140)

As already mentioned, some readers share some of the negative experiences described in the books, which seems to further their enjoyment into an impactful positive experience: “Although my experiences are vastly different to the characters in this book the feelings behind them weren’t all that different at all and so I connected with Noah and Echo in a rare and special way. [...] My emotions were totally invested in this book and so I really connected with it and can see it holding a special place in my heart for years to come.” (PushingTheLimi755149)

This pathos-laden component is accompanied by positive feelings towards the relationships, as indicated by the instances where readers talk about the love they share or the keyword “perfect”: “The plot was a perfect balance of an intense love story and a heartbreaking story of tragic loss & forgiveness.” (PushingTheLimi772578). This also applies to the sexual tension most readers positively report on as implied by the keyword “hot”, which is not at all perceived as incongruous to the more serious aspects of the story: “And WOW, Katie McGarry sure knows how to create some seriously HOT intimate moments. *Fans self* Their romance just felt right - two emotionally and physically scarred people who come together with a mutual understanding and the ability to heal each other's wounds.” (PushingTheLimi784911)

4.6 Cluster 12

Cluster 12 is the second Romance cluster containing eight reviews, six of which belong to the Romance genre (18 percent of all Romance reviews in the network) and two Science Fiction reviews (eight percent of all Science Fiction reviews). It is strongly tied to the other Romance cluster (11) and has a connection to the Horror/Thriller Science Fiction cluster (08). It can be noted that the direct connection is located between two science fiction nodes, which seem to be a more incidental match to the cluster as they are quite short.

Regarding the Absorption Dimensions, Attention, Emotional Engagement and Impact score the highest, the latter being slightly more represented than the two former. This is quite astonishing given the distribution of the genre in general. The Science Fiction reviews do not contain Emotional Engagement. But while they are mainly responsible for the rise of Impact above Emotional Engagement and contributing to the Attention count,

even without this influence the ratio between Emotional Engagement and Attention does not align with the general Distribution in the genre.

In the fine-grained annotation, there is a peak in Lingering Story Feelings in Impact, followed by General Sense of Absorption and Inability to stop reading in Attention and Parasocial Response and Emotional Connection in Emotional Engagement. Significantly, while Romance is the genre with the highest density of annotations, there is a vast difference between the annotation count of this cluster (29) and the other Romance cluster (11 ,65 annotations). After considering the difference in number of reviews and the inclusion of reviews from different genres, this still leaves us with 6.5 annotations per review in cluster 11 and only 4.8 in cluster 12. This is also reflected in the keywords where we find unique content words that point to plot descriptions, which can be found in approximately half of the reviews, but only one keyword describing reading experience ("love").

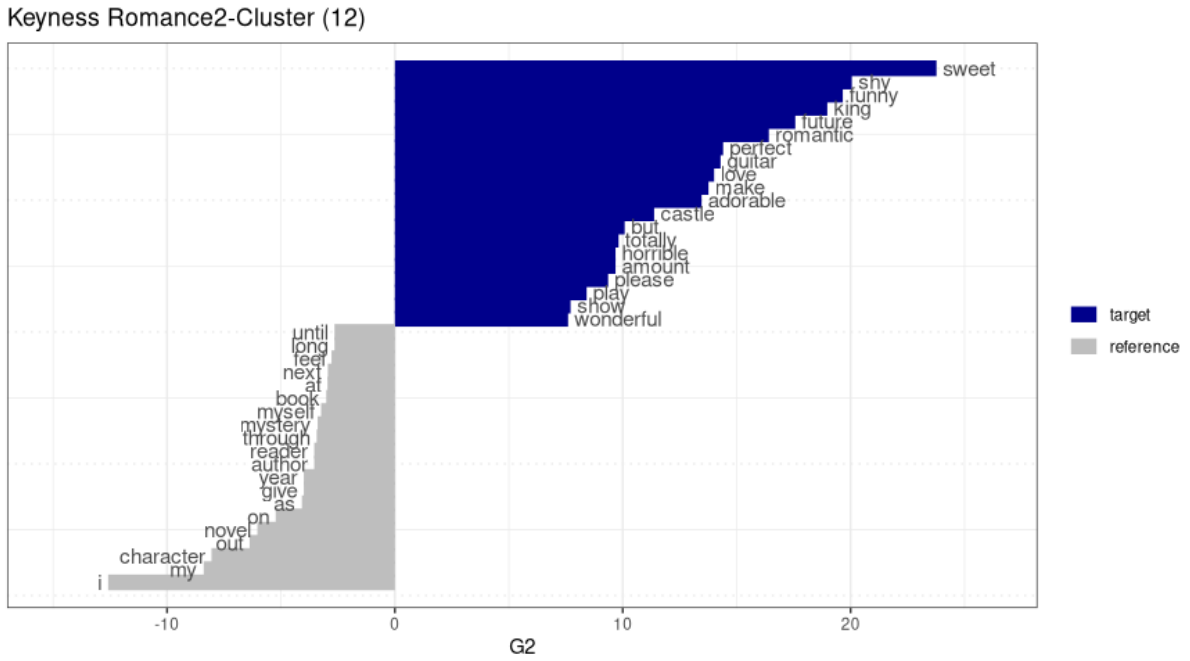

Figure 14: Keyness plot Romance cluster (12)

Genre:	none
Experience:	love
General descriptors:	funny, romantic, perfect, adorable, amount, wonderful, sweet, shy, horrible
Content:	king, future, guitar, *love,* castle, play, show
Additional information:	none
Other:	make, but, totally, please

Table 15: Categorized keywords for Romance cluster (12)

The reviews are very similar in style to those in cluster 11, employing the same informal language, capitalization, reaction descriptions in asterisks, commentary, slang and directly addressing the author, often by first name, while frequently adding “please” as found in the keywords.

What separates this cluster from other clusters that show a clear dominance in one genre is that no genre description appears in the keywords. Although “love” is among them, it is not mentioned in terms of genre (“love story”) and is not as defining for this cluster as it is in Cluster 11. In this cluster the term appears 38 times, seven times in regard to the characters, 13 times to express adoration for the book or parts of the story, once in regard to the author, 13 times to describe the content and once in relation to reading experience. By analogy to cluster 11, it is exclusively used by readers of the Romance genre and used more often in relation to the male than the female protagonists. In contrast to the other cluster, “love” serves less in the context of character descriptions and only in one case as a depiction of experience.

The most frequent function for “love" in this cluster is to express readers’ affective attitude towards the novel and to describe its content (see Table 19 in appendix). This reflects the fact that the connection to the characters tends to remain on a superficial surface level even though they play an important role in these reviews. This may be due to the fact that the romance novels in this cluster seem to correspond more to the genre of romantic comedies, as is also indicated in the keywords “funny”, “romantic”, "perfect", "adorable", "wonderful" and "sweet".[10] Another reason might be that readers in this cluster do not actually share the experiences that are depicted, but rather perceive them as a necessity to an engaging plot: “I like the awareness the writer sends here regarding domestic violence, an often element inserted in book with a certain purpose” (Collide775077); “I loved the humor and the fun but there was an undercurrent of emotion happening too and it was the prefect mix” (RoyallyMatched872433); “But those mistakes

[10] Except for “horrible”, which is in few instances used to describe the feelings of one character in regard to the prospect of becoming king and leaving behind his carefree life but is only addressed briefly: “But now that he's the future king, he knows exactly how horrible that life can be. He has so much political stuff to learn - and he hates that.” (RoyallyMatched818332)

made for great drama that really helped create this full, heartfelt story." (Sustained792596)

This is also underlined by the fact that "I" is the term deemed most unlikely to appear in this cluster, which – given a representation of 10.5 times per review in this cluster versus 17.8 times in cluster 11 – could be a marker of less self implication in writing the reviews.

4.7 Summary

Upon closer investigation of the Reviews, several claims can be made. The Romance and Mystery reviews in this corpus seem to be linguistically more distinguishable than reviews of the other clusters. This can not only be seen in the Network, but also in the closer investigation of clusters. Mystery readers exhibit a marked preference for a vocabulary that reflects their active cognitive involvement. Rather than focusing on what the story does to them, Mystery readers predominantly employ language that emphasizes their engagement in active pursuits such as seeking out clues, piecing together information, and engaging in reasoning to unravel the mystery. This linguistic contrast hints at a dichotomy between passive emotional immersion and the active participatory nature of the reading experience in the Mystery genre. Romance readers, however, emphasize the heavy emotional response the text, especially the characters, elicits in them. Moreover, Romance reviews share a distinct informal and personal style that indicates an enclosed community with its own conventions apart from that of other genres.

The clusters encompassing Science Fiction and Horror/Thriller on the other hand do not show such uniform patterns. One reason for this could be the overlaps in terms of genre, leading to a more diverse set of expectations and reactions. Another one could be that these genres attract a wider range of readers than for example the Romance readership, which could be identified as mostly straight and female. This again could lead to a wider array of modes of discussing literature, which is indicated by the different foci readers set in their reviews. The readers in cluster 01, for example, focus on "atmosphere" and the features by which it is achieved, cluster 08 on the topics addressed in the (Science Fiction) books, and cluster 07 on the literariness of the works discussed.

All clusters have rather distinct absorption patterns and words frequently appearing in the reviews overlap with those often used in absorption statements. While most of them

align with what would be expected by the genre that is best represented in the respective cluster, there are few exceptions. One is cluster 01, with an emphasis on characters and encompassing all Science Fiction reviews that contain Emotional Engagement. Another exception is cluster 12, which scores noticeably low on emotional engagement, although it mostly contains Romance reviews. This points to a subclass of Romance reviews which, even though focusing on characters, does not show the same emphasis on emotional connection with them as opposed to readers in cluster 11.

5 Conclusion

The presented research establishes network analysis as a useful tool for the identification of meaningful clusters in a collection of reader reviews. While genre was not the main factor in discriminating reviews, the network showed more connections within Mystery, Horror/Thriller and Science Fiction than these genres had to reviews of the Romance genre. This backs previous findings on Romance novels attracting a more homogeneous community of readers (Antoniak et al., 2021) and these tending towards specific vocabulary, such as the choice of body metaphors (Doche & Ross, 2022; Nuttall & Harrison, 2020). Apart from that, Mystery reviews exhibited a closer-knit linguistic similarity that revealed a unique reading approach in comparison with the other genres. They positively emphasized the challenging nature of the novels they reviewed, characterizing the nature of the reading experience they are seeking out as heavily reliant on active cognitive involvement.

Furthermore, it could be observed that each cluster demonstrated distinct absorption patterns, where reviews that deviated from the standard pattern of their genre tended to be sorted into separate clusters. This became apparent in the subdivision of Romance reviews in cluster 11 and cluster 12 as well as the character focused Science Fiction reviews in cluster 01. Moreover, terms that reflected prominent absorption dimensions for a cluster often appeared in the keywords. This suggests that the language used in absorption statements had high discriminatory power in the clustering process. However, this influence should be further explored within corpora that also contain reviews that do not report on absorbing reading experiences as well as within approaches that explicitly incorporate absorption in the clustering procedure.

While TF-IDF as measure of text similarity did produce meaningful results in this study, it proved to be susceptible to terms that were not as informative ("film", "movie"). Furthermore, it required an extensive process of data cleaning, which limits its applicability on larger datasets. In this regard, approaches that consider only the most frequent words should be explored. Another solution to this would be to combine reviews on the same book, which should decrease the discriminatory power of story-related terms and was already successfully implemented by (Chang et al., 2020). However, this would mean a heavy generalization of the individual reading experience, which, as this research has shown, does vary even in the discussion of the same novel.

Lastly, the necessity of excluding Fantasy from the analysis does suggest that this ascription as used by the members of Goodreads is much broader than the other genres in this study. This is consistent with the results of Antoniak et al. (2021), whose classifier was more likely to fail in the categorization of Fantasy reviews than reviews of other genres. This suggests that future research should focus on addressing its subgenres rather than the entirety of Fantasy. Alternatively, an intriguing approach would involve narrowing the scope to exclusively examine the multifaceted elements encompassed within the Fantasy genre.

References

Allington, D. (2016). 'Power to the reader' or 'degradation of literary taste'? Professional critics and Amazon customers as reviewers of The Inheritance of Loss. *Language and Literature*, *25*(3), 254–278. https://doi.org/10.1177/0963947016652789

Antoniak, M., Walsh, M., & Mimno, D. (2021). Tags, Borders, and Catalogs: Social Re-Working of Genre on LibraryThing. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), 1–29. https://doi.org/10.1145/3449103

Bail, C. (2016). *Graph-Based Automated Text Analysis* (0.1.1) [R]. https://github.com/cbail/textnets

Bálint, K., Hakemulder, F., Kuijpers, M., Doicaru, M., & Tan, E. S. (2016). Reconceptualizing foregrounding: Identifying response strategies to deviation in absorbing narratives. *Scientific Study of Literature*, *6*(2), 176–207. https://doi.org/10.1075/ssol.6.2.02bal

Benoit, K., & Matsuo, A. (2020). *spacyr: Wrapper to the 'spaCy' 'NLP' Library*. https://CRAN.R-project.org/package=spacyr

Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., & Matsuo, A. (2018). quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software*, *3*(30), 774. https://doi.org/10.21105/joss.00774

Bessi, A., & Briatte, F. (2016). *disparityfilter: Disparity Filter Algorithm for Weighted Networks*. https://CRAN.R-project.org/package=disparityfilter

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(P10008). https://doi.org/10.1088/1742-5468/2008/10/P10008

Bondi, M., & Scott, M. (2010). *Keyness in Texts*. John Benjamins Publishing.

Boot, P. (2023). 'A pretty sublime mix of WTF and OMG'. Four explorations into the practice of evaluation on online book reviewing platforms. *Journal of Cultural Analytics*, *7*(2). https://doi.org/10.22148/001c.68086

Busselle, R., & Bilandzic, H. (2009). Measuring Narrative Engagement. *Media Psychology*, *12*(4), 321–347. https://doi.org/10.1080/15213260903287259

Chang, K., Hu, Y., Shang, W., Sharma, A., Singhal, S., Underwood, T., Witte, J., & Wu, P. (2020). Book Reviews and the Consolidation of Genre. *ADHO 2020*. DH2020. https://hcommons.org/deposits/item/hc:31913/

Choi, Y., & Joo, S. (2020). Identifying Facets of Reader-Generated Online Reviews of Children's Books Based on a Textual Analysis Approach. *The Library Quarterly*, *90*(3), 349–363. https://doi.org/10.1086/708962

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal*, *Complex Systems*, 1695.

Cuilla, K. (2022). *reactablefmtr: Streamlined Table Styling and Formatting for Reactable* (R package version 2.0.0). https://CRAN.R-project.org/package=reactablefmtr

Dasaprakash, K., & Shaikh, N. (2019). *Entity Extraction and Classification using SpaCy*. https://kaggle.com/code/curiousprogrammer/entity-extraction-and-classification-using-spacy

Doche, A., & Ross, A. S. (2022). 'Here is my shameful confession. I don't really "get" poetry': Discerning reader types in responses to Sylvia Plath's *Ariel* on Goodreads. *Textual Practice*, 1–21. https://doi.org/10.1080/0950236X.2022.2082516

Driscoll, B. (2016). Readers of Popular Fiction and Emotion Online. In K. Gelder (Ed.), *New Directions in Popular Fiction* (pp. 425–449). Palgrave Macmillan UK. https://doi.org/10.1057/978-1-137-52346-4_21

Driscoll, B., & Rehberg Sedo, D. (2019). Faraway, So Close: Seeing the Intimacy in Goodreads Reviews. *Qualitative Inquiry*, *25*(3), 248–259. https://doi.org/10.1177/1077800418801375

Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, *19*(1), 61–74.

Eder, M. (2017). Visualization in stylometry: Cluster analysis using networks. *Digital Scholarship in the Humanities*, *32*(1), 50–64. https://doi.org/10.1093/llc/fqv061

Fabry, R. E., & Kukkonen, K. (2019). Reconsidering the Mind-Wandering Reader: Predictive Processing, Probability Designs, and Enculturation. *Frontiers in Psychology*, *9*, 2648. https://doi.org/10.3389/fpsyg.2018.02648

Fantasy Books | *Goodreads*. (2023, May 3). Goodreads. https://www.goodreads.com/genres/fantasy

Fiction Books | *Goodreads*. (2023, May 3). Goodreads. https://www.goodreads.com/genres/fiction

Gadamer, H.-G. (2010). Hermeneutik I. Wahrheit und Methode: Grundzüge Einer Philosophischen Hermeneutik. Mohr Siebeck.

Hall, G. (2008). Empirical research into the processing of free indirect discourse and the imperative of ecological validity. In S. Zyngier, M. Bortolussi, A. Chesnokova, & J. Auracher (Eds.), *Directions in Empirical Literary Studies: In honor of Willie van Peer*. John Benjamins Publishing.

Herrmann, J. B. (2018). In a test bed with Kafka. Introducing a mixed-method approach to digital stylistics. *Digital Humanities Quarterly*, *011*(4).

Hoffstaedter, P. (1987). Poetic text processing and its empirical investigation. *Poetics*, *16*(1), 75–91. https://doi.org/10.1016/0304-422X(87)90037-4

Holur, P., Shahsavari, S., Ebrahimzadeh, E., Tangherlini, T. R., & Roychowdhury, V. (2021). Modelling social readers: Novel tools for addressing reception from online book reviews. *Royal Society Open Science*, *8*(12), 210797. https://doi.org/10.1098/rsos.210797

Hyland, K. (1998). Boosting, hedging and the negotiation of academic knowledge. *Text & Talk*, *18*(3), 349–382. https://doi.org/10.1515/text.1.1998.18.3.349

Iser, W. (1976). Der Akt des Lesens. Theorie ästhetischer Wirkung. Wilhelm Fink Verlag.

Jacobs, A. M. (2016). The scientific study of literary experience and neuro-behavioral responses to literature: Reply to commentaries. *Scientific Study of Literature*, *6*(1), 164–174. https://doi.org/10.1075/ssol.6.1.08jac

Jannidis, F., Pielström, S., Schöch, C., & Vitt, T. (2015). Improving Burrows' Delta – An empirical evaluation of text distance measures.

Jauß, H. R. (1996). Literaturgeschichte als Provokation der Literaturwissenschaft. In *Texte zur Literaturtheorie der Gegenwart*. Philipp Reclam jun.

Jockers, M. L. (2013). *Macroanalysis: Digital Methods and Literary History*. University of Illinois Press. http://ebookcentral.proquest.com/lib/umainz/detail.action?docID=3414227

Klie, J.-C., Bugert, M., Boullosa, B., Eckart de Castilho, R., & Gurevych, I. (2018). The INCEpTION Platform: Machine-Assisted and Knowledge-Oriented Interactive Annotation. *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*, 5–9. https://www.aclweb.org/anthology/C18-2002

Koolen, M., Boot, P., & van Zundert, J. J. (2020). Online Book Reviews and the Computational Modelling of Reading Impact. *Proceedings of the Workshop on Computational Humanities Research (CHR 2020)*, *2723*, 149–169.

Kuijpers, M. M. (2022). Bodily involvement in readers' online book reviews: Applying Text World Theory to examine absorption in unprompted reader response. *Journal of Literary Semantics*, *51*(2), 111–129. https://doi.org/10.1515/jls-2022-2055

Kuijpers, M. M., Douglas, S., & Bálint, K. (2021). Narrative Absorption: An Overview. In *Handbook of Empirical Literary Studies* (pp. 279–304). De Gruyter. https://doi.org/10.1515/9783110645958-012

Kuijpers, M. M., Hakemulder, F., Tan, E. S., & Doicaru, M. M. (2014). Exploring absorbing reading experiences: Developing and validating a self-report scale to measure story world absorption. *Scientific Study of Literature*, *4*(1), 89–122. https://doi.org/10.1075/ssol.4.1.05kui

Kuijpers, M. M., Lusetti, M., Lendvai, P., & Rebora, S. (2023). *Annotating for absorption in online book reviews*. https://docs.google.com/document/d/1zVnsJ6h0fxHoMbzDF9Za59HBL-V8ZSst/edit?usp=embed_facebook

Kuijpers, M. M., Lusetti, M., Renner, L., Ruh, L., Tadres, J., Vogelsanger, J., Rebora, S., & Lendvai, P. (2023). *Absorption in Online Reader Reviews Annotation Guidelines*. https://docs.google.com/document/d/1S7tTblCvQ-AJ-dqJaOqxqxM74APGiBR4h/edit?usp=embed_facebook

Kuijpers, M. M., Rebora, S., Lendvai, P., Lusetti, M., Ruh, L., Vogelsanger, J., & Ternes, T. (2023). *Absorption in Online Book Reviews. Presenting the English-Language AbsORB Metadata Corpus and Annotation Guidelines.* [Data set]. osf.io/kr4v6

Kuiken, D., & Douglas, S. (2017). Chapter 11. Forms of absorption that facilitate the aesthetic and explanatory effects of literary reading. In F. Hakemulder, M. M. Kuijpers, E. S. Tan, K. Bálint, & M. M. Doicaru (Eds.), *Linguistic Approaches to Literature* (Vol. 27, pp. 217–249). John Benjamins Publishing Company. https://doi.org/10.1075/lal.27.12kui

Kukkonen, K. (2020). Probability designs: Literature and predictive processing. Oxford University Press, USA.

Lin, G. (2023). *reactable: Interactive Data Tables for R* (R package version 0.4.3). https://CRAN.R-project.org/package=reactable

Manning, C. D., Raghavan, P., & Schutze, H. (2008). *Introduction to Information Retrieval*.

Nuttall, L. (2017). Online readers between the camps: A Text World Theory analysis of ethical positioning in *We Need to Talk About Kevin*. *Language and Literature: International Journal of Stylistics*, *26*(2), 153–171. https://doi.org/10.1177/0963947017704730

Nuttall, L., & Harrison, C. (2020). Wolfing down the Twilight series: Metaphors for reading in online reviews. *Contemporary Media Stylistics*, 35–60.

Päpcke, S., Weitin, T., Herget, K., Glawion, A., & Brandes, U. (2022). Stylometric similarity in literary corpora: Non-authorship clustering and *Deutscher Novellenschatz*. *Digital Scholarship in the Humanities*, fqac039. https://doi.org/10.1093/llc/fqac039

Pedersen, T. L. (2022). ggraph: An Implementation of Grammar of Graphics for Graphs and Networks. https://CRAN.R-project.org/package=ggraph

Pedersen, T. L., Pedersen, M. T. L., LazyData, T., Rcpp, I., & Rcpp, L. (2020). Package "ggforce". *Accelerating "Ggplot2." Version 0.3*, *2*.

Peplow, D., & Carter, R. (2023). Stylistics and real readers. In M. Burke (Ed.), *The Routledge Handbook of Stylistics* (pp. 472–488). Taylor & Francis.

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. https://doi.org/10.1371/journal.pone.0226708

R Core Team. (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rebora, S., Boot, P., Pianzola, F., Gasser, B., Herrmann, J. B., Kraxenberger, M., Kuijpers, M. M., Lauer, G., Lendvai, P., Messerli, T. C., & Sorrentino, P. (2021). Digital humanities and digital social reading. *Digital Scholarship in the Humanities*, *36*(2), ii230–ii250. https://doi.org/10.1093/llc/fqab020

Rebora, S., Kuijpers, M., & Lendvai, P. (2020). Mining Goodreads. A Digital Humanities Project for the Study of Reading Absorption. *Sharing the Experience: Workflows for the Digital Humanities. Proceedings of the DARIAH-CH Workshop 2019*. DARIAH-CH Workshop 2019, Neuchâtel: DARIAH-CAMPUS.

Salgaro, M. (2021). The History of the Empirical Study of Literature from the Nineteenth to the Twenty-First Century. In *Handbook of Empirical Literary Studies* (pp. 515–542). De Gruyter. https://doi.org/10.1515/9783110645958-020

Savolainen, R. (2019). Sharing information through book reviews in blogs: The viewpoint of Rosenblatt's reader-response theory. *Journal of Documentation*, *76*(2), 440–461. https://doi.org/10.1108/JD-08-2019-0161

Serrano, M. Á., Boguñá, M., & Vespignani, A. (2009). Extracting the multiscale backbone of complex weighted networks. *Proceedings of the National Academy of Sciences*, *106*(16), 6483–6488. https://doi.org/10.1073/pnas.0808904106

Sievert, C. (2020). Interactive Web-Based Data Visualization with R, plotly, and shiny. Chapman and Hall/CRC. https://plotly-r.com

Steiner, A. (2008). Private Criticism in the Public Sphere: Personal Writing on Literature in Readers' Reviews on Amazon. *Participations*, *5*(2).

Swann, J., & Allington, D. (2009). Reading groups and the language of literary texts: A case study in social reading. *Language and Literature*, *18*(3), 247–264. https://doi.org/10.1177/0963947009105852

Thissen, B. A. K., Menninghaus, W., & Schlotz, W. (2018). Measuring Optimal Reading Experiences: The Reading Flow Short Scale. *Frontiers in Psychology*, *9*. https://www.frontiersin.org/articles/10.3389/fpsyg.2018.02542

Tselenti, D., Cardoso, D., & Carvalho, J. (2023). Constructing Sexual Victimization: A Thematic Analysis of Reader Responses to A Literary Female-on-Male Rape Story on *Goodreads*. *The Journal of Sex Research*, 1–15. https://doi.org/10.1080/00224499.2023.2172135

Walsh, M., & Antoniak, M. (2021). The Goodreads "Classics": A Computational Study of Readers, Amazon, and Crowdsourced Amateur Criticism. *Journal of Cultural Analytics*, *6*(2). https://doi.org/10.22148/001c.22221

Whiteley, S., & Canning, P. (2017). Reader response research in stylistics. *Language and Literature*, *26*(2), 71–87. https://doi.org/10.1177/0963947017704724

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Wickham, H. (2022). *stringr: Simple, Consistent Wrappers for Common String Operations* (R package version 1.4.1). https://CRAN.R-project.org/package=stringr

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wickham, H., François, R., Henry, L., & Müller, K. (2022). *dplyr: A Grammar of Data Manipulation* (R package version 1.0.10). https://CRAN.R-project.org/package=dplyr

Wickham, H., & Girlich, M. (2022). *tidyr: Tidy Messy Data* (R package version 1.2.1). https://CRAN.R-project.org/package=tidyr

Wickham, H., Hester, J., & Bryan, J. (2022). *readr: Read Rectangular Text Data* (R package version 2.1.3). https://CRAN.R-project.org/package=readr

Young Adult Books | Goodreads. (2023, May 3). Goodreads. https://www.goodreads.com/genres/young-adult

Zhang, C., Tong, T., & Bu, Y. (2019). Examining differences among book reviews from various online platforms. *Online Information Review*, *43*(7), 1169–1187. https://doi.org/10.1108/OIR-01-2019-0037

Appendix

column	content
link	Identifier for title-page on goodreads: https://www.goodreads.com/book/show/"link"
round	Annotation round
title	Booktitle
review	unique numeric identifier of review
mode	Absorption/AbsorptionMention
presence	Present/Absent
SWAS_category	Dimensions of the updated SWAS schema: "Attention", "Emotional_Engagement", "Mental_Imagery", "Transportation" and "Impact"
SWAS_tag	tags of the new schema: alphanumeric code and conceptualization of the absorption subcategory including SWAS-Rel. (e. g. "A7 Anticipation")
statement	the annotated part of the review
onset	String index of the onset of the annotation inside the review
offset	String index of the offset of the annotation inside the review
genre_1, genre_2, genre_3	include 1st, 2nd and 3rd genre as ascribed by goodreads-readers, including specified genres such as: "Historical: Historical Fiction"
votes_1, votes_2, votes_3	include number of votes for the respective genre ascription

Table 16: Structure of the dataset

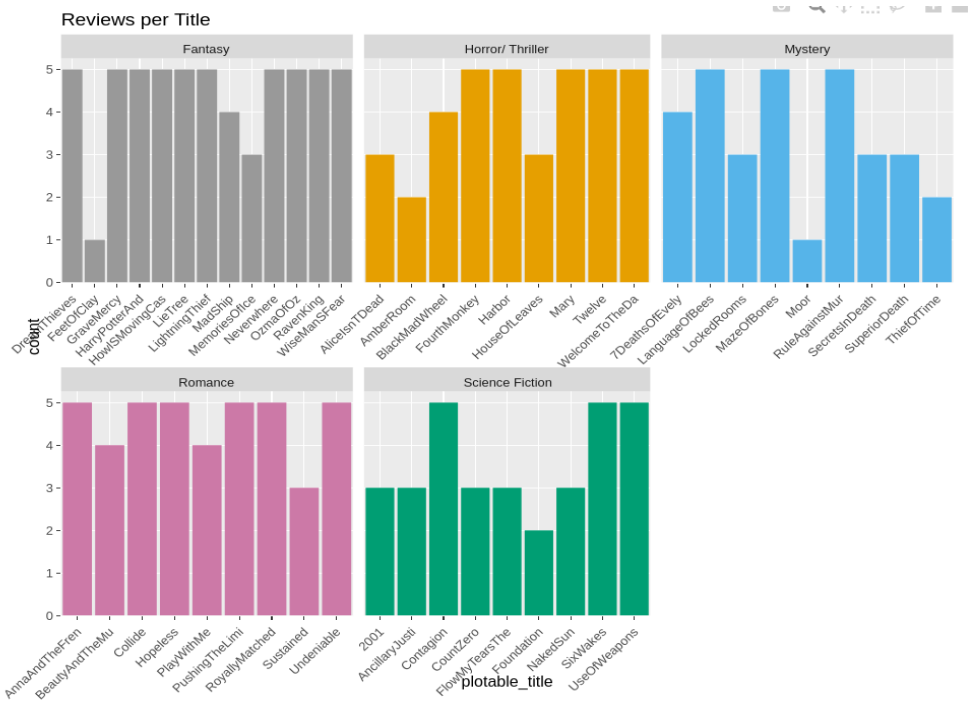

Figure 15: Number of reviews per title

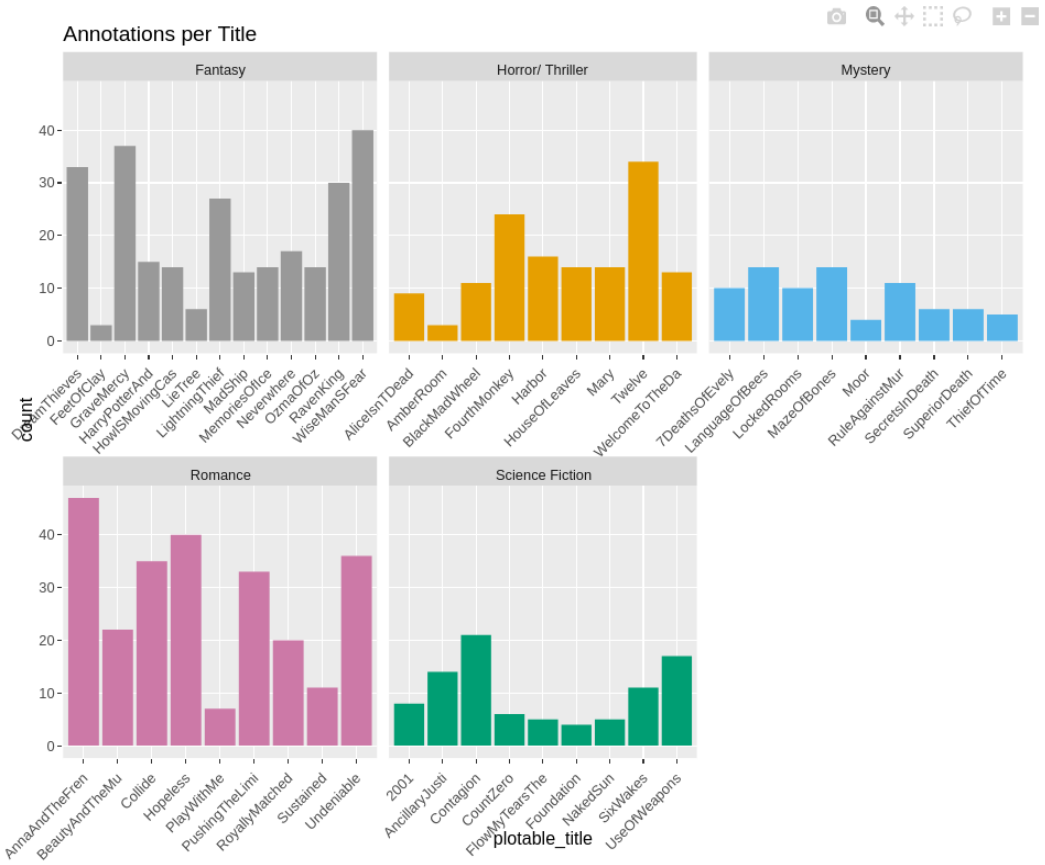

Figure 16: Number of annotations per title

docname	pre	key word	post
Main characters male			
Undeniable759679	Deuce is an asshole , yet I	love	him
Undeniable759679	And I	loved	watching him come to fully realise the hold she had
BeautyAndTheMu764001	would be the epitome of what I'd	love	in a man .
PushingTheLimi772578	(I loved Echo and was IN	LOVE	with Noah)
Hopeless781140	I	love	how Holder always took forever to think about every word
Hopeless781140	I	love	how Holder can be the breath that Sky so desperately
Hopeless781140	I	love	how hopelessly romantic and incredibly flirty Holder is
Hopeless781140	I	LOVE	DEAN HOLDER . I WANT HIM , I NEED HIM
PushingTheLimi784911	I frikkin '	LOVED	him to pieces (and so will you
BeautyAndTheMu991286	I fell in	love	right along with Ash
Main characters female			
Undeniable759679	I	loved	the strength and the resilience that she showed
PushingTheLimi772578	(I	loved	Echo and was IN LOVE with Noah)
Hopeless781140	I	love	how strong Sky is
Characters General			

PushingTheLimi755149	characters who I feel in	love	with .
PushingTheLimi772578	I	loved	them both equally and individually
Hopeless781140	I	love	how real Sky and Holder were
Hopeless781140	I	love	how in the depths of their despair
Hopeless794091	I am absolutely in	love	with Sky and Holder
Hopeless794091	I	love	how they never hold anything back
PlayWithMe796976	That's what I	loved	about the main characters in this book
BeautyAndTheMu967822	I	love	these characters and the way Penny writes
BeautyAndTheMu991286	I	love	the hillbilly brothers
Book			
PushingTheLimi755149	I wasn't expecting to	love	it quite as much as I did
PushingTheLimi755149	I think the reason I	loved	Pushing the Limits so much is that I could relate
Undeniable759679	I	loved	everything about it ! 5 fucked up , dysfunctional stars
PushingTheLimi772578	I could ramble on and on about how much I	loved	this book
Hopeless781140	I	love	everything about this book .
Hopeless781140	I	love	how honest and passionate , and raw and pure this book is.
Hopeless781140	I JUST	LOVE	THIS BOOK . . . plain and simple

Hope-less781140	I could tell you a million other reasons why I	love	this book and why you should read it
Beau-tyAndTheMu991286	every time I think I can't	love	the next book as much as I did the last
Undenia-ble759679	And as they get older , I	loved	seeing Eva fall deeper and deeper for Deuce
Hope-less781140	I	love	that in the mist of all the many firsts
Beau-tyAndTheMu764001	I	love	how different each KitC story is
Author			
PushingTh-eLimi755149	you're a fan of Simone Elkeles then you're going to	love	Katie McGarry !
Beau-tyAndTheMu764001	I	love	where her stories take me !
Hope-less781140	I	love	how Colleen Hoover can take a word like " live “
Content			
PushingTh-eLimi755149	with a soft side that is brought out by his	love	for his brothers and Echo
Undenia-ble759679	life-altering connection that de-velops into intense attraction and overwhelming	love	in the face of pain , betrayal , guilt ,
Undenia-ble759679	And I	loved	him still . That had never changed . "
Undenia-ble759679	I don't know , but Eva	loved	him through it all
Undenia-ble759679	and he	loved	her back intensely
Undenia-ble759679	to earn the	love	of a good woman
Beau-tyAndTheMu764001	hootiedoom (my new favorite word) , and	love	!

BeautyAndTheMu764001	She had a mother who	loved	her and rascal brothers
BeautyAndTheMu764001	brothers who annoyed her as siblings do but still always	loved	her
BeautyAndTheMu764001	trustworthy man who would do anything for the people he	loves	.
PushingTheLimi772578	He	loved	his friends and family so fiercely
PushingTheLimi784911	her innocence and pain and her desire to be	loved	.
PushingTheLimi784911	And the	love	he has for his brothers is the CUTEST damn thing
PushingTheLimi784911	Their	love	for each other slowly built up
PushingTheLimi784911	how Echo taught Noah to	love	again
Hopeless794091	The	love	that they have for each other is incredibly strong
Hopeless794091	trusting in the people that you	love	, and who love you .
Hopeless794091	and who	love	you . " I swear I will spend every last
Hopeless794091	thanking you for allowing yourself to	love	me . "
PlayWithMe796976	three months later with I'm sorry's and I	love	you'd , it was amazing .
PlayWithMe796976	it was full of	love	, romance , anticipation , surprise and passion
Genre			
PushingTheLimi755149	anyone who loves a smouldering , sexy , consuming	love	story to boot !

PushingTheLimi772578	The plot was a perfect balance of an intense	love	story and a heartbreaking story of tragic loss
PushingTheLimi772578	a beautiful	love	story , mystery , suspense , a story of loss
PushingTheLimi772578	I read tons of angst filled YA	love	stories
Experience			
Hopeless781140	I	love	how the smallest gesture of " pinkie holding " feels
Hopeless781140	your heart is being held in the palm of your	lovers	hand . I love , oh how I love
Hopeless781140	I	love	, oh how I love , how " no kissing “
Hopeless781140	how I	love	, how " no kissing " can turn out to
Hopeless781140	I	love	how every moment shared in the book between Sky
Hopeless781140	I	love	how my heart skipped a beat almost on every page
Hopeless781140	I	love	how I devoured this book up
Hopeless794091	laugh , cry , squee and swoon , and I	loved	every single freaking minute of it ! ! !
BeautyAndTheMu991286	I contemplate , I lust , I fall in	love	, i google , I envy and most of all
PushingTheLimi772578	brought on so many different emotions for me : young	love	giddiness , frustration , anger and intense sadness . I
Other			
PushingTheLimi755149	the praise it's received from amazing authors and bloggers I	love	, I wasn't expecting to love it quite as much
PushingTheLimi755149	the feeling of somebody who you	love	and trust intentionally hurting you

PushingTh-eLimi755149	to anyone who	loves	a smouldering , sexy , consuming love story to boot
Undenia-ble759679	Men I	loved	with all my heart
Beau-tyAndTheMu764001	I didn't remember how much I	loved	poetry until I met this mountain man .
Beau-tyAndTheMu967822	in search of a Winston brother or Ranger . I	loved	all the Nietzsche quotes , it took me right back
Beau-tyAndTheMu967822	to Google as much in this one . I would	LOVE	more from the Winston brothers ! I am ready to

Table 17: Concordance of "love" in cluster 11, blue highlighting indicates absorption statements

docname	pre	key word	post
Talking about men in General			
Undeniable759679	There are weak	men	; men who run and hide when life slaps them
Undeniable759679	There are weak men ;	men	who run and hide when life slaps them
Undeniable759679	Then there are	men	; men who have a backbone yet occasionally
Undeniable759679	Then there are men ;	men	who have a backbone yet occa-sionally
Undeniable759679	And then there are real	men	; men who don't cry or complain
Undeniable759679	And then there are real men ;	men	who don't cry or complain
Undeniable759679	they are the backbone .	Men	who make their own decisions and live with the consequences
Undeniable759679	, who accept responsibility for their actions or words .	Men	who , when life slaps them in the ass
Undeniable759679	slap back and move on .	Men	who live hard and die even harder

Undeniable759679	Men who live hard and die even harder .	Men	like my father and my uncles
Undeniable759679	Men like my father and my uncles .	Men	I loved with all my heart
BeautyAndTheMu764001	would be the epitome of what I'd love in a	man	. Quiet and humble on the surface
PlayWithMe796976	I think vulnerable	men	are my weak spot .
PlayWithMe796976	Once again , why haven't I met	men	like him ?
BeautyAndTheMu991286	Mountain	men	with beards - the new new age sexy man
BeautyAndTheMu991286	Mountain men with beards - the new new age sexy	man	- who'd da thunk it !
Talking about characters			
Undeniable759679	Men I loved with all my heart .	Men	like Deuce
Undeniable759679	the nature of the clubs , and the	men	and women who exist within this amazingly unique world
Undeniable759679	as a result of being tied to another	man	who " fucked me up so badly
Undeniable759679	" The most beautiful	man	I'd ever seen and still the biggest asshole
Undeniable759679	Not many chances a	man	has in his lifetime to do right
BeautyAndTheMu764001	Drew is a	man	after my heart
BeautyAndTheMu764001	until I met this mountain	man	. See how involved I am
BeautyAndTheMu764001	He seems like a	man	of contradictions
BeautyAndTheMu991286	I expected an alpha mountain	man	, but what I got was a soulful artist
BeautyAndTheMu764001	a strong , smart , loving , and trustworthy	man	who would do anything for the people he loves .

Table 18: Concordance of "man"/"men" cluster 11; blue highlighting indicates absorption statements

docname	pre	key word	post
Main characters male			
Sustained792596	Chase really showed Jake's emotions , subtly growing . I	loved	him , but he creates a bit of turmoil ...
RoyallyMatched809927	the perfection of Henry and Sarah ! I knew I'd	love	Henry , I knew Brent's position in my Emma's man
RoyallyMatched872433	so sweet and charming but wild and naughty , I	loved	his flirting and his dirty , teasing mouth . But
Main characters female			
RoyallyMatched848448	. Lady Sarah's story was really endearing . I really	loved	these shy girls with a romantic bone in their body
Characters General			
RoyallyMatched809927	the most amazing characters and I always end up in	love	and obsessed and crazy for them ! ! Royally Matched
RoyallyMatched818332	. They're so perfect for each other ! I just	loved	them together ! ♥ I laughed , I chuckled and
RoyallyMatched872433	were broken and suffering in their own way . I	loved	them remeeting and the connection they had . Henry is
Book			
Collide775077	it has a lot of controversy around it , I	loved	it . This is a story about betrayal and suffering
Sustained792596		Loved	this hard core ! Emma Chase does it again with
Sustained792596	down and enjoyed it from start to finish . I	loved	this so much and I think it hit on just
Sustained792596	, he would do anything for his friends . I	love	how he met Chelsea and that instant chemistry they had
Sustained792596	man Jake is . It was really wonderful . I	loved	getting to see Stanton , Sophia , and Brent as
RoyallyMatched809927	. * Sighhhhhh * I think you already know I	loved	this book , like crazy . So there's no need

RoyallyMatched809927	romantic and has a great amount of drama that I	loved	with all my heart . The characters from Royally Screwed
RoyallyMatched848448	got from this book . Some advice , right .	LOVED	IT ! Ok sooooo , Prince Henry and Lady Sarah
RoyallyMatched872433	of swooney moments . The writing was prefect . I	loved	the fantastic build it had . This was such a
RoyallyMatched872433	boy to love and grow and be fixed . I	love	that this is dual POV and we get both Henry
RoyallyMatched872433	Sarah's perspectives . They have so many hardships but I	loved	watching them come together and grow together , they're the
RoyallyMatched872433	, and becoming better versions of themselves together . I	loved	this from beginning to end . I laughed and cried
RoyallyMatched872433	, and solid story filled with incredible details . I	loved	the humor and the fun but there was an undercurrent
Author			
Sustained792596	back tears during the last chapter and epilogue . I	love	how Emma Chase really showed Jake's emotions , subtly growing
Content			
Collide775077	story about betrayal and suffering , but also one about	love	and second chances . The main plot surrounds Emily Cooper
Collide775077	inserted in book with a certain purpose . Slowly a	love	triangle is formed but there were circumstances regarding the events
Collide775077	Dillon's side , no tenderness , respect and apparently no	love	. In opposition to that , Gavin loves Emily sincerely
Collide775077	apparently no love . In opposition to that , Gavin	loves	Emily sincerely and will do anything to make her happy
Collide775077	you . This book had one of the most intense	love	scenes I have ever read , without being too much
Collide775077	, too exaggerated . It had the right amount of	love	, sensuality and passion that needed to be included in
Collide775077	hot . That intensity was felt not only in the	love	scenes , but also in the relationship between Emily and

Sustained792596	was such a wonderful story with caring characters who find	love	and hope amidst chaos . So heart-felt , super sexy
RoyallyMatched818332	is now the future King , because his brother chose	love	over royalism ☺ But Henry never wanted the burden his
RoyallyMatched809927	got to met the real Henry , the man who	loved	to play guitar , who admire his brother and grandmother
RoyallyMatched809927	no other but was too afraid to disappoint , who	loved	his people and his country , and most of all
RoyallyMatched872433	sweet , and witty but so full of heart and	love	. I couldn't wait to start and dive right in
RoyallyMatched872433	woven romantic tale about two broken people , falling in	love	, growing , and becoming better versions of themselves together
Experience			
Sustained792596	the story . This was just so fantastic ! I	love	when the world can fade away and I'm pulled into
Other			
RoyallyMatched809927	is get your copy of Royally Matched and fall in	love	with Henry and Sarah . And if by any chance
RoyallyMatched872433	it to end all at the same time . I	loved	Royally Screwed and Prince Nicholas was fantastic but there was
RoyallyMatched872433	and learn everything . I wanted that broken boy to	love	and grow and be fixed . I love that this

Table 19: Concordance of "love" cluster 12; blue highlighting indicates absorption statements

Acknowledgements

I would like to take a moment to express my heartfelt appreciation to all those who have contributed to the completion of this thesis.

First, I want to thank the SHARD team at the university of Basel for providing me with this fascinating dataset. Furthermore, I want to express my deepest gratitude to Moniek Kuijpers for her amazing supervision. I am grateful for her guidance and encouragement, her expertise, insightful feedback, and the countless discussions that have shaped this work. I would also like to acknowledge the assistance and support received from Simone Rebora and Anastasia Glawion, who were kind enough to share their technical expertise and have played a significant role in the success of this research. A special thank you goes out to Anna Monschau and Patrick Teichmann, who dedicated their time and meticulous attention to proofread this work. Lastly, I want to thank my friends and family, especially my husband Maurice, for their unconditional love, belief in my abilities, and unwavering encouragement.

JOHANNES GUTENBERG UNIVERSITÄT MAINZ

ERKLÄRUNG

gemäß § 19 Abs. 3 i. V. m. § 15 Abs. 9 der Ordnung der Fachbereiche 02, 05 und 07 der Johannes Gutenberg-Universität für die Prüfung in den Masterstudiengängen der geltenden Prüfungsordnung

Hiermit erkläre ich,

Name, Vorname: Ternes, Tina

Matrikelnummer: 2754879

dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen oder Hilfsmittel (einschließlich elektronischer Medien und Online-Quellen) benutzt habe. Mir ist bewusst, dass ein Täuschungsversuch oder ein Ordnungsverstoß vorliegt, wenn sich diese Erklärung als unwahr erweist.

31.06.2023 — Datum

 — Unterschrift

Auszug aus der Ordnung der Fachbereiche 02, 05 und 07 der Johannes Gutenberg-Universität für die Prüfung in Masterstudiengängen

§ 19 - Versäumnis, Rücktritt, Täuschung, Ordnungsverstoß

(3) Versucht die Kandidatin oder der Kandidat das Ergebnis einer Prüfung durch Täuschung oder Benutzung nicht zugelassener Hilfsmittel zu beeinflussen, oder erweist sich eine Erklärung gemäß Absatz 5 als unwahr, gilt die betreffende Prüfungsleistung als mit „nicht ausreichend“ (5,0) absolviert. Stört eine Kandidatin oder ein Kandidat den ordnungsgemäßen Ablauf einer Prüfung, kann sie oder er von der jeweiligen Prüferin oder dem jeweiligen Prüfer oder Aufsichtführenden in der Regel nach Abmahnung von der Fortsetzung der Prüfungsleistung ausgeschlossen werden; in diesem Fall gilt die betreffende Prüfungsleistung als mit „nicht ausreichend“ (5,0) absolviert. In schwerwiegenden Fällen kann der zuständige Prüfungsausschuss die Kandidatin oder den Kandidaten von der Erbringung weiterer Prüfungsleistungen ausschließen.

(4) Die Kandidatin oder der Kandidat kann innerhalb einer Frist von einem Monat verlangen, dass Entscheidungen nach Absatz 3 Satz 1 und 2 vom zuständigen Prüfungsausschuss überprüft werden. Belastende Entscheidungen sind der Kandidatin oder dem Kandidaten unverzüglich schriftlich mitzuteilen, zu begründen und mit einer Rechtsbehelfsbelehrung zu versehen. Der Kandidatin oder dem Kandidaten ist vor einer Entscheidung Gelegenheit zur Äußerung zu geben.

(5) Bei schriftlichen Prüfungsleistungen gemäß § 13 mit Ausnahme von Klausuren hat die oder der Studierende bei der Abgabe der Arbeit eine schriftliche Erklärung vorzulegen, dass sie oder er die Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt hat. Erweist sich eine solche Erklärung als unwahr oder liegt ein sonstiger Täuschungsversuch oder ein Ordnungsverstoß bei der Erbringung von Prüfungsleistungen vor, gelten die Absätze 3 und 4 entsprechend.