

Exploring Politics Books Bestsellers on Amazon IT: Major topics and emotions expressed in reviews

Enrique Taietta
Matricola 257230

Data Science, University of Trento
13/06/2025

Introduction

Books represent a fundamental vehicle for disseminating ideas, concepts, and worldviews. In the political sphere, the social repercussions of books are even more evident and significant: they can serve as tools of propaganda, catalysts for change, or vehicles of dissent, influencing public opinion, consolidating dominant ideologies, or fueling movements of resistance and social transformation.

The digital context has profoundly altered how cultural content is produced, distributed, and received, introducing new dynamics of visibility, recommendation, and user interaction [1]. In this scenario, certain global platforms play a central role not only in the commercialization of books but also in constructing the social and political meanings surrounding them. Among these, Amazon stands out, having deeply influenced how books are approached as cultural products in recent years.

It is therefore essential to delve into the concept of platformization and its implications. Cultural production is becoming increasingly dependent on platforms, and some of its forms are progressively shaped by the logic these platforms impose. By their nature and multiperspective orientation, platforms can be conceived as networks, within which actions and interactions at one node can significantly affect other nodes within the same network. For example, interactions between cultural content producers and new users joining the platform, as well as user reviews, can influence additional users and even the producers themselves.

Over time, various sentiments or shared attitudes can emerge around a book, even creating an actual myth surrounding it. Books have the potential to foster political mobilization, regardless of their specific roles, dynamics, or narratives within communities or social struggles. While acknowledging the importance of these factors, it's worth noting that books can serve as communication tools for critical communities both within and outside their respective settings. The narratives built around a book can also influence the mobilization, development, and cohe-

sion of social movements. Books may, in fact, offer a shared language and serve as a reference, providing resources and arguments that can circulate within a movement [2].

This study focuses on political books and Amazon as a platform for the consumption and dissemination of books.

The aim is to analyze how Amazon influences the dissemination and reception of cultural content, as well as the visibility and circulation of politically charged books. In particular, this study investigates whether Amazon, through its architecture and algorithms, may contribute to generating forms of radicalization or polarization, as has already been observed on other digital platforms.

This study assumes the premise that Amazon can be considered a full-fledged social network, characterized by users who not only purchase products but also actively interact within a social network. The platform allows users to interact through product ratings and reviews, to which others can respond, for example, by marking them as "useful."

It is also possible to view each user's reviewed items, which can be likened to a personal bulletin board or homepage. The recommendation systems for additional products resemble, in algorithm logic, those present on platforms like YouTube.

This set of activities constitutes a network of social signals which, though not configured as direct or dialogic interactions, contribute to the construction of reputation, influence, and visibility of content, with potentially significant effects on sales trajectories and public perception of books.

In reading cultures, books have historically served as markers of taste and status, visibly displayed on personal bookshelves. This symbolic role persists today and can also be observed in digital practices, for example, on the Goodreads platform through its "shelves" function [3].

The interaction between books and the reading world—intrinsically social and shared—highlights how the future of writing and publishing will inevitably be influenced by platformization. In this regard, Amazon's publishing ecosystem represents an emblematic case of the cultural production processes

described in the existing literature, which identifies five specific processes: resorting, transmutation, enclosure, surveillance, capture [4].

Amazon thus acts as a mediator between the stakeholders involved on its platform—publishers, authors, and readers—reordering cultural content through algorithms such as sales rankings, personalized recommendations, and evaluation systems like user reviews, shaping the content itself and not just facilitating the interaction between the actors. This aspect becomes particularly significant when extreme or polarizing content, by generating more activity on the platform, receives greater visibility and algorithmic promotion.

The phase of transmutation proves crucial, as the platform's logics end up transforming the cultural forms they are meant to merely mediate. An example of this is the adaptation of literary genres to the tastes detected through the platform's continuous data analysis. In this context, the role of self-publishing assumes growing importance, both for the accelerated pace of production and for content serialization, thus also altering writing and production modes of the book itself.

All these aspects are particularly relevant considering Amazon's global reach and the resources at its disposal, as these mechanisms also apply to the production and distribution of politically charged books and cultural content.

Given the social nature of the platform, especially regarding the circulation of political or ideological content, it is necessary to investigate the potential emergence of echo chambers or filter bubbles due to the platform's algorithmic characteristics, as already observed in studies conducted on other digital platforms, as YouTube [5].

Building on these premises and combining them with theories that view books as objects of cultural capital, this research aims to investigate whether the content of politically charged books, analyzed through user reviews, reflects or amplifies the current dominant cultural climate, and whether it is possible to extract emotions or collective attitudes towards different topics, identifying potential phenomena of polarization or the formation of ideological bubbles. Reviews are an important source of information, given that, the political ideology of reviewers can significantly influence the language used in online reviews [6].

A central case study is represented by the social phenomenon surrounding the book *Il mondo al contrario* by Roberto Vannacci. The objective is to verify whether this phenomenon can be detected through observable dynamics on Amazon and whether there are elements to interpret it as a possible echo cham-

ber. In particular, the study seeks to analyze whether the polarization of content and reactions—geared toward reinforcing group identity and marking differences from otherness—can be read through the mechanisms proposed by social identity theory.

The main goal is not to qualitatively or formally assess the contents of Vannacci's book but to verify whether it exhibits the characteristics of a cultural product capable, on the one hand, of dividing and demarcating social groups and, on the other, of uniting by providing a common political language and resource. Books may serve as shortcuts for defining collective identity within a movement [2].

This conception of book as a cultural product intersects with theories of cultural capital, which manifest in how publishers and authors create and manipulate such capital throughout the various stages of the publishing process [7].

The analysis of Amazon reviews thus provides a privileged lens through which to observe the dynamics shaping public opinion in digital environments, highlighting emerging forms of political participation, mechanisms of selective visibility, and processes of symbolic exclusion.

This study contributes to the broader debate on the algorithmic governance of information and the influence of digital platforms in shaping the contemporary public sphere, offering a critical reflection on the role these cultural infrastructures play in defining what becomes visible, relevant, and legitimate in collective discourse.

Literature Review

Unlike other platforms, Amazon has been relatively understudied, particularly when conceptualized as a social network. Concerning the role of books, much of the existing literature focuses predominantly on reviews posted on Goodreads. In the present study, one of the guiding hypotheses is that Amazon reviews differ significantly from those found on other book review platforms such as Goodreads. While Goodreads heavily emphasizes user interaction, through the creation of personal bookshelves [3], on Amazon, quantitative aspects such as the number of ratings and reviews assume greater relevance in terms of interaction. These reviews may focus both on the book as an object and on the reading experience, and, in the case of politically charged books, may also contain evaluative commentary on the themes addressed by the work as political stance, with diverse emotional level.

As highlighted in the article **Goodreads vs Amazon** [8], different platforms offering similar activities

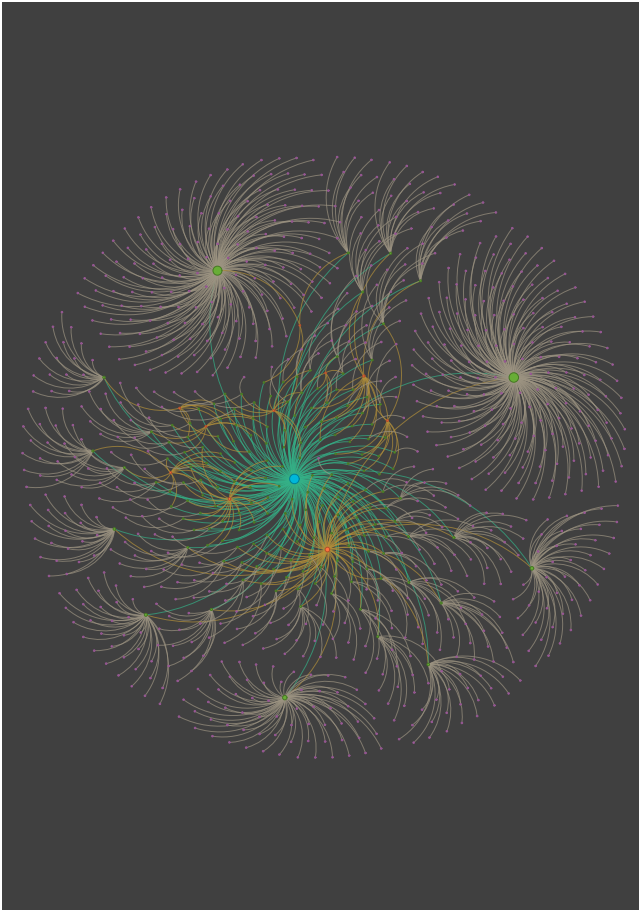


Figure 1: Subnetwork rooted at the node regarding the book "La banalità del male". The Blue Node is the B (book) node, the green nodes are the R (review) nodes, the Orange node are the T (Topic) nodes and the Purple nodes are the Nodes regarding the reaction to the review

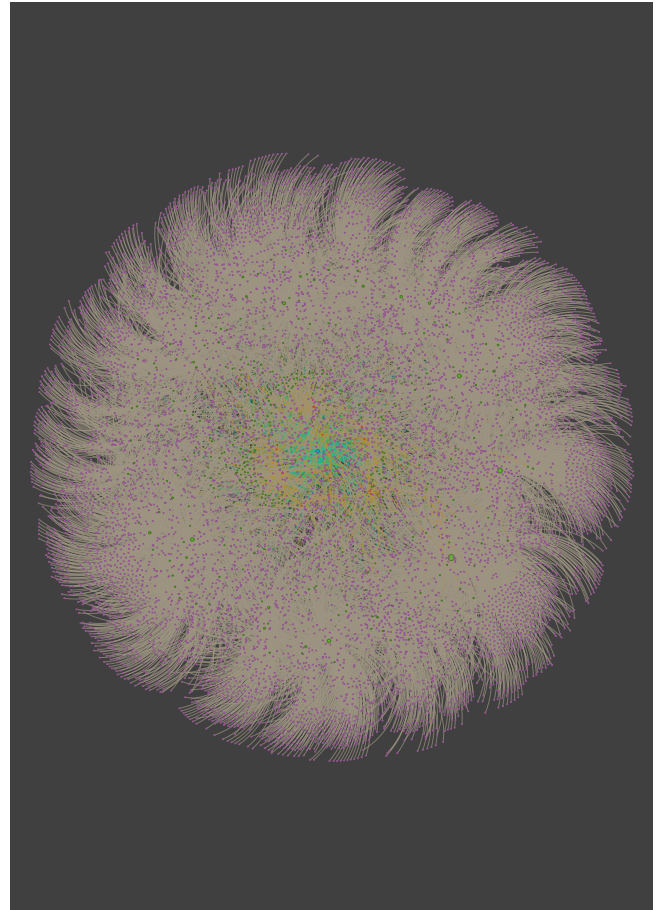


Figure 2: Subnetwork rooted at the node regarding the book "Il mondo al contrario". The Blue Node is the B (book) node, the green nodes are the R (review) nodes, the Orange node are the T (Topic) nodes and the Purple nodes are the Nodes regarding the reaction to the review

may elicit distinct user behaviors depending on the specific context in which they operate.

Given Amazon's commercial nature as a marketplace, reviews posted on the platform may reflect emotions and topics that diverge from those found on Goodreads. It is also hypothesized that users who choose to review or engage with one platform over another may belong to different reference groups in terms of attitudes, emotional expressiveness, motivations, and sociocultural backgrounds.

It is important to emphasize, as already noted in **Goodreads vs Amazon**, that users who engage with reviews or browse platform content may do so for varying reasons, often seeking different types of information.

In the study by Stefan Dimitrov and colleagues, data collection was carried out through web crawling of both Amazon and Goodreads, acquiring books and corresponding reviews in order to create two comparative datasets, which were then analyzed through sentiment analysis. The dataset was initially built on Goodreads, limiting the sample to the Biography category. Using ISBN codes, corresponding reviews from both Goodreads and Amazon were retrieved, along with user identifiers associated with those reviews. Among the strengths of this study is its treatment of Goodreads and Amazon as social networks, and the substantial volume of collected data, which enables comparisons in terms of review quantity, length, and user engagement. One of the key findings is that Amazon reviews tend to be significantly longer than those on Goodreads, while the sentiment polarity remains generally stable across both platforms. In this study, review length—one of the variables under examination—will receive further interpretation.

A common limitation shared by this and other sentiment analysis-based studies lies in the difficulty of conducting more nuanced social analyses when focusing solely on positive or negative polarity. It is posited that an emotion-based classification approach may offer richer interpretative possibilities, allowing for the detection of contextually relevant nuances [9].

Another relevant contribution that addresses the interaction between Goodreads and Amazon through review analysis is the study by Wang Kai **Exploring Goodreads Reviews for Book Impact Assessment** [10]. In this case, researchers investigate how reviews may influence the potential impact of books, emphasizing—also regarding prior studies—how reviews serve not only as reflections of reader attitudes and opinions but also as valuable sources

of information and reflection for future readers.

Similarly to Goodreads vs Amazon, the dataset was constructed starting from a book list, in this case, titles from the Book Citation Index (BKCI), followed by web crawling of Goodreads via book title searches. Unlike the previous study, users were categorized into three groups: *ordinary users*, *Goodreads authors*, and *Goodreads librarians*, with subsequent segmentation of reviews based on ratings.

Sentiment analysis was once again employed, using dedicated libraries to assess polarity. Among the findings relevant to the present study is the observed correlation between review length and sentiment: shorter reviews tend to express emotions more explicitly, whereas longer ones are generally more descriptive. This aspect will be further explored in the present analysis. As with previous studies, a key limitation is the inherent challenge of sentiment analysis in capturing the full complexity of textual polarity [11].

Another significant study focusing on Goodreads and, more broadly, on the role of digital platforms in reading practices, is the research by Mariana O. Silva and colleagues [12]. This work is valuable for the present discussion as it underscores how reading practices and book consumption are deeply shaped by social, economic, and cultural factors. Reading merges with ideas and influences from diverse social perspectives, and an economic standpoint, and functions as a form of consumption directly affected by socioeconomic variables.

Methodologically, this study is distinguished by its use of Network Analysis, employing a Multipartite Network model that maps user reading preferences by literary genre and geographical location. The application of structural network analysis is particularly significant in deriving conclusions. A notable limitation, however, is the absence of review-based and sentiment analysis data, which could have further enriched the study's insights.

An additional contribution directly relevant to the present work, particularly within the Italian context, is the study **A Lexicon-based Approach for Sentiment Classification of Amazon Books Reviews in Italian Language** [13]. In this research, the authors collected a dataset of 8,255 Italian-language reviews for 85 books across diverse authors and topics, employing a lexicon-based approach for classifying review polarity as either positive or negative. Although the study primarily focuses on lexicon-based sentiment analysis, it provides valuable

insights into Amazon book reviews, demonstrating once again that these reviews differ substantially from those of other product categories: they tend to be longer, more articulate, and more complex.

For instance, even low-rated reviews often contain positive remarks about Amazon's service, and vice versa. This phenomenon highlights a persistent limitation of sentiment analysis when applied to binary polarity classification.

Finally, as an additional methodological reference for the study of social phenomena through a computational and theoretically grounded approach, the article **Exploring the Impact of Social Network Structures on Toxicity in Online Mental Health Communities** [14] is considered.

Drawing from Rheingold's definition of online communities as social aggregations, the study integrates this concept with social capital theory, analyzing network connections, configurations, and participatory mechanisms. Of particular interest is the exploration of *cognitive social capital*, defined as "encompassing the shared norms, values, attitudes, and beliefs that promote mutual understanding among members of the community, essentially representing the collective consciousness." Within this theoretical framework, it is pertinent to explore whether it is possible to identify bounded communities through the analysis of attitudes and values expressed toward a cultural product, where individuals may cultivate shared beliefs and perspectives within the emergent community.

Of additional interest in this study is the data collection phase, which employed Communalistic for extensive data harvesting from Reddit, combined with external APIs such as Perspective API to enrich the dataset by assigning toxicity scores to texts. Notably, the authors focused their analysis on the r/MentalHealth subreddit, selected for its supportive characteristics, which offered a fitting context for investigating online toxicity.

Research Question

The study is structured around two principal lines of inquiry: one of an exploratory nature and the other of a descriptive-interpretative character.

The Emergence of Topics and the Structure of the Amazon Review Network

RQ1 : *What are the themes that emerge from reviews of Italian bestselling books on Amazon's category 'politics', and how are these elements distributed and interconnected*

within the network of user interactions?

From an exploratory perspective, the investigation aims to identify the main themes that emerge from the corpus of reviews of political books on Amazon, specifically focusing on Italian bestsellers, in the Politics category, and to assess whether these themes reflect, reinforce, or deviate from the dominant media narrative within the Italian political landscape. Furthermore, the study seeks to explore how and why certain reviews become more "active" or central within the network, that is, why they receive a higher number of user interactions (such as helpfulness votes). Employing Network Analysis techniques and descriptive statistical tools, the study examines the network structure composed of books, reviews, and user reactions. The following hypotheses are proposed:

- H1A: Books conveying extreme or right-wing ideological content exhibit higher levels of network activity and are connected to a greater number of topics.
- H1B: Books that share similar meta-paths toward specific topics tend to share a common ideological basis.

Emotions and Interactions: The Impact of Reviews within the Network

RQ2 : *How do textual and emotional characteristics of reviews influence the number of user interactions?*

In parallel, the interpretative component of the study focuses on the emotions expressed within the review texts, with the objective of assessing whether these emotions vary according to the topics addressed, and whether they may serve as indicators of collective sentiment, forms of social consciousness, or recurring patterns in relation to specific themes conveyed by the books.

For each review, the dataset includes the token count, the eight primary emotions as defined by Plutchik's theory, and any resulting emotional dyads. The following hypotheses are formulated:

- H2A: There is a correlation between the number of user reactions to a review and the word count of the review.
- H2B: There is a dependency between the number of reactions and the presence of strongly expressed emotions within the review.

Emotional Variation Across Themes

RQ3 : *How do the emotions expressed in reviews vary according to the themes emerging from the selected set of books?*

- H3: Different themes elicit distinct emotional profiles in the reviews, which can be identified and analyzed through a syntactic-semantic and emotional analysis approach.

Finally, it is hypothesized that the nature of the themes addressed influences the emotional intensity of the associated reviews. Since the books may cover a wide range of subjects (historical, political, personal, ideological), significant emotional differentiation across topics is anticipated.

The objective of the study is to assess whether Amazon's online reviews offer a valuable lens through which to observe social, ideological, and emotional phenomena related to reading, and in particular, whether they constitute a meaningful form of interaction capable of revealing positionalities, polarizations, and emerging forms of collective consciousness.

The study also seeks to explore whether certain types of content, particularly those with high ideological or emotional charge, generate more intense or polarized responses within the reader community.

Project Design

As previously mentioned, the research design of this study reflects a dual objective: to conduct an exploratory analysis of the structure of interactions and content within Amazon's review system and to develop a descriptive-interpretative investigation into the cognitive and emotional processes underlying users' reading experiences and participatory behaviors.

To meet these two objectives, the project is organized into two distinct phases:

- The first phase, focused on network analysis, employs an approach grounded in Social Network Analysis to represent Amazon's review system in the form of a network.
- The second phase utilizes tools from the field of Cognitive Data Science to examine the semantic, emotional, and textual characteristics of the reviews, relating these features to user behavior.

The integration of the results from both phases enables a comprehensive analysis of the interdependence between social dynamics and the cognitive and emotional aspects associated with the production and reception of reviews.

Network Representation In the first phase of the study, a Multilayer Network or Heterogeneous Information Network (HIN) was constructed to represent the review system of the platform. The network consists of four types of nodes:

- B: Book
- R: Review
- T: Topic (assigned through topic modeling)
- U: Reaction to review (labeled as "I Found it useful")

And three types of edges:

- B-R: represents the association between book and review
- R-T: association between a review and its prevailing topic
- R-U: association between a review and the user's helpfulness evaluation

A HIN is defined as a directed graph whose nodes and edges are drawn from multiple types, together with a network schema that specifies valid object and relation types, thus preserving rich semantic information across diverse entity and relation classes [15]. The theoretical framework of the network has been defined as a HIN, as it enables the explicit modeling of multiple types of nodes and relationships, with a particular focus on meta-paths that capture significant semantic dependencies among heterogeneous entities.

Topic and emotion Labeling of Reviews For the second part of the research, each review was enriched with a set of data and information to examine its textual and semantic features. Based on the texts obtained via web crawling on Amazon, the token count of each review was calculated to measure its length. Two labeling systems were then applied.

The first set of labels is based on Plutchik's model of eight primary emotions. Emotions were annotated by distinguishing whether their presence in the text was statistically significant, according to values above or below the confidence threshold (e.g., *joy-under*, *joy-above*, *anger-under*).

The second labeling system is based on the identification of so-called emotional dyads. These represent complex emotions emerging from the combination of two significantly expressed primary emotions. For example, the combination of *"trust-above"* and *"joy-above"* is interpreted as the complex emotion "Hope." This step was implemented using the EmoAtlas library, a tool that integrates psychological lexicons, semantic networks, and artificial intelligence elements for emotional text analysis.

EmoAtlas builds a network of words syntactically and semantically related to a target term, and derives its emotional profile by analyzing these contextual associations, including synonyms and hierarchical relations (hypernyms/hyponyms).

Emotional salience is assessed via z-scores, computed

by comparing the observed frequency of emotion-related words against a null distribution obtained through random sampling. Z-scores exceeding ± 1.96 (at $\alpha=0.05$) indicate statistically significant over- or under-expression of specific emotions relative to expected baseline levels [16] [17].

Finally, through topic modeling through latent Dirichlet allocation (LDA), each review was assigned the topic with the highest probability, thus connecting the texts to the semantic and ideological dimensions of the reviewed books.

Data Collection Strategy

The first step of the data collection—conducted independently without relying on pre-existing datasets or third-party services—involved defining the scope of the investigation. Given the vastness of Amazon's catalog, the focus was placed on the Bestsellers in the "Politics" category on Amazon Italy.

This decision was based on two primary considerations. First, the "Politics" category is likely to reflect a variety of themes and emotional reactions that would not emerge as strongly in other categories.

Second, one of the research objectives is to verify whether, by analyzing the reviews of political bestsellers, it is possible to obtain an overview of contemporary cultural consumption in Italy, including its emotional dimension.

Once the field of inquiry was defined, an initial dataset was created, composed of 100 records, each corresponding to a book listed in the bestseller ranking as of May 7, 2025. Each record includes the book's title, author, and ASIN (Amazon Standard Identification Number) code, all data gathered through web crawling and scraping methods.

It is important to consider the temporal constraint of the study: Amazon's bestseller rankings are dynamic and subject to frequent changes, which is why the analysis must be contextualized according to the date of data collection.

Based on the initial dataset, a script was developed to retrieve the reviews associated with each book ASIN on the dedicated product-reviews page. Due to limitations imposed by Amazon, access is restricted to a maximum of 100 reviews for each star rating (from 1 to 5), up to a total of 500 reviews per book. However, consistent with the nature of bestsellers, there is always a structural imbalance toward positive reviews, as 1-star ratings and their corresponding reviews rarely reach 100 entries. Despite this limitation, the final dataset—comprising 3,748 reviews—is considered sufficiently robust in both quantity and quality for meaningful analysis.

Each record on this dataset includes the ASIN of the book, the title of the review (Mandatory), the descrip-

tion of the review, and the number of reactions to the review, labeled as 'X persons find it useful'.

Also, the reviews were sorted by 'importance', in this phase of Data Collection, when retrieved. By default, Amazon retrieves reviews that have received the most interaction first; this mechanism suggests that the most relevant reviews—the primary focus of the investigation—tend to appear among the first 100 for each rating. It is important to note, however, that some highly polarizing books, such as "Il mondo al contrario" by Roberto Vannacci, display an anomalous number of reviews, significantly impacting global statistics—a phenomenon discussed in detail in the conclusions. The crawling and scraping scripts were written in Python, utilizing the Selenium and BeautifulSoup libraries.

In the subsequent phase, a prevailing topic was assigned to each review via Topic Modeling. After a preliminary phase of text cleaning and pre-processing including in the dictionary common and proper nouns (removal of high-frequency words, for example, 'book' that were observed as outliers, for a total of 5 words), an LDA (Latent Dirichlet Allocation) model was trained. After testing several training configurations, the number of topics was set at 10, based on qualitative considerations regarding the semantic interpretability of the resulting clusters. In the final section, it will be discussed how the phase of text cleaning and pre-processing could be improved, for example, for the misspelled words, using dedicated frameworks or LLMs APIs.

This phase was conducted using Python libraries and frameworks as Spacy and Gensim.

The identified topics cover areas such as geopolitics, identity and society, fascism, the book's product quality, and one topic entirely centered on Vannacci. Some topics are related more to the purchase experience and the book object itself rather than to ideological content, an interesting finding that will be further discussed.

The results of the topic modeling are summarized in the table below.

Topics extracted from LDA model

| Topic | Description | Top words |
|-------|--|--|
| 0 | Society and Identity | famiglia, idea, psicologia, destra, società, capitolo, pensiero, senso, figlio, d |
| 1 | Reading Experience | contenuto, racconto, genere, verità, acquisto, stella, insegnamento, berlinguer, libricino, pensiero |
| 2 | Geopolitics | Travaglio, situazione, fatto, guerra, Fabbri, Israele, evento, conflitto, realtà, geopolitica |
| 3 | Vannacci | pensiero, Vannacci, cosa, problema, opinione, argomento, concetto, persona, capitolo, società |
| 4 | Moral Judgment and Memory | punto, vista, processo, Eichmann, stile, argomento, riflessione, fatto, scuola, politica |
| 5 | Israeli-Palestinian Conflict | palestinese, Israele, Albanese, cultura, conflitto, Hamas, aspettativa, affermazione, pace, guerra |
| 6 | Book Quality | qualità, pagina, copertina, edizione, prezzo, carattere, prodotto, legge, potere, regalo |
| 7 | Ideological Analysis of Fascism | fascismo, analisi, Eco, Orsini, viaggio, pagina, informazione, prof, mente, passato |
| 8 | Purchase Experience | tempo, pagina, condizione, argomento, spedizione, consegna, palestina, spunto, tratto, situazione |
| 9 | History and Memory of Fascism in Italy | fascismo, Italia, Mussolini, volume, periodo, guerra, donna, violenza, Cazzullo, Scurati |

Following topic modeling, each review was assigned a topic, completing the B - R - U - T Multi-layer Network or Heterogeneous Information Network (HIN) structure.

For the network visualization was used Gephi, and for the network analysis it was used the NetworkX library from the Python environment.

For the second part of the study, focusing on cognitive and emotional perspectives, EmoAtlas was employed to detect the emotions present in the texts through z-score calculations based on Plutchik's eight primary emotions.

The EmoAtlas library was chosen because it works also on Italian language, if specified in the parameters.

For each review, two types of labels were generated:

- **Primary emotions**, whose z-score deviates statistically from the mean ($||z|| > 1.96$), indicating a pronounced emotional presence (positive or negative);
- **Complex emotions (dyads)** are obtained by combining pairs of significant primary emotions. Out of 28 theoretically possible combinations, only those in which both emotions displayed a significantly positive z-score were selected.

Note that, based on the context, different complex emotions can be expressed through the same dyads, in this study, for example, joy+fear, is expressed and differentiated into Excitement and Guilt.

This exploratory approach captures the complexity

of readers' emotional engagement, moving beyond a simple positive-negative review classification. The objective is to gain a deeper understanding of which emotions emerge in the reception of culturally and politically relevant content.

Methodology

The analysis will begin with an initial general exploration of the dataset, aiming to describe the characteristics of the analyzed reviews and books.

Quantitative aspects such as the average number of words per review, the distribution of reviews per book, and the number and type of emotions expressed in the texts will be assessed. This will allow for the identification of which books elicit the highest emotional engagement and which reviews are the richest in content.

Subsequently, attention will shift to network analysis, using Network Analysis tools (NetworkX, Gephi) to understand the importance and role of different nodes (books, reviews, topics). In particular, the weighted degree of topics and books will be examined, assessing the influence of reviews by also considering the number of received interactions (reactions), in order to identify whether and how the weight of reviews alters the relative importance of topics within the network. Through the number of meta-paths, the distribution of topics across books will be analyzed.

Additionally, the PathSim method will be applied

to measure the similarity between books based on shared topics through the meta-path B - R - T - R - B. This will help identify clusters of thematically similar books that may be relevant to ongoing social debates. PathSim is a meta-path-based similarity measure tailored for symmetric meta-paths in an HIN, designed to find peer objects by jointly considering both their direct connectivity and their individual visibilities [18]. Formally, for a symmetric meta-path P, PathSim between two nodes x and y is computed as

$$\text{PathSim}(x, y) = \frac{2 |\{\text{paths } x \rightarrow y\}|}{|\{\text{paths } x \rightarrow x\}| + |\{\text{paths } y \rightarrow y\}|}, \quad (1)$$

which normalizes the cross-node path count by their self-connectivities to emphasize balanced, peer-like relationships, such as books with comparable topic distributions.

From a socio-theoretical perspective, topics with a high degree represent those that attract more opinions and interactions within the network, while books with high similarity may play a significant role in disseminating ideas and stimulating reactions in the analyzed social context.

In a subsequent phase, it will be assessed whether the texts elicit significant emotions and which review characteristics (such as length and emotional intensity) influence or correlate with the number of user reactions. Furthermore, a statistical analysis will be conducted to evaluate whether the range of emotions expressed varies significantly across different topics, highlighting any notable differences between the topics addressed.

Results

Note: Tables 5 to 11 and Figures 3 to 9 are shown at the end of the document

Table 1 reports the top 10 books ranked by the degree of node B (corresponding to the book) within the network. Table 2 also presents the top 10 books, but in this case, the ranking is based on the number of B-R-U meta-paths (Book, Review, and User Reactions to the review).

In both tables, "Il mondo al contrario" occupies the first position; however, the overall order differs between the two cases, indicating that some books received a higher number of reactions to reviews despite having a similar number of reviews. It is worth noting that the B-R-U meta-path count for "Il mondo al contrario" by Vannacci is more than ten times higher than that of the second-ranked book ("La banalità del male" by Hannah Arendt). Although the

degree of type node B for Vannacci's book is approximately twice that of Arendt's work, the difference remains substantial. Figures 1 and 2, show the sub-network for the two books, "La banalità del Male" and "Il mondo al contrario".

Table 3 provides a summary of the final dataset, displaying for each book the description of its main topic, the list of primary emotions (if present), complex emotions (expressed as dyads), review length (in tokens), and number of reactions. The table is ordered by the number of reactions to reviews. Notably, all of the top 10 positions are occupied by "Il mondo al contrario".

Table 4 presents key distribution statistics for the number of reactions to reviews (USEFUL-NUM) and review length (in tokens). The minimum value of zero is due to the inclusion of reviews consisting solely of symbols or characters.

The maximum review length exceeds 2000 words, with an average of approximately 50. For both variables (USEFUL-NUM and token-len), percentiles reveal highly skewed distributions. For example, while the maximum number of reactions reaches 824, 50% of reviews have no reactions at all; the 95th percentile for review length is 164 tokens, compared to a maximum of 2139 tokens. Both distributions are strongly right-skewed.

This dynamic suggests that reviews accumulate reactions over time. A plausible interpretation, requiring further investigation, is that Amazon grants greater visibility to reviews with higher numbers of reactions, thereby creating a cumulative exposure effect that fosters additional reactions.

Table 5 displays the percentage of reviews for each book exceeding the 90th percentile. There is a marked predominance of "Il mondo al contrario", which accounts for approximately 30% of the longest reviews. This may also indicate heightened emotional engagement with the book, motivating users to write more detailed and elaborate reviews. As also observed in the study **A Lexicon-based Approach for Sentiment Classification of Amazon Books Reviews in Italian Language** [13], reviews of books often appear highly elaborate and complex, as though users were assuming the role of literary critics or political analysts.

Table 6 reports the percentage distribution of topics based on the number of T-R-U meta-paths, while Table 7 presents the degree centrality calculated for the nodes corresponding to topics.

Figure 3 shows the topic distribution percentage on 10 books. Tables 8, 10, and 11 display the results of the PathSim calculations for the symmetric meta-path B - R - T - R - B for three books, showing the top

| ASIN | Book title, Author | B node Degree |
|------------|---|---------------|
| B0CF4BJN3D | Il mondo al contrario, Roberto Vannacci | 421 |
| 8893881284 | Le 48 leggi del potere, Robert Greene | 263 |
| 8807897458 | La banalità del male, Hannah Arendt | 181 |
| 8804751169 | Mussolini il capobanda, Aldo Cazzullo | 178 |
| 8830104957 | M. L'uomo della provvidenza, Antonio Scurati | 171 |
| 8804665289 | Gomorra. Viaggio nell'impero economico..., Roberto Saviano | 168 |
| 8893442418 | Il fascismo eterno, Umberto Eco | 158 |
| B0CLML6FM8 | Israele e i palestinesi in poche parole, Marco Travaglio | 150 |
| 8896985625 | Le parole sono finestre (oppure muri), Bertram Rosenberg Marshall | 144 |
| B076H8MFMJ | Sovietistan: Un viaggio in Asia centrale, Erika Fatland | 133 |

Table 1: First 10 books ordered by B-type node degree

| ASIN | Book title, Author | Meta-path B-R-U count |
|------------|--|-----------------------|
| B0CF4BJN3D | Il mondo al contrario, Roberto Vannacci | 9652 |
| 8807897458 | La banalità del male / Hannah Arendt | 845 |
| 8804751169 | Mussolini il capobanda, Aldo Cazzullo | 739 |
| 8893881284 | Le 48 leggi del potere, Robert Greene | 724 |
| 8804776455 | Grazie, Occidente, Federico Rampini | 634 |
| B0CB237DYY | J'accuse, Francesca Albanese | 525 |
| B00DJ2KYQQ | Il capitale (eNewton Classici), Karl Marx | 379 |
| 8807900521 | L'arte della guerra, Tzu Sun | 342 |
| B0CLML6FM8 | Israele e i palestinesi in poche parole, Marco Travaglio | 334 |
| 8893442418 | Il fascismo eterno, Umberto Eco | 296 |

Table 2: The table shows the first 10 book sorted by meta-path B-R-U counts

five books with the highest similarity scores. While the values of Vannacci's and Arendt's book are less interpretable, for J'accuse of Francesca Albanese a clear thematic affinity emerges with other volumes addressing various aspects of the Israeli-Palestinian conflict.

Table 9 presents the distribution of topics by book. For Vannacci's book, the predominant topics are T0 (Society and Identity) and T3 (Vannacci). It is also noted that, for all books, there is always a significant component linked to the reading experience, book quality, and purchasing experience.

Figures 4 and 5 present, via boxplots, the overall distribution of primary and complex emotions by topic. In both cases, an outlier appears in topic T0 – Society and Identity. This may be explained by certain limitations in the topic modeling process.

From the heatmap analysis of primary emotions (Figure 6), it is observed that the most commonly expressed emotions across topics are Anticipation, Joy, and Trust (all 'above'). This aligns with the fact that most reviews are positive, given that the analyzed

books are bestsellers. However, the distribution of Sadness and Surprise above is also noteworthy, suggesting the presence of emotional dissonance across topics.

The cluster map of primary emotions (Figure 8) highlights, for the Vannacci topic, that the emotion Disgust is uniformly distributed above and below the z-score threshold but remains higher than in other topics. This may indicate strong emotional polarization around this topic.

For the purchase experience topic, the data on Anticipation above may reflect the fact that Amazon users, especially bestsellers, expect timely delivery and product quality.

In the Israeli-Palestinian conflict topic, Disgust (above) is prominent, while in the history and memory of fascism in Italy topic, Anger, Fear, and Sadness (above) dominate. Cognitively, these emotions may represent a strong aversion toward these themes; in the case of the Israeli-Palestinian conflict, considering the May 7th data collection date, this may reflect an early alignment of political and media discourse on

| ASIN | Useful | Topic Description | Meaningful Emotions | Complex Emotions | Token Len |
|------------|--------|---|--|------------------|-----------|
| B0CF4BJN3D | 824 | Vannacci | [anger_under, trust_above, fear_under] | [] | 786 |
| B0CF4BJN3D | 593 | Giudizio morale e memoria | [trust_above, disgust_under, anticipation_above] | [hope] | 673 |
| B0CF4BJN3D | 461 | Società e identità | [anticipation_above] | [] | 136 |
| B0CF4BJN3D | 417 | Storia e memoria del fascismo in Italia | [trust_above, disgust_under] | [] | 1052 |
| B0CF4BJN3D | 413 | Vannacci | [fear_under, anticipation_above] | [] | 746 |
| B0CF4BJN3D | 318 | Società e identità | [disgust_under] | [] | 417 |
| B0CF4BJN3D | 242 | Conflitto Israelo-Palestinese | [trust_above] | [] | 1366 |
| B0CF4BJN3D | 229 | Vannacci | [trust_above] | [] | 131 |
| B0CF4BJN3D | 222 | Società e identità | [] | [] | 131 |
| B0CF4BJN3D | 220 | Vannacci | [trust_above] | [] | 860 |

Table 3: *Analisi per topic e contenuto emozionale su nodi del libro B0CF4BJN3D*

La tabella riassume le metriche principali su 10 nodi selezionati, inclusi topic trattati, emozioni estratte e dimensione del contenuto.

| | Mean | Max | Min | Std | Q1 | Q2 | Q3 | P90 | P95 |
|------------|-------|---------|------|--------|-------|-------|-------|--------|--------|
| P99 | | | | | | | | | |
| Useful Num | 4.69 | 824.00 | 0.00 | 25.95 | 0.00 | 0.00 | 2.00 | 8.00 | 16.00 |
| 83.18 | | | | | | | | | |
| Token Len | 50.54 | 2139.00 | 0.00 | 103.11 | 11.00 | 26.00 | 51.00 | 102.00 | 164.00 |
| 461.53 | | | | | | | | | |

Table 4: *Statistiche descrittive per Useful Num e Token Len.*

the conflict in Italy in recent weeks.

Regarding complex emotions (Figure 9), the most represented are Friendliness, Hope, and Optimism, originating from the positive interaction between Joy, Trust, and Anticipation, which is consistent with the bestseller book context.

Within the Vannacci topic, all 28 emotion combinations emerge, albeit with varying frequency, indicating a highly polarizing impact of the reviews.

In the Israeli-Palestinian conflict topic, there is a greater prevalence of Remorse, Shame, and cynicism, while in the reading experience topic, Contempt and Hate emerge, potentially reflecting negative judgments on the quality and style of the book.

Additionally, for the Vannacci topic, higher levels of Unbelief and Despair are observed, likely associated with low-rated reviews conveying a general sense of shock and bewilderment. The topic also includes Aggressiveness, Ambivalence, and Morbidity, which are more complex to interpret, as they may express either strongly negative judgments or mixed and nuanced reactions to controversial or divisive content.

To further validate the analyses, a Spearman correlation test was conducted between the number of reactions to reviews and review length, this test was chosen due to its robustness on outliers and because it doesn't assume a linear relationship, but a monotonic relationship.

The results show a statistically significant positive correlation between the two variables, allowing rejection of the null hypothesis H_0 since $p < 0.05$.

All the test results are shown as the output object from the dedicated methods from the library `scipy.stats`:

Spearman Correlation Test Hypotheses:

- H_0 : There is no monotonic correlation between number of reactions and review length.
- H_1 : There is a monotonic correlation (positive or negative) between number of reactions and review length.

`SpearmanResult(statistic=0.44915107029970447, pvalue=1.5207482880738465e-185).`

Subsequently, attention focused on the role of primary and complex emotions in driving interaction with reviews. Given the asymmetry of the variables, normality was tested using the Shapiro-Wilk test for two groups: reviews with no expressed complex emotions and reviews with at least one expressed complex emotion:

Shapiro-Wilk Test Hypotheses

- H_0 : The variable is normally distributed.
- H_1 : The variable is not normally distributed.

Values for the Complex Emotion-Useful Reaction groups

```
ShapiroResult(statistic=0.15093621639863475,
               pvalue=6.877567317608895e-77)
ShapiroResult(statistic=0.12669054329346574,
               pvalue=1.5243093212406935e-55)
```

Given the rejection of H_0 (non-normality), the Mann-Whitney U test (right-tailed) was applied to assess whether the distribution of the first group (no complex emotions) tends to have higher values than the second group (with complex emotions):

Mann-Whitney U Test Hypotheses (right-tailed):

- H_0 : The distributions of the two populations are the same or the "no complex emotions" group tends to have equal or lower values than the "complex emotions" group.
- H_1 : The "no complex emotions" group tends to have higher values than the "complex emotions" group.

```
MannwhitneyuResult(statistic=np.float64(1462916.0),
                    pvalue=np.float64(0.0014242348351336017))
```

The result allows rejection of the null hypothesis, indicating a significant difference between the two groups, with the distribution of reviews without complex emotions likely to exhibit higher values. However, for the groups comparing reactions to reviews with or without primary emotions, the null hypothesis H could not be rejected:

Mann-Whitney U Test Hypotheses (right-tailed):

- H_0 : The distributions of the two populations are the same or the "no primary emotions" group tends to have equal or lower values than the "primary emotions" group.
- H_1 : The "no primary emotions" group tends to have higher values than the "primary emotions" group.

```
MannwhitneyuResult(statistic=np.float64(1595972.5),
                    pvalue=np.float64(0.5125328167509862))
```

This result requires further investigation, as no interpretation has yet been found for the difference between significant complex emotions and primary emotions about the number of reactions. However, in general, this evidence may suggest that excessively emotional reviews provide less useful information for purchasing decisions, leading users to rate emotionally neutral reviews as more useful. Finally, contingency tables for topics and times of emotions expressed in reviews (primary and complex) were created and subjected to chi-square tests for independence. The results are as follows:

Chi-square Test Hypotheses:

- H_0 : Topics and emotional expression are independent.
- H_1 : Topics and emotional expression are not independent.

Primary emotions:

Chi-square statistic: 214.6378609174688, P-value: 4.735719708192554e-09, dof: 108

Complex emotions:

Chi-square statistic: 358.212539244621, P-value: 1.2096627862377577e-05, dof: 252

In both cases, the null hypothesis can be rejected, confirming the existence of a statistically significant relationship between topic and complex emotional expression. A note on the contingency table, emotions that were not expressed in any review, are not inserted as columns in the final table.

Conclusions Based on Empirical Analysis

Hypothesis H1A is supported by the network analysis, particularly by the comparison of activity within the book-review-user subnetworks, as measured through the number of meta-paths $B \rightarrow R \rightarrow U$.

In the case of "Il mondo al contrario" by Roberto Vannacci, clearly positioned within the far-right wing of Italian politics, especially in light of the author's subsequent political career, the network appears substantially more active compared to other titles. However, it is important to note that, within the bestseller list considered, few books have an explicit

political connotation: most belong to the genres of essays, philosophy, or journalism, addressing current events with more analytical than ideological tones.

A possible interpretation of the phenomenon surrounding Vannacci's book, beyond the quantitative data, concerns the strong emotional impact generated by the book, an impact amplified by the media coverage it received.

On one hand, many users seem compelled to express an opinion on the book, given its centrality in public debate; on the other hand, supporters of the book appear to feel the need to defend it against external criticism, viewing it as an identity symbol and an element of ideological cohesion. This dual mechanism contributes to the formation of a particularly dense network of interactions, characterized by a high level of polarization and participation.

Regarding hypothesis H1B, the analysis did not provide conclusive results. The exploration of topic distribution across the top 10 books revealed notable differences, but it was not possible to quantify their concentration or dispersion. The application of the PathSim method on the meta-path $B \rightarrow R \rightarrow T \rightarrow R \rightarrow B$ produced ambivalent results: for Vannacci and Arendt, no clear thematic similarities emerged, while for J'accuse by Francesca Albanese a higher degree of macro-thematic affinity with other books was observed. However, the limited size of the analyzed sample, which will be further explored in subsequent sections, prevents drawing robust conclusions.

Regarding hypothesis H2A, a positive and statistically significant correlation was found between the number of reactions received by a review and its length (measured in word count). This finding aligns with other studies on Amazon user behavior: longer reviews tend to be perceived as more informative and thus generate greater interaction. In the context of book reviews, where the primary objective is to assist other users in making purchase decisions, length may reflect a descriptive rather than an emotional or evaluative intent.

Hypothesis H2B produced different results depending on the type of emotions analyzed. For primary emotions, no statistically significant difference was observed in the average number of reactions between reviews expressing such emotions and those that did not.

Conversely, for complex emotions, the Mann-Whitney test indicated a significant difference: reviews lacking complex emotions tend to receive slightly more reactions. This result may be related to what was observed under hypothesis H2A: reviews with complex emotions—potentially more concise and more judgment-oriented than descriptive—might be less informative and thus less appre-

ciated by users.

Finally, hypothesis H3 is confirmed by the chi-square tests conducted on contingency tables linking identified topics with the emotions (both primary and complex) expressed in the reviews. The results show a non-random distribution of emotions across topics, suggesting that certain subjects tend to elicit specific emotional responses. This allows us to assert that it is possible to infer the emotional content of reviews based on the themes addressed in the books.

Critical Analysis of the Adopted Strategy

Strengths One of the main strengths of this study lies in its focus on a platform that differs from traditional social networks such as X, Instagram, or YouTube. Although Amazon is not generally classified as social media, it exhibits numerous social network dynamics, especially in the interactions between users, content, and products. This perspective allows for the exploration of ideological and affective dynamics within a consumption context, opening avenues often overlooked by more conventional research.

Another distinctive element of the study is the use of a heterogeneous and complex network composed of multiple types of nodes and relations, enabling a more realistic representation of the interactions occurring on the platform. Despite not explicitly including user nodes, it was still possible to analyze three fundamental relational dimensions: the book–review link, the review–social interaction (reactions) link, and the review–topic link.

A particularly innovative aspect of the project is the approach to textual analysis, which focuses more on emotional analysis rather than the more common sentiment analysis. The use of the EmoAtlas library enabled the detection of a broad range of emotions, going beyond the simple positive/negative dichotomy.

This approach allows for the capture of complex nuances and lends itself well to integration with theoretical models from social psychology and the sociology of emotions. Unlike polarity analysis, which is often overly simplistic and reductive, emotional analysis provides richer insights that are more useful for understanding the cognitive and relational contexts in which opinions emerge.

Finally, the project stands out for its critical and interdisciplinary perspective, combining tools and concepts from cognitive and social sciences. Special attention was given to the book as not only a consumer product but also a cultural and ideological object through which affiliations, oppositions, and identity positions are expressed. In this sense, the analysis of bestsellers in Amazon's Politics category

highlighted the central role of the book in constructing and polarizing collective imaginaries, as well as the platform's algorithmic and commercial logic.

Weaknesses The main limitation of the research concerns the dataset size. Due to Amazon's restrictions during web scraping (rate limits, anti-bot systems, dynamic page structures), it was only possible to analyze a limited number of books and a maximum of 100 reviews for each.

This constitutes a significant methodological constraint, especially for books such as *"Il mondo al contrario"* by Vannacci, which have hundreds or thousands of reviews. Thus, the analysis is based on a potentially unrepresentative sample, introducing possible biases in evaluating emotional and social dynamics.

Another weakness concerns the quality of the extracted text: many reviews contain spelling mistakes, incorrectly written proper names or informal expressions that are difficult to process with standard NLP tools. Although a spellchecker was used to attempt a correction, the process proved too costly in terms of time and resources. The topic modeling was also affected by these issues, leading to the emergence of hard-to-interpret or overly generic topics.

Finally, the absence of user nodes in the network limits the possibility of conducting behavioral or longitudinal analyses on user profiles and consumption habits. Although this choice was intentional due to its complexity, it remains a point of reflection for future developments.

Potential Improvements Among the possible improvements, the first involves expanding the dataset, both in terms of the number of books and the number of reviews per title. Another promising direction would be the explicit inclusion of user nodes in the network to investigate:

- how much exposure users have to similar content (in thematic or ideological terms),
- how active users are in producing reviews,
- and whether there are common consumption patterns among users with similar ideological orientations.

This would also allow for the study of individual polarization dynamics, for instance, whether a user consistently reviews books of the same orientation, varies over time, or is influenced by media trends.

Further developments could focus on analyzing Amazon's recommendation system. It is plausible to hypothesize that the algorithm suggests content similar to what has already been purchased or viewed, potentially generating echo chambers or filter bubbles similar to those observed on YouTube (as highlighted in the referenced South Korean study) [5].

The visibility architecture of reviews may also play a role in creating cognitive biases, favoring highly reactive or positive content.

Finally, it would be valuable to examine the role of self-publishing through Amazon, as in the case of Vannacci's book. If the book's success was amplified by the platform's internal logic, then Amazon itself may bear indirect responsibility for the visibility and dissemination of divisive or extremist content.

Final Conclusions

The initial goal of this study was to explore Amazon through the lens of social networks and to verify the possibility of adopting a sociological approach to platform activities, focusing particularly on emotional aspects rather than mere ideological polarization.

The results suggest that activities conducted and mediated on Amazon indeed reflect social dynamics, from which meaningful interpretations can be drawn, especially in the context of politically oriented book consumption—a field particularly revealing of opinions, reflections, and ideological tensions.

The case of *"Il mondo al contrario"* by Vannacci clearly demonstrates how a social phenomenon can take shape within the platform. The intensity and variety of emotional reactions expressed in the reviews suggest a strong ingroup-outgroup division, visible precisely in the distribution and nature of the expressed emotions. This confirms the importance of the book not only as a cultural object but also as a symbolic catalyst of collective identity—almost functioning as a fetish around which a movement may coalesce.

It is also important to consider that the book was self-published, leveraging the infrastructures and visibility mechanisms offered by Amazon. In this sense, the book's success may be interpreted, at least in part, as an outcome of the platform's technical and commercial architecture.

The adopted approach, which combines emotional analysis and topic modeling, has proven effective in generating interesting interpretations that may also be relevant for understanding certain aspects of the current political discourse in Italy. Although the results cannot be generalized or empirically confirmed exhaustively, the study offers valuable insights, particularly when integrated with future interdisciplinary research.

Ethical Considerations

In the present study, no identifying information related to the users who authored the analyzed reviews was intentionally collected or stored in the dataset.

Due to the absence of publicly available APIs provided by Amazon, it was necessary to employ web crawling techniques directly from the website. However, this approach may raise both legal and ethical concerns.

While it may appear paradoxical that a platform such as Amazon, with its substantial IT infrastructure, does not provide data or public APIs for academic research purposes, data collection from this source remains subject to privacy constraints. In particular, the inclusion of user-related data could allow for the extraction of additional personal information, thus raising significant ethical issues.

For these reasons, the present work deliberately avoided collecting or utilizing any user-identifiable data. Nevertheless, future studies should carefully assess, at minimum, the anonymization of such data and undertake a more thorough evaluation of the associated ethical and legal considerations.

References

- [1] David Nieborg and Thomas Poell. "The platformization of cultural production: Theorizing the contingent cultural commodity". In: *New Media Society* 20 (Nov. 2018), pp. 4275–4292. DOI: 10.1177/1461444818769694.
- [2] David Meyer and Deana Rohlinger. "Big Books and Social Movements: A Myth of Ideas and Social Change". In: *Social Problems* 59 (Feb. 2012), pp. 136–153. DOI: 10.1525/sp.2012.59.1.136.
- [3] LISA NAKAMURA. "'Words with Friends': Socially Networked Reading on 'Goodreads'". In: *PMLA* 128.1 (2013), pp. 238–243. ISSN: 00308129. URL: <http://www.jstor.org/stable/23489284> (visited on 06/11/2025).
- [4] Mark Davis. "Five processes in the platformization of cultural production: Amazon and its publishing ecosystem". In: *Australian Humanities Review* 66 (2020), pp. 83–103.
- [5] Han Park and Sejung Park. "The filter bubble generated by artificial intelligence algorithms and the network dynamics of collective polarization on YouTube: the case of South Korea". In: *Asian Journal of Communication* 34 (Feb. 2024), pp. 1–18. DOI: 10.1080/01292986.2024.2315584.
- [6] Graf-Vlachy L., Goyal T., and Ouardi Y et al. "Reviews Left and Right: The Link Between Reviewers' Political Ideology and Online Review Language". In: *Business Information Systems Engineering* 63 (2021), pp. 403–417. DOI: <https://doi.org/10.1007/s12599-020-00652-1>.
- [7] Leslie Howsam. "Introducing Book Networks and Cultural Capital: Space, Society and the Nation". In: *Mémoires du livre / Studies in Book Culture* 2.1 (2010).
- [8] Stefan Dimitrov et al. "Goodreads Versus Amazon: The Effect of Decoupling Book Reviewing And Book Selling". In: *Proceedings of the International AAAI Conference on Web and Social Media* 9 (Aug. 2021), pp. 602–605. DOI: 10.1609/icwsm.v9i1.14662.
- [9] Samuel Bestvater and Burt Monroe. "Sentiment is Not Stance: Target-Aware Opinion Classification for Political Text Analysis". In: *Political Analysis* 31 (Apr. 2022), pp. 1–22. DOI: 10.1017/pan.2022.10.
- [10] Kai Wang, Xiaojuan Liu, and Yutong Han. "Exploring Goodreads reviews for book impact assessment". In: *Journal of Informetrics* 13 (Aug. 2019), pp. 874–886. DOI: 10.1016/j.joi.2019.07.003.
- [11] Suman Maity, Abhishek Panigrahi, and Animesh Mukherjee. "Analyzing Social Book Reading Behavior on Goodreads and how it predicts Amazon Best Sellers". In: (Sept. 2018). DOI: 10.48550/arXiv.1809.07354.
- [12] Mariana O. Silva et al. "Brazilian Reading Preferences in Goodreads: Cross-state and Cross-region Analyses". In: *iSys - Brazilian Journal of Information Systems* 15.1 (Dec. 2022), 25:1–25:20. DOI: 10.5753/isys.2022.2411. URL: <https://journals-sol.sbc.org.br/index.php/isys/article/view/2411>.
- [13] Franco Chiavetta, Giosuè Lo Bosco, and Giovanni Pilato. "A Lexicon-based Approach for Sentiment Classification of Amazon Books Reviews in Italian Language". In: (Jan. 2016), pp. 159–170. DOI: 10.5220/0005915301590170.
- [14] Ezgi Akar. "Exploring the Impact of Social Network Structures on Toxicity in Online Mental Health Communities". In: *Computers in Human Behavior* 165 (Apr. 2025), p. 108542. DOI: 10.1016/j.chb.2024.108542.
- [15] Yizhou Sun et al. "Heterogeneous information networks: the past, the present, and the future". In: *Proceedings of the VLDB Endowment* 15 (Aug. 2022), pp. 3807–3811. DOI: 10.14778/3554821.3554901.
- [16] Alfonso Semeraro et al. "EmoAtlas: An emotional network analyzer of texts that merges psychological lexicons, artificial intelligence, and network science". In: *Behavior Research Methods* 57 (Jan. 2025). DOI: 10.3758/s13428-024-02553-7.

- [17] Alfonso Semeraro, Salvatore Vilella, and Giancarlo Ruffo. “PyPlutchik: Visualising and comparing emotion-annotated corpora”. In: *PLOS ONE* 16 (Sept. 2021), e0256503. DOI: 10.1371/journal.pone.0256503.
- [18] Yizhou Sun and Jiawei Han. “Meta-Path-Based Search and Mining in Heterogeneous Information Networks”. In: *Tsinghua Science and Technology* 18 (Aug. 2013). DOI: 10.1109/TST.2013.6574671.

| | ASIN | Count | Ratio on Total |
|---|------------|-------|----------------|
| 1 | B0CF4BJN3D | 108 | 0.282723 |
| 2 | 8807897458 | 31 | 0.081152 |
| 3 | 8830104957 | 21 | 0.054974 |
| 4 | 8804751169 | 16 | 0.041885 |
| 5 | 8893881284 | 14 | 0.036649 |

Table 5: Percentage of the number reviews for each book exceeding the 90th percentile in length

| Topic | Value |
|--------|----------|
| TOPIC3 | 0.256649 |
| TOPIC0 | 0.169414 |
| TOPIC4 | 0.134406 |
| TOPIC5 | 0.068595 |
| TOPIC6 | 0.068140 |
| TOPIC9 | 0.067856 |
| TOPIC2 | 0.066606 |
| TOPIC7 | 0.066095 |
| TOPIC8 | 0.058081 |
| TOPIC1 | 0.044158 |

Table 6: Percentage distribution of topics based on the number of T-R-U meta-paths

| Topic | Value |
|--------|----------|
| TOPIC0 | 0.034635 |
| TOPIC4 | 0.021630 |
| TOPIC3 | 0.019905 |
| TOPIC6 | 0.017947 |
| TOPIC2 | 0.017528 |
| TOPIC7 | 0.016828 |
| TOPIC8 | 0.013658 |
| TOPIC9 | 0.013146 |
| TOPIC1 | 0.009929 |
| TOPIC5 | 0.009510 |

Table 7: Degree Centrality for the Topic Node Type

| ASIN | Valore |
|------------|--------|
| B0CF4BJN3D | 1.0000 |
| 8893881284 | 0.6945 |
| 8804665289 | 0.6020 |
| 8896985625 | 0.4680 |
| 8804751169 | 0.4167 |
| 8807897458 | 0.3956 |

Table 8: PathSim - B0CF4BJN3D

| | T0 | T1 | T2 | T3 | T4 | T5 | T6 | T7 | T8 | T9 |
|------------|------|------|------|------|------|------|------|------|------|------|
| B0CF4BJN3D | 0.24 | 0.08 | 0.04 | 0.29 | 0.08 | 0.05 | 0.1 | 0.05 | 0.04 | 0.04 |
| 8807897458 | 0.13 | 0.06 | 0.05 | 0.03 | 0.4 | 0.02 | 0.08 | 0.05 | 0.1 | 0.07 |
| 8804751169 | 0.16 | 0.06 | 0.08 | 0.07 | 0.09 | 0.04 | 0.05 | 0.09 | 0.05 | 0.3 |
| 8893881284 | 0.22 | 0.06 | 0.04 | 0.1 | 0.04 | 0.03 | 0.27 | 0.05 | 0.14 | 0.05 |
| 8804776455 | 0.12 | 0.15 | 0.03 | 0.16 | 0.12 | 0.07 | 0.07 | 0.15 | 0.12 | 0.03 |
| B0CB237DYY | 0.13 | 0 | 0.19 | 0.06 | 0.15 | 0.19 | 0.07 | 0.04 | 0.13 | 0.06 |
| B00DJ2KYQQ | 0.17 | 0.05 | 0.03 | 0.11 | 0.08 | 0.03 | 0.35 | 0.04 | 0.1 | 0.04 |
| 8807900521 | 0.16 | 0.09 | 0.08 | 0.05 | 0.06 | 0.06 | 0.23 | 0.07 | 0.14 | 0.05 |
| B0CLML6FM8 | 0.11 | 0.02 | 0.32 | 0.12 | 0.15 | 0.09 | 0.09 | 0.04 | 0.05 | 0.01 |
| 8893442418 | 0.11 | 0.06 | 0.05 | 0.08 | 0.09 | 0.01 | 0.02 | 0.48 | 0.05 | 0.04 |

Table 9: Distribuzione dei topic per i primi 10 libri classificati secondo il suo degree attraverso il conteggio delle meta-path

| ASIN | Valore |
|------------|--------|
| 8807897458 | 1.0000 |
| 8858043057 | 0.6705 |
| B076H8MFMJ | 0.6211 |
| 8804665289 | 0.6178 |
| 8893780712 | 0.5828 |
| 8804751169 | 0.5636 |

Table 10: PathSim 8807897458

| ASIN | Valore |
|------------|--------|
| B0CB237DYY | 1.0000 |
| 885815553X | 0.8940 |
| 8836000185 | 0.8763 |
| B0CTS4LK18 | 0.8213 |
| 8858051793 | 0.8057 |
| 1798606364 | 0.7386 |

Table 11: Pathsim B0CB237DYY

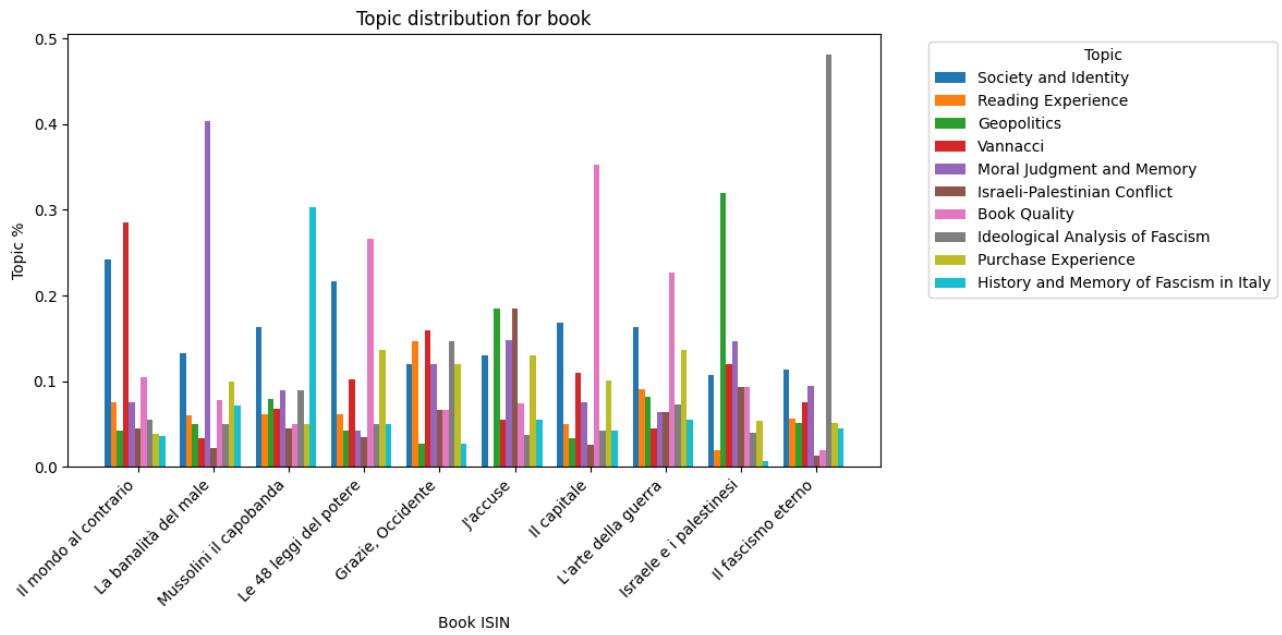


Figure 3: Topic Percentage for first 10 Books

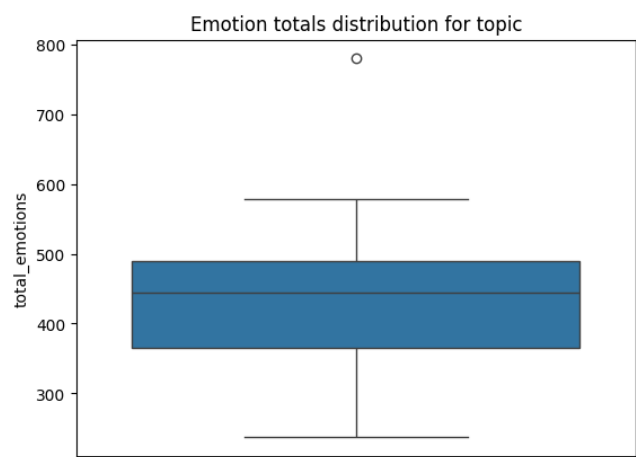


Figure 4: Boxplot for primary emotions totals for topic

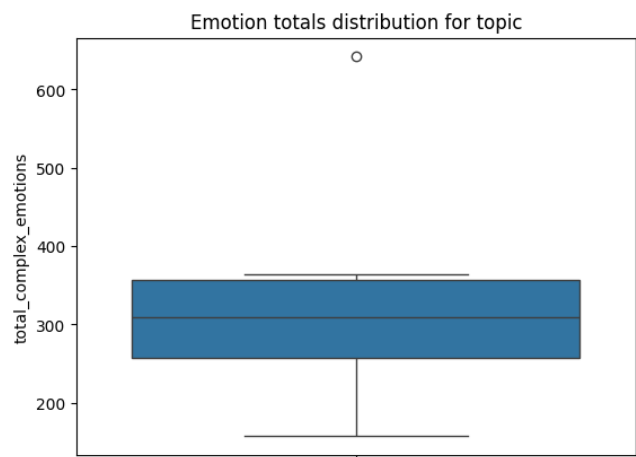


Figure 5: Boxplot for complex emotions totals for topic



Figure 6: Heatmap of primary emotions for topic

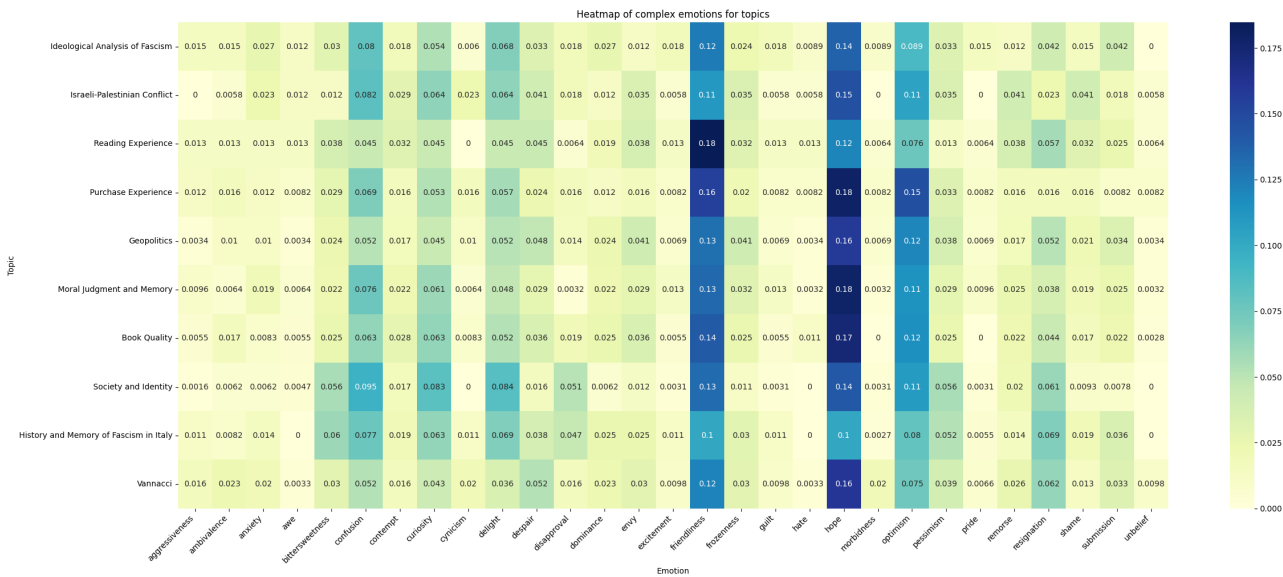


Figure 7: Heatmap of complex emotions for topic

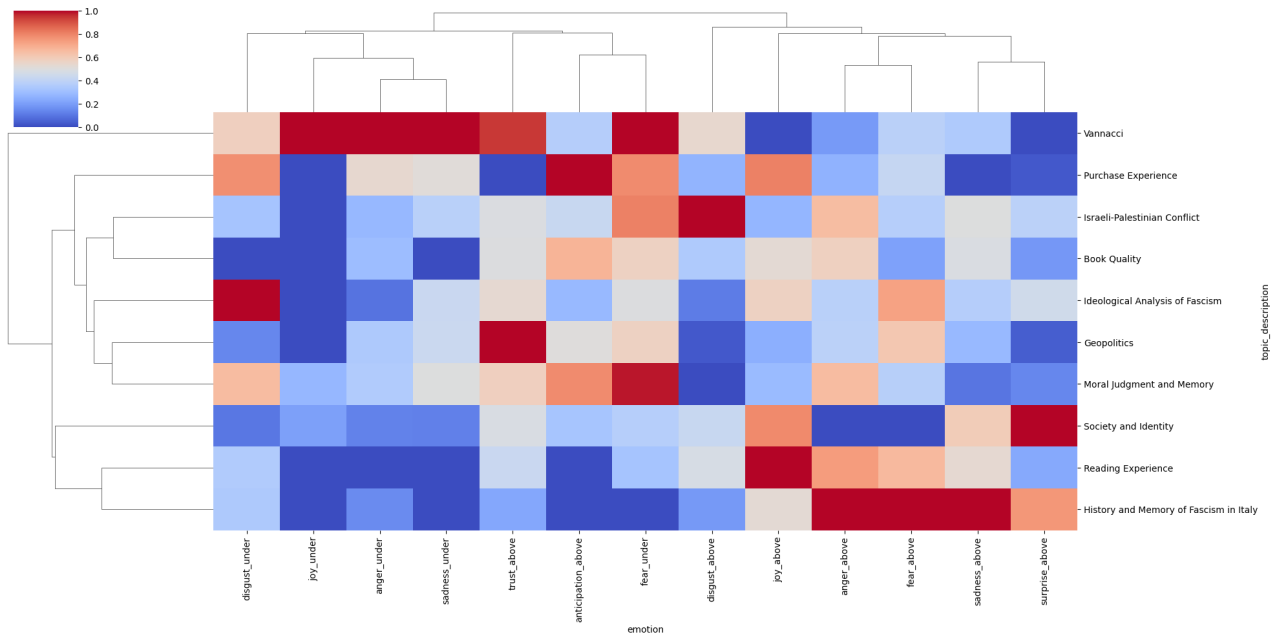


Figure 8: Clustermap of primary emotions for topic

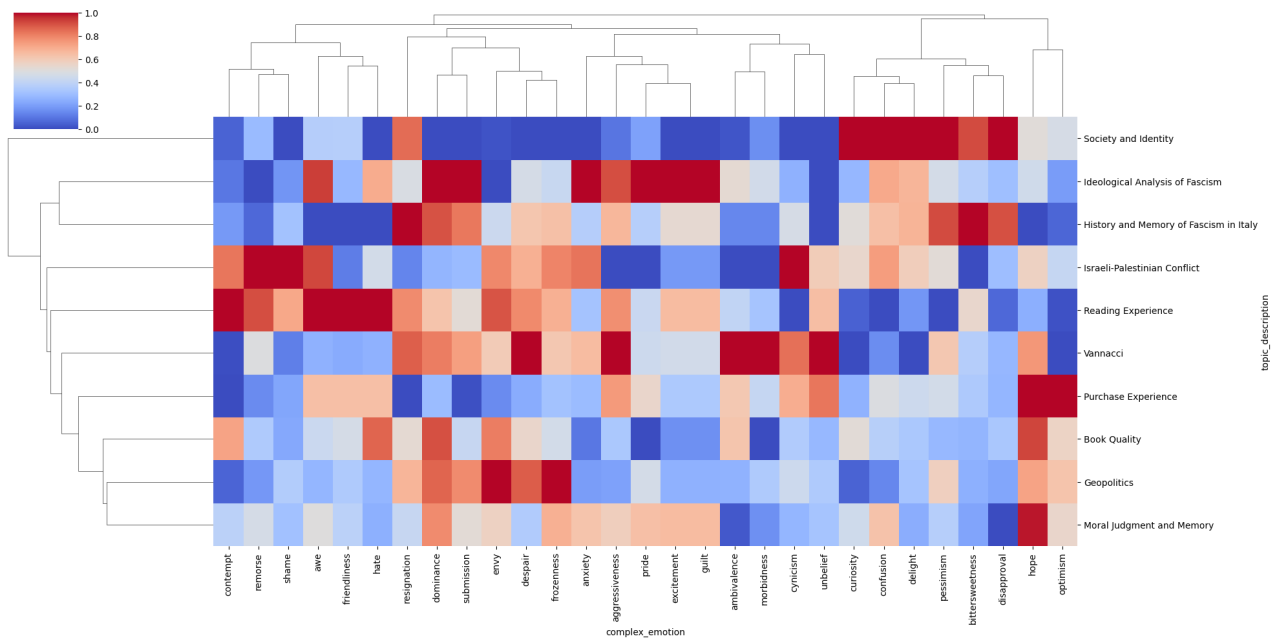


Figure 9: Clustermap of complex emotions for topic