Exploring Brazilian Cultural Identity
Through Reading Preferences

Mariana O. Silva[1]**, Clarisse Scofield**[1]**, Gabriel P. Oliveira**[1]**,**
Danilo B. Seufitelli[1]**, Mirella M. Moro**[1]

[1]Department of Computer Science
Universidade Federal de Minas Gerais (UFMG) – Belo Horizonte, MG – Brazil

`{mariana.santos,clarissescofield,gabrielpoliveira}@dcc.ufmg.br`

`{daniloboechat,mirella}@dcc.ufmg.br`

Abstract. *In Brazil, each region has its own cultural identity regarding accent, gastronomy, traditions, all of which may reflect in its literature. Specially, we believe that country's background and contextual features are directly related to what people read. Hence, we perform a cross-state comparison analysis based on Brazilian reading preferences through a multipartite network model. Also, we explore the effects of socioeconomic and demographic factors on favorite books and literary genres. Such cross-state analyses highlight how the country is culturally wealthy, where each region has its own distinctive personality. Our findings offer great opportunities for the Brazilian book industry by enhancing current knowledge on social indicators related to reading preferences.*

1. Introduction

Brazil is the largest country in Latin America and the fifth-largest nation in the world. It is also known as one of the most diversified countries, and such a diverse background highlights its cultural wealth. Indeed, the Brazilian states have their own vocabulary, accent, gastronomy, fashion, music, and literature preferences.

Especially in literature, reading books is a cultural behavior shaped by social, economic, and local backgrounds. People choose what to read according to their ideas and intellectual realizations, plus their surrounding influence from a social perspective. From an economic perspective, cultural access and consumption are directly affected by socioeconomic status. Then, from a location perspective, each state or region within a country has characteristics that may interfere with reading preferences.

As a practical example, Figure 1 shows cross-state information regarding the top three most-read genres and books in Brazil (dataset in Section 3). *Romance*, *fantasy* and *classic* are the preferred genres in the country for regions northeast, southeast and south; whereas *mystery* and *adventure* are preferred in north and central-west regions. For the two main genres, the most read books are the same. However, the south elects "1984" as the most read of the *classic* genre.

Studying if and how cultural and socioeconomic factors influence reading preferences is a valuable tool for book-selling and understanding current cultural trends. Moreover, such insights may improve recommendation systems of books, sales projections, and customer relationship management. With so many perspectives, the book market is
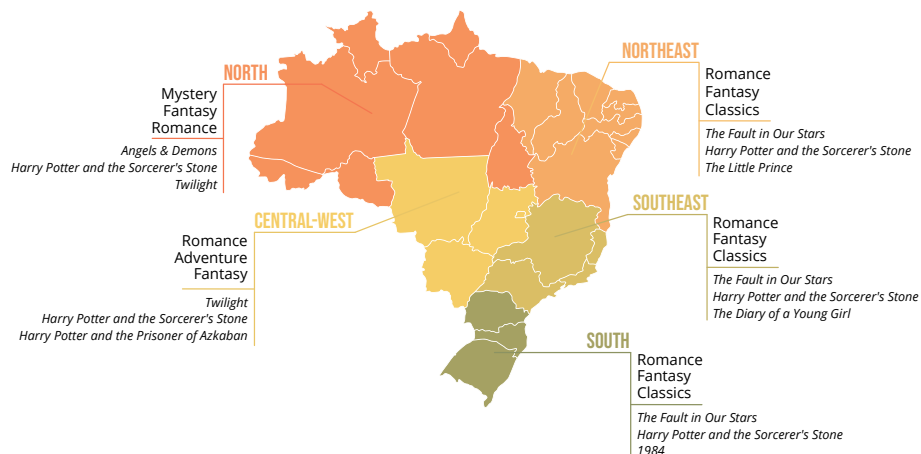
Figure 1. Brazilian Literary map with the top-3 most read genres by region.

bound to prosper. In this context, we analyze similarities and differences among the 26 Brazilian states and the Federal District based on reading preferences by using data from Goodreads.[1] We also use a bipartite configuration and a community detection model to explore cross-state relationships based on their book and genre preferences. After Related Work (Section 2), our contributions are summarized as follows.

- We build a dataset for genre and book preferences in Brazilian states (Section 3);
- We model a multipartite network and propose two metrics to assess the cross-state cultural identity (Section 4); and
- We analyze the networks from three different perspectives regarding reading preferences (Section 5): differences between Brazilian states and regions, demographic and socioeconomic relationships, and state identities.

2. Related Work

As the world becomes more connected with the Internet and other technological advances, people are frequently in touch with foreign cultures. Such a multicultural environment is studied by mainly focusing on how cultural differences affect how people interact with new technologies, such as social media [Nascimento et al. 2018], online music [Liu et al. 2018, Mondelli et al. 2018], and e-learning [Zhao et al. 2021]. Nonetheless, even a single country may have different cultural aspects within it. For instance, although Brazil has a strong cultural identity, it contains several cultural differences between its states and regions due to its continental dimensions. Digiampietri et al. assessed this regional factor by analyzing the relationships between Brazilian scientists using a cross-state social network [Digiampietri et al. 2014]. In addition, Borges de Souza et al. identify a significant correlation between cultural differences and how Brazilians from the five geographical regions interact with Web interfaces [Borges de Souza et al. 2015].

Reading preferences are also affected by cultural identities because what people choose to read is highly connected to their language, background, and personal beliefs. In such a context, Goodreads is one of the leading social networks for book readers. Current research on such a platform includes using book reading behavior to predict best-sellers [Maity et al. 2017], measuring book impact by user reviews

---

[1]Goodreads: https://www.goodreads.com/

[Wang et al. 2019], and using such reviews to extract meaningful characters and their relationships [Shahsavari et al. 2020]. Data extracted from Goodreads is also used to understand cultural differences between countries. Sabri et al. use a bipartite network model to explore cross-country relationships, grouping them in communities based on reading preferences [Sabri et al. 2020]. The results indicate that such preferences are highly associated with factors such as geographical distance, language, and individualism.

This paper represents an important step towards understanding the relationship between cultural differences and reading habits in the face of all such related work. Brazil's continental nature is a perfect setup for different cultural practices across the 26 states and the Federal District. Hence, we wonder if such cultural diversity may also influence Brazilians' reading habits. To the best of our knowledge, we are the first to study book preferences and cultural identity in Brazil through a network-based model. Overall, our cross-state and cross-region analysis may improve the existing knowledge on Brazilian reading preferences, providing benefits for both the academy and the book industry.

3. Brazilian Readers' Data at Goodreads

In this exploratory study, we present an enhanced dataset comprising cultural, geographical, and socioeconomic information. We chose the Goodreads website as our principal data source due to the sheer volume of data available, as well as its organized and easy-access API. We collect Brazilian readers' data through the *goodreads*[2] library, which provides a Python interface for the Goodreads API.[3] Specifically, we collect members of two of the largest Brazilian reading groups: the "Clube de Leitores em Português" – "Club of Readers in Portuguese" (4,229 members), and "Goodreads Brasil" (3,222 members). For all members of both groups, we also collect data from their friends. Then, we filter only those containing Brazil as the location information from the final users' set. Finally, with the same library, we gather users' bookshelves to assess their reading preferences.

To investigate the Brazilian reading identity, we consider a medley of demographic and socioeconomic data from the Brazilian Institute of Geography and Statistics (IBGE):[4] territorial area, population estimate, demographic density, Human Development Index (HDI), Gross Domestic Product (GDP), and monthly household income per capita. All indicators refer to the year 2020, except the HDI that refers to 2017. The data was collected from February 23 to March 04, 2021. Our final dataset is publicly available and comprises 38,231 Brazilian Goodreads users, containing 75,093 distinct books belonging to 80 literary genres and six IBGE indicators regarding the 27 federative units of Brazil [Silva et al. 2021]. Next, we describe the data cleansing process (Section 3.1) and present a condensed exploratory analysis of our final dataset (Section 3.2).

3.1. Data Cleansing

In Goodreads, a book's genre is determined by crowd-sourcing users' bookshelves. For example, if multiple users shelve a book as *Science*, then that genre is assigned to the book by the algorithm. Such a system may include fuzzy and/or noisy tags in the genre list for a particular book. Hence, we apply a data cleansing process to consider only meaningful

---

[2]goodreads package: `https://github.com/sefakilic/goodreads/`
[3]Goodreads API: `https://www.goodreads.com/api`
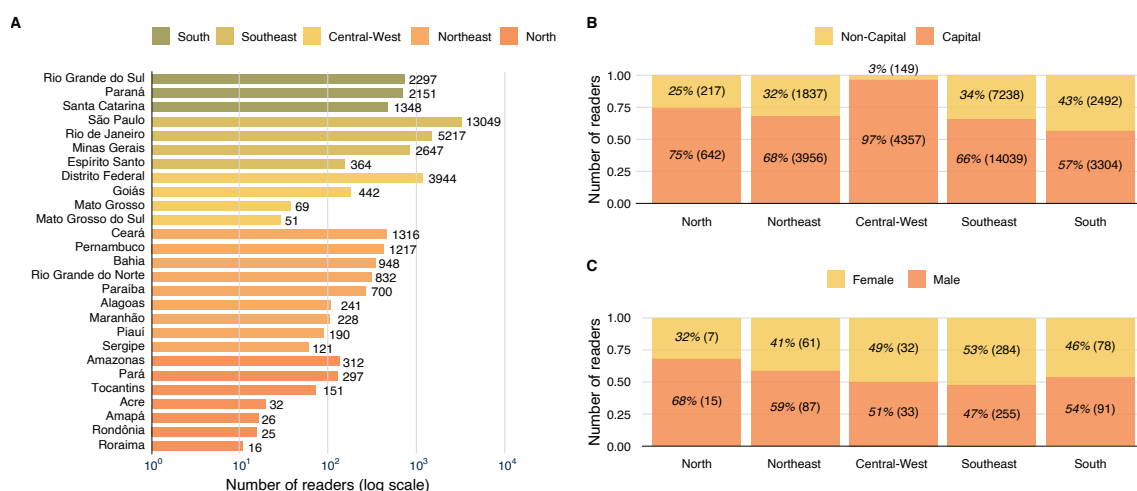[4]IBGE: `https://www.ibge.gov.br/en/cities-and-states`

Figure 2. Goodread readers-related statistics grouped by Brazilian region. (A) Number of readers ranked by federative unit within each region. (B) Percentage of readers in state capitals. (C) Percentage of readers by gender.

tags for each book. To do so, we filter the book's tag list to contain only genres related to *Fiction* and *Nonfiction* (see Figure 3). Next, we manually detect and remove inaccurate records from the user's dataset's location field to keep only Brazilian readers.

3.2. Exploratory Data Analysis

We now present a brief Exploratory Data Analysis (EDA) to investigate our processed dataset and summarize its main characteristics. Although Goodreads remains a valuable source of information on reading habits, we can determine through EDA the best means to manipulate and discover patterns in such data. Specifically, we start by analyzing the data related to the collected Brazilian readers, and next, we investigate the genres present in the dataset. In both analyses, we work with multivariate visualizations for describing data and mapping interactions between different variables. There are varied techniques for graphical EDA, including histograms, scatter and box plots. However, the most common approach used to describe categorical variables is bar plots.

Figure 2A ranks a grouped bar chart where each group represents a Brazilian region, and bars within a group (i.e., region) represent the total number of Brazilian readers. Figures 2B and 2C show the percentage of readers who live in state capitals and by readers' gender, respectively. Such analyses reveal an evident dominance of the Southeast region, with the state of São Paulo leading the ranking. Such a region compresses approximately 56% of the collected readers, followed by the South and Central-West regions corresponding to 15% each. Moreover, the vast majority of readers reside in capital cities. Such behavior is present in all regions, mainly in the Central-West, where the capitals are equivalent to 97% of the data. Finally, regarding gender information, there is a balanced outcome in general.

We now focus on the collected book-related information. With more than seven thousand distinct books, our dataset includes 80 different genres. Each genre is classified into *Fiction* or *Nonfiction* categories. Figure 3 breaks down all the collected books by category then genre. Within *Fiction*, most books fall into the *Novels* genre (125,907 books), making it the most popular genre. Indeed, in literature, fiction usually refers
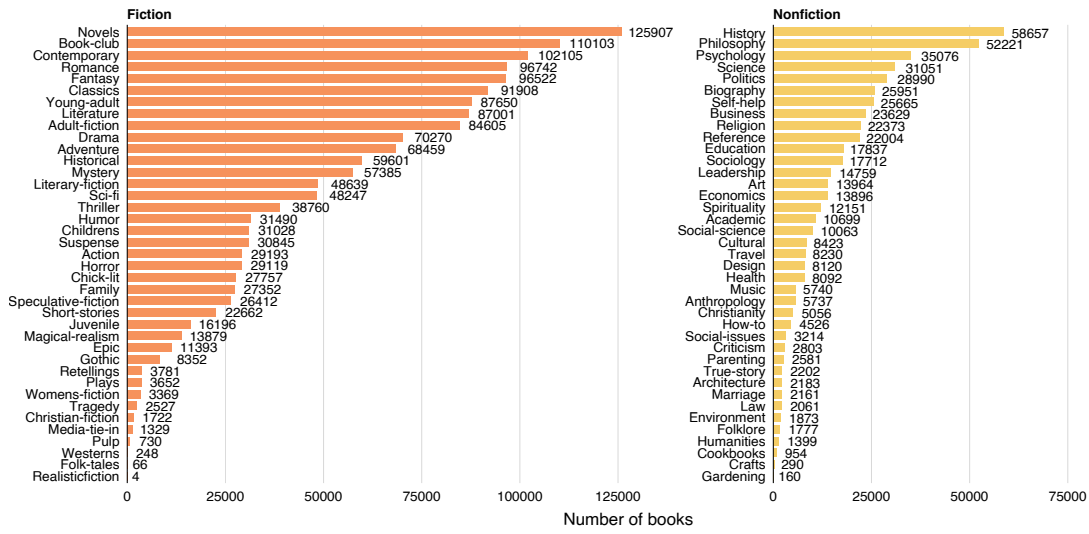
Figure 3. Breakdown of collected books into genre categories of (left) Fiction and (right) Nonfiction.

to a novel, one of the longest forms of literary prose. In the second place, *Book-club* has 110,103 books, which is more than twice the number of books in Sci-fi, the 15th place with 48,247 books. Previous research shows the growing popularity of the *Fiction* category [Yucesoy et al. 2018]. Likewise, there is explicit predominance of such a category in our dataset, comprising about 76% of the books. To conclude, among nonfiction books, approximately 45% of the collected books belong to *History*, *Philosophy*, *Psychology*, *Science*, *Politics*, and *Biography*.

4. Methodology

Having collected the data, our methodology follows by modeling two bipartite networks to assess the cross-state reading identity (Section 4.1). Then, we apply a graph-level statistic test to explore the significance of demographic and socioeconomic indicators on reading preferences (Section 4.2).

4.1. Multipartite Network Modeling

To investigate reading preferences between Brazilian states, we need to calculate cross-state cultural distances. To do so, we create a $u \times b$ matrix (denoted by UB) in which each row represents a Goodreads user, and each column represents a unique book. $UB[i][j]$ is set to one if book j is present in the bookshelf of user i. As previously mentioned, each book includes multiple tags that its readers assign. Therefore, after performing the cleansing steps specified in Section 3.1, we connect each book in the *user-book* bipartite network to the list of its Top-3 genres[5]. In addition, we connect each user to its Brazilian state, finally creating a *state-user-book-genre* multipartite network, as illustrated by a generic example in Figure 4.

The multipartite network connects a Brazilian state to a Goodreads user, and each user is connected to one or more books. In turn, each book is connected to a list of genres. Such a complete composition can be projected into six different bipartite networks,

---

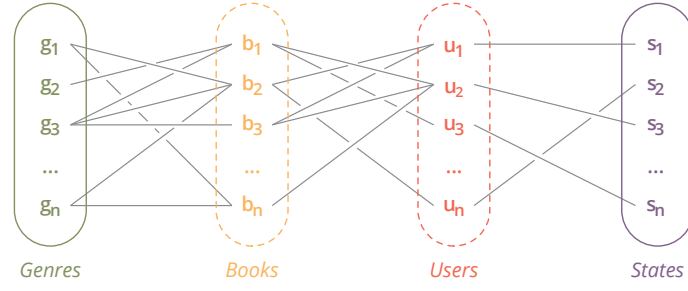[5]The length of the genres list ranges from one to three items.

Figure 4. The *state-user-book-genre* multipartite network diagram.

each generated by omitting one node type from the multipartite projection. Here, we are interested only in two bipartite networks: one connecting states and books and another connecting states and genres. We then create the $s \times b$ adjacency matrix (denoted by SB), with rows and columns representing the states and books, respectively. $SB[i][j]$ is set to one if book j is present in the users' bookshelves residing in the Brazilian state i. Likewise, we also create the *state-genre* matrix SG.

Finally, after creating the adjacency matrices, we proceed to calculating the cross-cultural distances: book and genre distances. To do so, we compute the cosine similarity for each adjacency matrix created. The cosine similarity between two states i and i' is given by $1 - \frac{u \cdot v}{\|u\|_2 \|v\|_2}$, where $u \cdot v$ is the dot product of vectors u and v, which represent two 1-D arrays $[i][:]$ and $[i'][:]$ of the respective adjacency matrix (SB or SG).

4.2. Cross-state distances

In order to examine the significance of the relationship between the Brazilian socioeconomic and demographic landscape and the reading preferences (cultural distances), we define the following cross-state distance measures:

- **Geographical Distance**: considers the latitude and longitude of each state capital and measures the geodesic distance between pairs of states.
- **Demographic and Socioeconomic Distances**: calculated between state pairs using the Canberra distance [Lance and Williams 1966, Lance and Williams 1967], a weighted version of the Manhattan distance. Such distances are separately calculated for each IBGE indicator listed in Section 3.

Having calculated all the cross-state distances, we apply the Quadratic Assignment Procedure (QAP) [Simpson 2001] test to examine whether such measures are significantly associated with the reading preferences. The QAP test is commonly used in social network analysis and achieves high statistical powers when performed on different types of complex networks [Krackardt 1987, Choi et al. 2006, Liu et al. 2018, Fredrickson and Chen 2019]. Moreover, it is a useful tool for analyzing dyadic datasets, i.e., datasets where pairs of entities are analyzed, which is our case. To perform the QAP test, we employ the *qaptest* function provided in R's SNA package [Butts 2020].

5. Results

In this section, our main goal is to provide meaningful insights into the Brazilian reading identity and associations between socioeconomic and demographic indicators. In particular, we are interested in the following research questions (RQs):
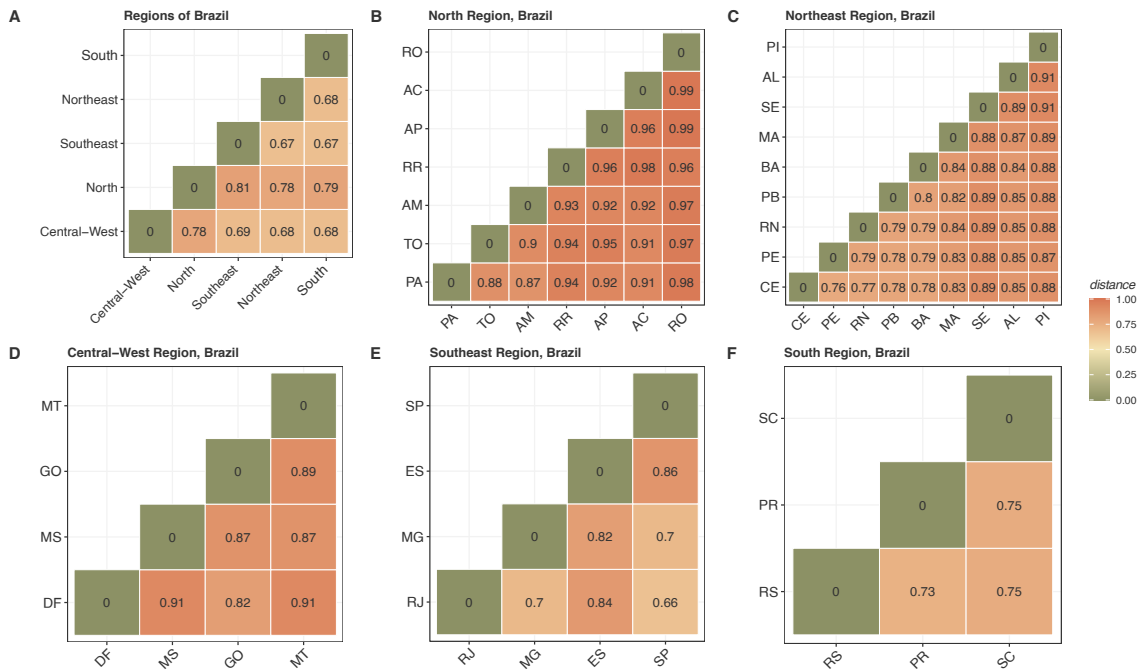
Figure 5. Cross-state Cultural Distance based on shared books.

RQ1. *Are there significant differences between Brazilian states/regions concerning reading preference?* (Section 5.1)

RQ2. *Is there a demographic or socioeconomic relationship in the reading preference in the Brazilian population?* (Section 5.2)

RQ3. *Is it possible to identify distinct cultural identities in Brazil based on the states' reading preferences?* (Section 5.3)

5.1. Cross-state Cultural Analysis

Initially, we perform a cross-state comparison to explore the similarities and differences based on people's favorite books and genres. To avoid bias results, we grouped the Brazilian states into regions. Figures 5 and 6 present the cross-cultural distance (cosine distance) of books and genre preferences, respectively. The cosine distance is bounded within $[0, 1]$, where a small value indicates a smaller distance between the respective pairs of states. In both figures, shades of green suggest cultural proximity, whereas orange tones indicate a cross-cultural distance. Overall, book-based distances are considerably larger than those based on genres, as the preference for books is naturally more specific. Besides, the available book search space is more extensive (75,093 books versus 80 genres).

From an inter-cluster perspective, the north region is more distant from the others regarding personal bookshelves (Figure 5A). Being the most extensive region in Brazil, covering approximately 45% of the entire national territory, its demographic indicators may be directly related to such a wide area. On the intra-cluster scale, the north (0.94 on average), central-west (0.88 on average), and northeast (0.84 on average) regions are the most diverse. The south and southeast regions have a milder level of diversity compared to an average cosine distance equal to 0.76 and 0.74, respectively. Even so, the state of Espírito Santo (ES) is notably more distant than other states within the southeast region.
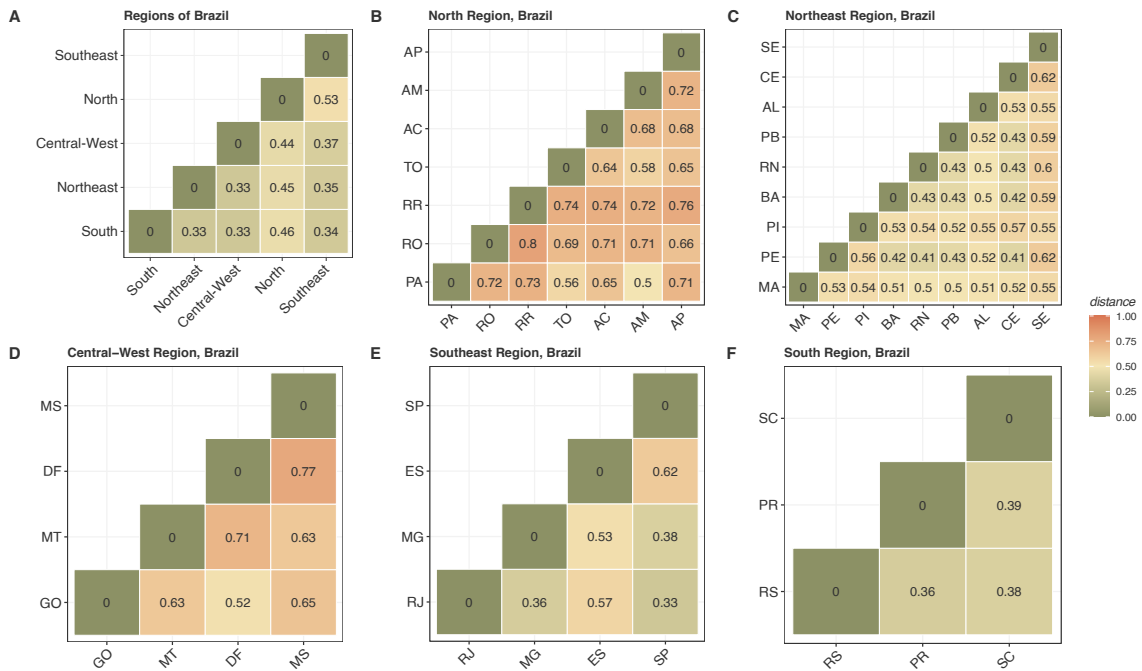
Figure 6. Cross-state Cultural Distance based on shared genres.

Regarding distances based on genres, Figure 6 reports similar results, although not as pronounced as in book preferences. The north region remains the most distant among all. Also, on the intra-cluster scale, north (0.68 on average), central-west (0.65 on average), and northeast (0.51 on average) regions are the most diverse. Such results support the presence of differences (and similarities) between Brazilian states and regarding reading preference (**RQ1**). Next, we assess which factors may significantly contribute to the cross-state cultural discrepancy by applying statistical tests.

5.2. Cross-state Demographic and Socioeconomic Analysis

To answer our second research question, we perform QAP tests to investigate the significance of socioeconomic and demographic factors on reading preferences in different states. In particular, the QAP test is commonly used in social network analysis to determine whether there is any statistical association between two adjacency matrices [Krackardt 1987]. Here, the dependent variables are matrices of cross-state cultural distances in book and genre preferences. The independent variables are matrices of cross-state geographical, demographic, and socioeconomic distances. Table 1 reports the QAP correlation coefficients among each adjacency matrix.

To evaluate the correlation results, we use Cohen's convention [Cohen 2013] to interpret the effect size: small (0.1 to 0.3), medium (0.3 to 0.5), and large (0.5 to 1.0). As expected, we confirm a strong and significant association between the cross-state cultural distances. Consequently, both analyses resulted in similar relationships between the vast majority of indicators analyzed, except for the GDP and the population estimate, which is stronger in the genre-based relationship. We also performed the same statistical analysis for each Brazilian region separately.[6] However, the outcomes revealed no significant

---

[6]Full results are available in Project Bàde homepage: `https://bit.ly/proj-bade`

Table 1. The QAP correlation results. Cross-marked values indicate the correlation is not statistically significant ($p \geq 0.05$).

	Factors	corr			Factors	corr
Books	Genre distance	0.914		*Genres*	Book distance	0.914
	Geographical locations	0.313			Geographical locations	0.344
	Gross Domestic Product	0.276			Gross Domestic Product	0.453
	Demographic density	0.200			Demographic density	0.291
	Human Development Index	×			Human Development Index	×
	Monthly household income	×			Monthly household income	×
	Territorial area	×			Territorial area	×
	Population estimate	0.337			Population estimate	0.409

association at the regional level. A possible explanation is that there may be distinct social contexts within each region and, therefore, the cross-region socioeconomic distances cannot explain the diversity in reading preference.

Overall, most of the significant relationships are positive and moderate, including the geographical distance (0.313 and 0.344), the GDP (0.453), and the population estimate (0.337 and 0.409). Such findings support that demographic and socioeconomic factors can directly impact the literary choices of the Brazilian population (**RQ2**), as given by Goodread users. However, the remaining indicators (demographic density, HDI, monthly household income, and territorial area) showed non-significant associations, with weak or insubstantial effect sizes. Considering some indicators individually (as a unique factor) may explain such a lack of solid and significant correlations. Perhaps studying the combination of certain external factors could reflect an improved association.

5.3. Brazilian Reading Identities

To explore the Brazilian cultural identity through reading preferences, we first model two unimodal projections based on the state-book and state-genre bipartite networks. In both graphs, two states have a weighted edge when shared books/genres are in their population's bookshelves. Next, with the unipartite networks modeled, we can apply a community detection algorithm to identify clusters of states with their own cultural identities. However, to do so, we need a method for determining significant associations in the state's projection network.

A simple approach to finding significant edges in unimodal projections is to preserve only links for which the weight beats a specific threshold. This approach, however, systematically overlooked nodes with a small strength (the sum of weights attached to ties belonging to a node). As a result, compelling features and structures below the cut-off scale are drastically removed. To address such a problem, we use the disparity filter proposed by [Serrano et al. 2009]. The filtering method extracts the relevant backbone from a weighted network and, at the same time, retains most of its nodes, total weight, global connectedness, small-world property, heterogeneous degree distribution, and clustering. Figure 7 shows the pre-filtering and post-filtering of the book-based unipartite network.

We use the 'backbone' function provided in R's *disparityfilter* package[7] to apply the disparity filter algorithm in both unipartite networks. Specifically, the 'backbone' function identifies the "backbone structure" of a weighted graph using the disparity filter

---

[7]disparityfilter: `https://CRAN.R-project.org/package=disparityfilter`
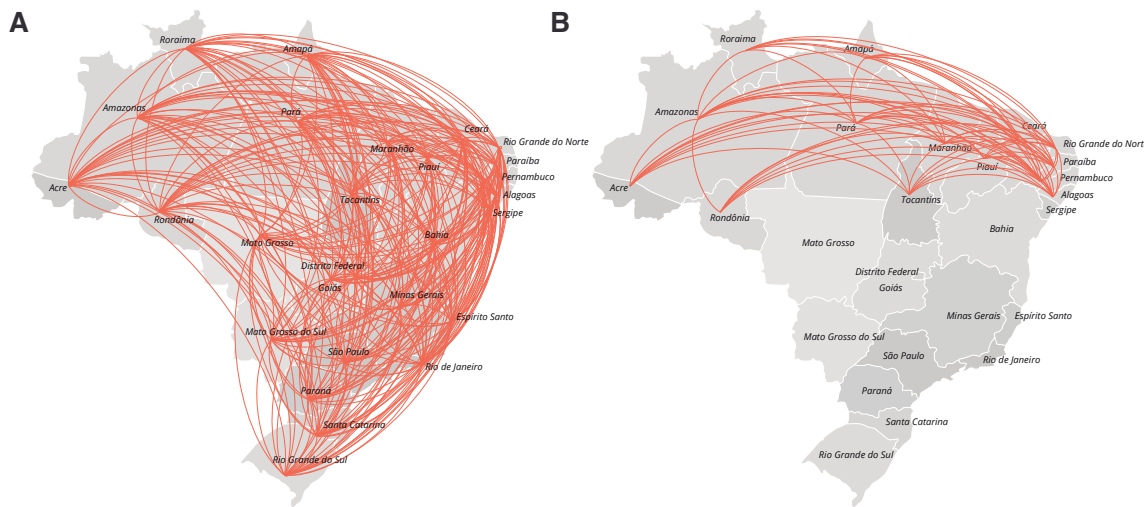
Figure 7. Comparison of the book-based unipartite network in (A) the pre-filtering and (B) the post-filtering.
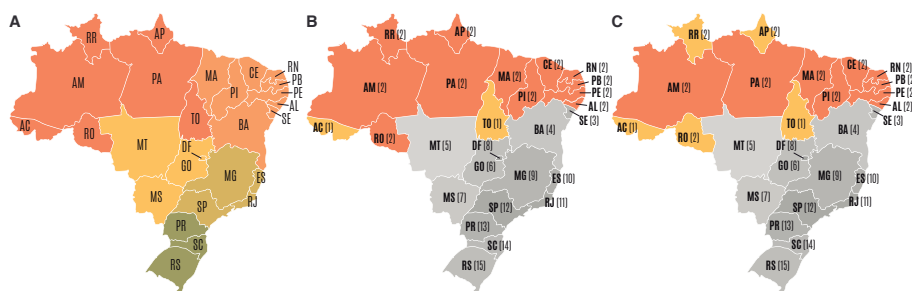

Figure 8. Groupings of Brazilian states: (A) regions, (B) book-based, and (C) genre-based unipartite network communities. Shades of gray indicate isolated communities.

algorithm. In the algorithm, the preserved edges (i.e., relevant edges) have weights statistically different from what is expected by chance at the significance level $alpha$. Thus, we chose an alpha that maximizes the number of remaining nodes in the network and minimizes the average node degree. Finally, we performed the Louvain community detection algorithm [Blondel et al. 2008] on the resulting networks.

Figure 8 shows three different groupings of the Brazilian states: **(A)** regions, **(B)** and **(C)** communities detected on the unipartite projections after removing statistically insignificant edges. The states with identical colors belong to the same community. For both book and genre resulting networks, the Louvain algorithm returned 15 different communities. Both grouping results reported a union of most of the north and northeast regions (except for the states of Bahia and Sergipe), as opposed to previous results. Moreover, some northern states gathered into a separate community, while the other Brazilian states, on the other hand, appear in unique communities. The disparity filter and, as a result, the statistical connections explain such controversial classification.

Answering our **RQ3**, we can assume it is possible to classify distinct reading identities in Brazil based on the states' reading preferences. However, the grouping of states depends directly on how cross-state connections are explored. By relying exclusively

on the cultural distances, the northern region stands out culturally, whereas the others are more similar to each other. On the other hand, by evaluating the statistically significant connections, the north and northeast regions form two distinct communities, with the other states forming unique communities. Overall, according to our findings, most Brazilian states do have their characteristics in the form of reading preferences.

6. Conclusion

In this paper, we explored the diverse Brazilian cultural identities through reading preferences. We shed light on cross-state differences and similarities based on their book and genre preferences. Additionally, through QAP tests, we revealed significant relationships between cultural distances and geographical, demographic, and socioeconomic indicators. Finally, we also used a bipartite projection and a community detection method to assess statistical significance relationships between states based on their reading preferences. Overall, our findings reinforce the cultural richness of Brazil, emphasizing that each state and region holds its own culture when it comes to what people chose to read. Such results provide insights into the driving forces behind Brazilian reading choices, indicating great opportunities for both the academy and the book industry.

From an academic view, this work represents a step further in interdisciplinary research fields involving Computer Science, Anthropology, and Sociology. For instance, knowledge from cultural behavior patterns may improve the accuracy of recommendation systems and prediction models. Besides understanding regional reading preferences, the Brazilian editorial market could properly direct its efforts to promote specific book releases for each region and state. Similarly, the government could benefit from such knowledge by investing in tax incentive policies to encourage reading habits since Brazil has lost more than 4.6 million readers from 2015 to 2019.[8] Overall, such investments directly impact the population by guaranteeing easy access to culture as well as raising literacy and other educational indicators.

Limitations and Future Work. Although Goodreads remains a valuable source of information on reading habits, our findings can be biased since our final dataset does not reflect the preferences of the whole Brazilian population (See Figure 2). Therefore, as future work, we plan to expand our data collection to better cover readers from all Brazilian regions, enhancing future analysis. Furthermore, we plan to explore other social indicators and consider combinations of such factors, aiming at stronger and more significant associations between cultural distances.

References

Blondel, V. D. et al. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008.

Borges de Souza, T. R. C. et al. (2015). Brazilian cultural differences and their effects on the web interfaces user experience. In *Int'l Conf Cross-Cultural Design Methods, Practice and Impact*, pages 209–220.

---

[8]Agência Brasil, (April 15, 2021). `https://bit.ly/34EijlY`

Butts, C. T. (2020). *sna: Tools for Social Network Analysis*. R package version 2.6.

Choi, J. H. et al. (2006). Comparing world city networks: a network analysis of internet backbone and air transport intercity linkages. *Global Networks*, 6(1):81–99.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Academic press.

Digiampietri, L. et al. (2014). Análise da rede dos doutores que atuam em computação no brasil. In *BraSNAM*, pages 33–44.

Fredrickson, M. M. and Chen, Y. (2019). Permutation and randomization tests for network analysis. *Social Networks*, 59:171–183.

Krackardt, D. (1987). Qap partialling as a test of spuriousness. *Social networks*, 9(2):171–186.

Lance, G. N. and Williams, W. T. (1966). Computer Programs for Hierarchical Polythetic Classification ("Similarity Analyses"). *The Computer Journal*, 9(1):60–64.

Lance, G. N. and Williams, W. T. (1967). Mixed-data classificatory programs i - agglomerative systems. *Australian Computer Journal*, 1(1):15–20.

Liu, M. et al. (2018). The relation of culture, socio-economics, and friendship to music preferences: A large-scale, cross-country study. *PloS one*, 13(12):e0208186.

Maity, S. K. et al. (2017). Book reading behavior on goodreads can predict the amazon best sellers. In *ASONAM*, pages 451–454.

Mondelli, M. L. B. et al. (2018). O que os países escutam: Analisando a rede de gêneros musicais ao redor do mundo. In *BraSNAM*. SBC.

Nascimento, M. L. et al. (2018). Uma análise do fator cultural em tecnologias persuasivas: um estudo de caso da rede social facebook. In *BraSNAM*. SBC.

Sabri, N. et al. (2020). A cross-country study on cultural similarities based on book preferences. *Soc. Netw. Anal. Min.*, 10(1):86.

Serrano, M. Á. et al. (2009). Extracting the multiscale backbone of complex weighted networks. *Proceedings of the National Academy of Sciences*, 106(16):6483–6488.

Shahsavari, S. et al. (2020). An automated pipeline for character and relationship extraction from readers literary book reviews on goodreads.com. In *WebSci*, pages 277–286.

Silva, M. O., Scotfield, C., Oliveira, G. P., Seufitelli, D. B., and Moro, M. M. (2021). BraCID: Brazilian Cultural Identity Information Through Reading Preferences.

Simpson, W. (2001). Qap: The quadratic assignment procedure. North American Stata Users' Group Meetings 2001 1.2.

Wang, K. et al. (2019). Exploring goodreads reviews for book impact assessment. *J. Informetrics*, 13(3):874–886.

Yucesoy, B., Wang, X., Huang, J., and Barabási, A. (2018). Success in books: a big data approach to bestsellers. *EPJ Data Sci.*, 7(1):7.

Zhao, Y. et al. (2021). Do cultural differences affect users' e-learning adoption? A meta-analysis. *Br. J. Educ. Technol.*, 52(1):20–41.