NOVA IMS Information Management School

MDSAA

Master Degree Program in
Data Science and Advanced Analytics

What are the structural differences in character networks within Portuguese-language literary works across various genres?

Catarina Ribeiro Duarte

Master Thesis

presented as partial requirement for obtaining a Master's Degree in Data Science and Advanced Analytics

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

What are the structural differences in character networks within Portuguese-language literary works across various genres?

by

Catarina Ribeiro Duarte

Master Thesis presented as partial requirement for obtaining the Master's degree in Data Science and Advanced Analytics, with a specialization in Data Science

Supervised by

Flávio Pinheiro, PhD, NOVA Information Management School
João L. M. Pereira, PhD, Universidade de Évora

July, 2024

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledged the Rules of Conduct and Code of Honor from the NOVA Information Management School.

Lisboa, 15th July 2024

DEDICATION

Para os meus Pais, Ana Isabel e João Paulo, que levo sempre comigo mesmo sem ser Joana.

ACKNOWLEDGEMENTS

I am deeply grateful to my primary supervisor, Professor Flávio Pinheiro, for his continuous support, invaluable guidance, and ever-present good mood. My thanks also extend to Professor João Pereira for joining as a co-advisor and enriching this work with his insights and knowledgeable point of view. Additionally, I would like to express my heartfelt appreciation to NOVA IMS for providing an excellent educational environment throughout my studies.

A distinctive thanks to Tiago Canário, an indisputable surprise and a fantastic teammate. Not only for the hard work, motivation, and companionship in hard challenges but also for the music recommendations that became the soundtrack for most of my writing sessions.

If this experience were a book, it would be a mix of genres, but mainly it was an adventure, a tale of perseverance, and an ode to the academic journey that brought me here. I'm so grateful for the characters in this narrative, whose network would be vast and dense, and I would like to thank the dear and important characters in this story.

Firstly, my whole family, whose trust and support I have always had and who passed on to me a love for books. Mãe, Pai, Mano, and Magali, you have been my home where I could always rest and find comfort in the middle of the journey in so many ways. Your support was vital to this narrative and kept the story moving forward.

Marta, who has been present and essential since the Prologue, even though from a different background, celebrated so deeply every conquest and struggle with me, the only one able to make 300 km seem so close. And Carolina, so intrinsically present in every chapter, always offering words of encouragement and teaching me in so many forms the meaning of 'unconditional'.

I also extend my gratitude to all my closest friends who added a touch of comedy to this tale; it would be a mystery how it could have been without all and each of you.

Lastly, thanks to Avó Celeste, whose belief in me endures.

ABSTRACT

This study investigates the structural differences in character networks within Portuguese-language literary works across various genres, including Fantasy, Mystery, and Romance. By exploring how character interactions and relationships differ across genres, we aimed to understand the underlying narrative structures that define each genre. Through Social Network Analysis, specific metrics were extracted for each genre, analyzed, and a Principal Component Analysis (PCA) was conducted, with real-life networks added for comparison and reference. Our analysis revealed that, although no extremely prominent metric differences were observed among the genres, there were notable structural tendencies within each genre. Fantasy narratives tend to focus on a single main character that interacts with many other characters, exhibiting lower density and clustering than other genres, and have less focus on relationship building. Mystery narratives feature moderately sized, denser networks with key bridging characters and the highest centrality on a relationship, translating into a balance of relationship and character building. In contrast, Romance narratives present the most variable network size, emphasizing multiple central characters and cohesive subgroups, focusing especially on relationship building. This makes Romance and Fantasy the most different genres, while Mystery acts as a mix of these structure strategies and is simultaneously closer to what is observed in real-life networks. The study acknowledges limitations related to sample size and data processing, highlighting the need for future research with larger and more diverse datasets to validate these findings.

KEYWORDS

Social Networks Analysis (SNA); NLP; Genre Theory; Portuguese Literature

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

Understanding the structural differences in character networks within Portuguese-language literary works across various genres is a complex yet intriguing challenge. Despite most readers being able to identify their favorite book genre, the literary academic field lacks consensus on the precise characteristics that define each genre (Chandler, 1997). This study aims to fill this gap by exploring how character networks vary across genres within Portuguese literature, leveraging the power of Natural Language Processing (NLP) and Social Network Analysis (SNA).

The Portuguese language boasts a literary tradition spanning eight centuries, originating from Galician roots and evolving through several significant periods, each leaving a distinct mark on its development (Earle et al., 2013). This rich history includes medieval poetry, chronicles of Portuguese society, theatrical pieces, and epics narrating overseas expansion, extending to modern works (dos Reis, 1993). Portuguese-language literature, a fundamental aspect of Portuguese culture, has been the subject of scholarly investigation over the years regarding its progression and heritage (Tamen & Buescu, 2013).

Given the unique characteristics of Portuguese literature and its genres, there is substantial potential for applying advanced analytical techniques to enhance our understanding. Natural Language Processing (NLP) provides innovative solutions for data-driven challenges by enabling the automatic analysis and representation of human language (Cambria & White, 2014). NLP is essential for analyzing and understanding literature across various linguistic contexts, significantly enhancing the creation and analysis of character networks (Labatut & Bost, 2020), named entity recognition (Naseer et al., 2021; Amaral et al., n.d.), and semantic analysis (Maulud et al., 2021), that allows this study to delve deeper into the dynamics within literary texts and analyse them.

To address the gap in understanding the structural differences in character networks across genres within Portuguese-language literary works, this study utilizes Social Network Analysis (SNA) (Newman, 2010). SNA provides a framework to study how the web of relationships between social actors influences social phenomena (Borgatti et al., 2018). By applying graph theory and network science methods, SNA has been applied in a broad range of disciplines, demonstrating wide-reaching implications (Otte & Rousseau, 2002). From analyzing complex social structures to understanding individual behaviors, SNA aids diverse fields such as health research (Fernández-Peña et al., 2022), animal welfare (Norman et al., 2023), and scientific collaboration (Fonseca et al., 2016). In literary studies, SNA helps map the relationships between characters (Park et al., 2013), offering a deeper understanding of how intricate dynamics and relationships drive literary narratives.

Inspired by Silva et al. (2023) and Mamede & Chaleira (2004), we utilize the TAGGUS (Canário & Duarte, 2024) information retrieval pipeline to identify and extract characters and their relationships from Portuguese literary works. This extracted information is then used to infer the social networks of characters and analyze the characteristics of each work across various genres. This study assesses how character networks compare between genres in the Portuguese language using network properties and graph specific measures to compare networks, also conducting a principal component analysis (PCA) and adding real life networks to draw a comparison and reference values for the metrics.

2. RELATED WORK

Network analysis is a powerful framework commonly used to study the interconnected nature of physical and social systems. It offers a means to study the organization and functionality of complex systems (Freeman, 2004; Borgatti et al., 2009; Mata, 2020; Bento et al., 2020; Newman, 2010). Thus, it is unsurprising that network analysis has been applied to entertainment. Examples range from biblical texts (Massey, 2016) to novels (Jarynowski & Boland, 2016) but also include manga (Sugishita & Masuda, 2023), films (Mourchid et al., 2019), and TV shows (Fronzetti Colladon & Naldi, 2019).

This application of network analysis is particularly intriguing when considering the complex landscape of Genre Theory. At first glance, the delineation between literary genres such as fiction and non-fiction might appear straightforward: non-fiction adheres to factual narratives of real events or phenomena, and fiction, despite occasionally incorporating non-fiction elements, primarily weaving stories from imagination and depicting events that are not factual. However, the landscape of Genre Theory is markedly complex, having been the subject of academic exploration for decades (Wellek & Warren, 1956; Fowler, 1982). Wellek & Warren (1956), introduce a theoretical framework for understanding literary genres, emphasizing that genres are dynamic, evolving constructs rather than static categories. This perspective is echoed and expanded upon by Tzvetan Todorov (1978), who articulated that the emergence of new literary genres often entails the transformation of existing genres, whether by inversion, displacement, or combination.

Lopes (2010) states the importance of thematic, structural, and formal elements in genre classification in her examination of both literary and journalistic genres. This notion of fluidity and subjectivity in genre delineation is further supported by Chandler (1997), who remarks, "*One theorist's genre may be another's sub-genre or even super-genre, and what is technique, style, mode, formula, or thematic grouping to one may be treated as a genre by another*." Bordwell (1989) argues that themes alone are inadequate for defining genres since any theme can manifest across different genres. Genre theory's intricate and evolving nature challenges categorizing literary works within rigid boundaries.

Genre theory is key for developing recommendation systems. Indeed, such information filtering systems can broadly predict whether a particular user would prefer an item based on the user's profile (Isinkaye et al., 2015). However, when focusing on entertainment, models benefit from metadata such as the genre (Middha et al., 2022). Reddy et al. (2019) and Kim and Moon (2011) achieved notable results in developing content-based movie recommendation systems using genre correlation, similarity, and preferred genres. Similarly, in book social media platforms like Goodreads[1] or StoryGraph[2] which have their own

[1] https://www.goodreads.com
[2] https://app.thestorygraph.com

recommending systems for literary works, the genre is a significant indicator and factor for better performance and user satisfaction and experience.

In this context, the analysis of texts, scripts, and images (Park et al., 2009) becomes essential as it allows for the inference of fictional social networks, which can underscore the significance of character networks in dissecting and understanding fictional narratives (Labatut & Bost, 2020). The extraction process of these networks and their characteristics vary according on the narrative's source, the network's analysis goal, and other factors. However, indeed, character networks can simplify a story's plot by unveiling its details and summarizing it through characters and their interactions, making them useful tools to aid genre theory.

Ardanuy et al. (2015) explored the correlation between character networks within novels and their genre classification, highlighting how novelistic subgenres are flexible in style and structure. By constructing character-based social networks, the study applied two clustering methods: one by genre and the other by author. The results revealed the inherent challenges of genre-based clustering, reflecting the broader complexities and ambiguities within genre theory itself. Conversely, author-based clustering provided more definitive results, suggesting that a novel's network structure more accurately reflects an author's stylistic fingerprint than its genre classification.

The pioneer work by Alberich et al. (2002) on the fictional Marvel Universe demonstrates that characters' social networks exhibit many properties commonly found in real-world social networks. More recently, Gessey-Jones et al. (2020) mapped the web of relationships between the characters of "*A Song of Fire and Ice,*" showing that not only did the network have similarities with real-world social networks but that it also helped predict character deaths and understand plot development dynamics.

Holanda et al. (2018) categorized genres as pure fiction, legendary, and biographical by analyzing character networks and network-specific measures such as density, clustering coefficient, degree, betweenness, closeness, and the Lobby index to analyze. Rahul et al. (2021) used data from a Harry Potter fan fiction community (FanFic.net), which consisted of 820,000 works of fiction written in over 50 languages and 20 genres, to achieve an F1-score was 80.52% in a genre classification task supported by machine learning. Their findings also underscored the challenging nature of genre classification based solely on social network analysis, but also the nuanced reality that a single work might embody multiple genres.

The exploration of character networks and genre classification is not confined to English-language literature. Hettinger et al. (2015) combined name-entity recognition techniques and network centrality measures to identify important characters and interactions in German novels. His work highlights the potential and challenges to genre classification in German literature. Fan & Li (2022) proposed innovative methods for extracting and analyzing character relationships in Chinese literature, emphasizing the complexity of character networks, showcasing the use of metrics like degree distribution and clustering coefficient for literary analysis. The study highlighted the role of network structures in revealing genre-specific patterns, suggesting that such analytical approaches could be valuable across linguistic boundaries. Silva et al. (2023) applied structural network analysis and character importance metrics to examine the intricate relationships between characters in Portuguese literature. The study also highlights the challenges of automated text processing in Portuguese due to linguistic complexities and limited natural language processing (NLP) resources compared to English. These studies emphasize the intricate relationship between character networks and literary genre classification. Despite advances in methodology and the application of machine learning, the multifaceted nature of genres, compounded by subjective labeling and the inherent diversity of literary styles, continues to pose significant challenges to definitive classifications.

Finally, despite its importance and application, Moretti (2011) highlights that transforming plots into networks in literature presents unique challenges. A significant portion of literature's character actions, interactions, and dialogues are conveyed through narration rather than direct discourse. This aspect can render network representations less precise, especially given that direct discourse constitutes only a small segment of the entire plot, potentially skewing the understanding of narrative dynamics. Labatut and Bost (2020) emphasize that character networks help in summarization, classification, and role detection, but these tasks are harder due to the specific properties of fictional works. They provide a survey on the extraction and analysis of these networks, illustrating the relevance of character networks through various applications while identifying limitations and promising perspectives for future research.

3. METHODOLOGY

This section outlines the methodology adopted to analyze the structural differences in character networks within Portuguese-language literary works across various genres. This work necessitates a meticulous selection of genres and literary works, which form the foundational framework for translating textual elements into character networks. These networks will then be measured using relevant metrics chosen to serve as comparison points among genres. The processes detailed below aim to capture the unique narrative dynamics that characterize each genre and highlight their differences.

We start by outline the genres that were chosen for comparison. The focus was on subgenres of Fictional Narratives to create richer networks and set more coherent expectations for these subgenres. Sources differ in categorizing the same novels; some novels are labeled with multiple genres, while others are not categorized at all. To address this challenge and enable a more robust and distinct comparison, we selected genres where the reader's expectations are clearly defined. For example, readers anticipate a love story in a romance novel or a crime to be solved in a mystery novel. This contrasts with genres like adventure and fantasy, which may have more overlapping themes and less distinct reader expectations (Foster, J. n.d.).

Regarding book selection, the sources for genre tagging included “Projeto Adamastor”[3] and the Goodreads[4] website.

“Projeto Adamastor” is a digital library of public domain Portuguese-language works and it has been utilized in Silva et al. (2023) research. Each book is tagged with its genre, allowing for extraction and genre classification for use in this study. This platform is continuously updated, and it is noteworthy not only for converting texts into digital format but also for its meticulous review of each available work to minimize errors and adhere to the current Portuguese Spelling Reform. This is crucial for this study as it provides a base text that is the least error-prone and thus will produce the most accurate character networks. According to the “Collaborator Guide for Projeto Adamastor” by Lourenço, R.F. (2014), each collaborator identifies a genre during the selection and book treatment process, which is then validated by a peer review involving at least one collaborator not involved in the initial review. They conduct a comprehensive reading of the file to ensure the quality of the book and the accuracy of genre tagging.

“Goodreads” is a book-based social website where members share, review, and rate literary works and connect with other readers (Thelwall & Kousha, 2017). On this platform, genres are attributed in a manner that considers several factors but remains largely undisclosed to users. It is known, however, that user influence plays a role due to genre tags reflecting the self-created names by its community. Shelves are a method for users to organize and group their

[3] https://projectoadamastor.org
[4] https://www.goodreads.com

books, often named after genres, such as romance or crime. These names influence genre assignments for the books, as the genre lists on Goodreads book home pages include up to ten of the most popular assigned book genres (Thelwall, 2019).

For the book selection within each genre, it was ensured that the book, whether on “Projeto Adamastor” or among the first five genre tags on Goodreads, was identified with that genre. Although no genre assignment strategy can perfectly categorize genres and ensure a book fits only one category, these precautions and methodological approaches were adopted for this study.

The chosen genres, all within the realm of fiction, in the Portuguese language, including Brazilian Portuguese, are presented in table 1 with their respective definitions from Goodreads. This alignment provides a preview of reader expectations and the sample collected focused on presenting more contemporary books.

Regarding the pre-processing required for the book text before passing it into the pipeline to extract the networks, only a simple cleanup is needed to remove non-narrative text from the digital versions of the books. Therefore, to correct this issue, the initial or ending unnecessary chunks were manually removed from the TXT files of the books.

The pipeline used for this work, TAGGUS (Canário & Duarte,2024) was specifically designed to extract character networks from Portuguese novels, addressing and overcoming all the impediments and constraints associated with the language that have been previously mentioned. Detailed information about the entire pipeline process and the decisions made can be found in it’s Github repository[5].

Nevertheless, Labatut & Bost (2020), in their widely cited work, outlined the extraction process, which also served as the foundation for the pipeline's creation and it’s represented in figure 1.

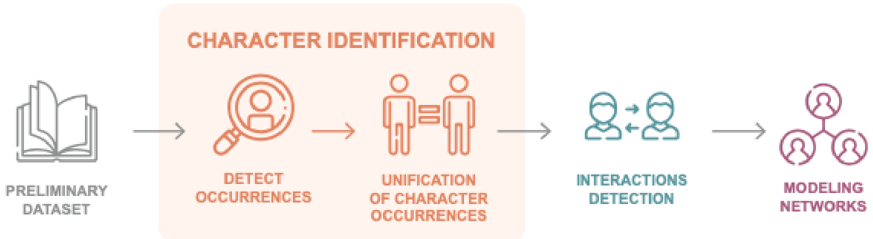

Figure 1 - Overview of the generic character network extraction methodology

[5] https://github.com/ticadu/taggus.

Table 1 - Chosen Genres and selected books for each

Book Genre	Goodreads definition	Author, Book title and year
Romance[6]	"Two basic elements comprise every romance novel: a central love story and an emotionally satisfying and optimistic ending." Both the conflict and the climax of the novel should be directly related to that core theme of developing a romantic relationship, although the novel can also contain subplots that do not specifically relate to the main characters' romantic love"	Valter Hugo Mãe, *"O Filho de Mil Homens"* (2011)[7] Lídia Jorge, "*A Costa dos Murmúrios*" (1988)[8] Miguel Sousa Tavares, *"Equador"* (2003)[9] João Ricardo Pedro, *"O teu Rosto será o último"* (2012) [10]
Fantasy[11]	Fantasy is a genre that uses magic and other supernatural forms as a primary element of plot, theme, and/or setting. Fantasy is generally distinguished from science fiction and horror by the expectation that it steers clear of technological and macabre themes, respectively [...].	Thiago d'Evecque, *"Limbo"* (2015) [12] Eduardo Spohr, *"A batalha do Apocalipse"* (2007) [13] Sandra Carvalho, "*A Última Feiticeira*" (2005) [14] Filipe Faria, "*Oblívio*" (2011)[15]
Mystery[16]	The mystery genre is a genre of fiction that follows a crime [...] from the moment it is committed to the moment it is solved. Mystery novels often [...] turn the reader into a detective trying to figure out the who, what, when, and how of a particular crime.	Pedro Garcia Rosado, *"A Cidade do Medo"* (2015) [17] Nuno Nepomuceno, "*A Célula Adormecida*" (2007) [18] Raphael Montes, "*Dias perfeitos*" (2005)[19] Francisco Moita Flores, *"O Bairro da Estrela Polar"* (2012)[20]

[6] https://www.goodreads.com/genres/romance
[7] https://www.goodreads.com/book/show/12395665-o-filho-de-mil-homens
[8] https://www.goodreads.com/book/show/3371637-a-costa-dos-murm-rios
[9] https://www.goodreads.com/book/show/1140942.Equador
[10] https://www.goodreads.com/book/show/13597497-o-teu-rosto-ser-o-ltimo
[11] https://www.goodreads.com/genres/fantasy
[12] https://www.goodreads.com/book/show/25844730-limbo
[13] https://goodreads.com/book/show/18299638-a-batalha-do-apocalipse
[14] https://www.goodreads.com/book/show/6333188-a-ltima-feiticeira
[15] https://www.goodreads.com/book/show/10474666-obl-vio
[16] https://www.goodreads.com/genres/mystery
[17] https://www.goodreads.com/book/show/12742439-a-cidade-do-medo
[18] https://www.goodreads.com/book/show/32574062-a-c-lula-adormecida
[19] https://www.goodreads.com/book/show/21405408-dias-perfeitos
[20] https://www.goodreads.com/book/show/16155430-o-bairro-da-estrela-polar

The process begins with the preliminary dataset, a single literature work, which is fed into the model's pipeline. The first step is character identification, divided into two sequential stages. Initially, all the different character names are identified and stored. However, not every name represents a unique character, as narratives, like real life, often use various names to refer to the same character. For example, in the work *Project Hail Mary*[21], *Ryland Grace* may also be referred to in the same story as *Mr. Grace*, representing the same character but by another expression. Therefore, a co-reference resolution step is employed to unify character occurrences. The pipeline used in this work takes precautions to ensures accurate unification by matching gender, nicknames, likelihood of the most popular character, and anaphoric resolution.

Following character identification, the next step is to detect interactions between characters. In this work, co-occurrence is used to count interactions. This method decomposes the narrative into smaller units. These units may vary from project to project, considering two characters to interact when they appear together in the same unit (Labatut & Bost, 2020). Although this technique is widely used due to its simplicity and ease of implementation, it has the disadvantage of potentially labeling simply two characters mentioned in the same unit as interactions, leading to false positive interaction counts. However, this remains as the most widespread approach in the literature for this issue (Labatut & Bost, 2020). In this pipeline the narrative unit chosen to detect interaction was two characters mention in the same sentence. Finally, after unifying the characters and detecting their interactions, a graph is constructed.

The graph that represents a character network is built with nodes and edges, each having its own definition. The nodes represent the characters, and the edges represent the relationships among them. Moreover, these graphs can also incorporate temporal integration, resulting in either a static network or a dynamic network (Labatut & Bost, 2020).

The static network, which is the approach most commonly presented in the literature, captures interactions between characters over the entire narrative period. In contrast, the dynamic network integrates the progression of interactions over time or across different sections of the story. While the static network allows for a better visualization of the novel, it does not capture the evolution of characters throughout the narrative (Labatut & Bost, 2020). For this study, given the goal of understanding and analyzing the global story and the lack of necessity for temporal dimension analysis, static networks were used.

The simpler element, nodes, can represent individual characters or groups of characters and also present character characteristics, such as character type (e.g., human, dog, etc.), gender, race, abilities, etc. Moreover, weighted nodes can give us an idea of the relevance of the characters throughout the story. On the other hand, the edges represent the interactions and connections among characters and can contain more information. Interaction intensity can be shown through weighted nodes. Additionally, the edges can have directions, creating a

[21] https://www.goodreads.com/book/show/54493401-project-hail-mary

directed graph that transmits the speaker/addressee in the interaction. Furthermore, assigned graph properties can present the interaction polarity that conveys the sentiment of the interaction/relationship (Labatut & Bost, 2020).

In this work, the static networks will feature weighted nodes, where each node represents a distinct character. While the nodes are weighted to indicate the relevance of each character, the edges in our network will be weighted as well, undirected, unsigned, and unattributed. An example can be seen in figure 2 and a representative network of each genre can be found in Appendix A as well as its respective adjacency matrix.

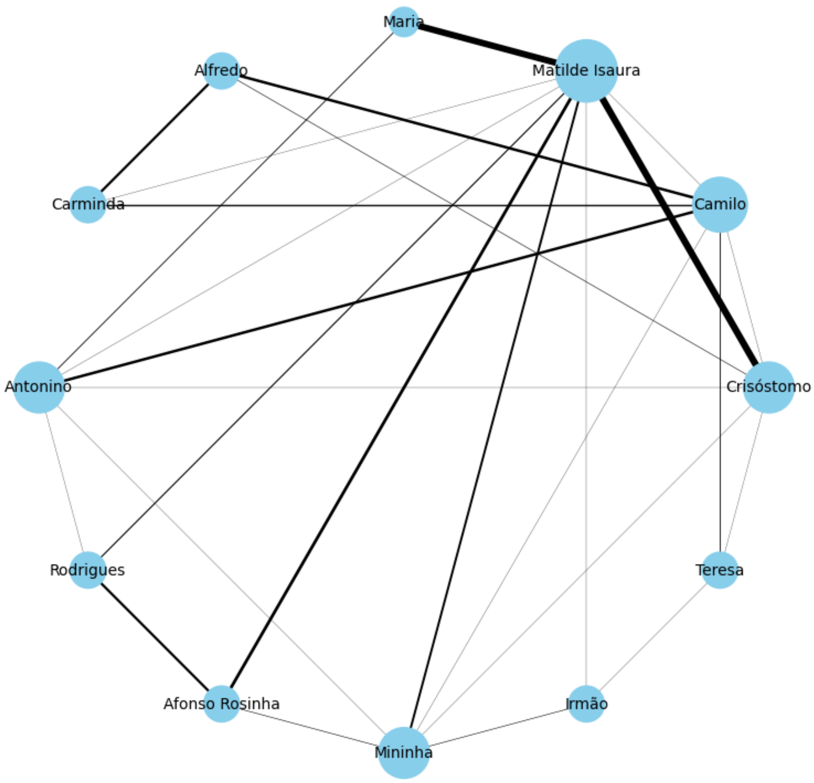

Figure 2 - Character Network for "O Filho de Mil Homens"

When comparing character networks, we focused on measurable aspects that could be translated to narrative significance and analyzed for genre-specific differences. For each extracted character network, we computed the network size (N) that informs on the number of nodes of the network, its size. In our context it gives us the number of characters of each book (Silva et al., 2023; Suen et al., 2013). Silva et al. (2023) analysis showed that the network size varies little between Portuguese literary works, with an average of nine characters in her sample. In the case of figure 2 it would be 12 nodes so, 12 characters.

Furthermore, the number of components counts the subset of the vertices of a network that are connect, this is, that there exists at least one path from each member of that subset to each other member (Newman, 2010). In the context of character networks, analyzing the number of components can reveal how fragmented or cohesive the narrative structure is. For example, a high number of components might indicate multiple isolated subplots or character groups that do not interact, while a low number of components, and especially a single one, suggests a more integrated and interconnected story.

Next, maximum degree shows the highest number of connections a single character has, this is the highest degree of any node in the network (Silva et al., 2023; Labatut & Bost, 2020). This metric helps identify the most central or influential character within the narrative and how connected it is, providing insights into the structure and dynamics of the character interactions. Sudhahar & Cristianini (2013) found in their work that the hero of a narrative always had the highest degree in a network, showing their central position in the narrative. In figure 2 it's clear that these would be the number of connections *Matilde* has, due to being the most connected character.

In our work, we propose a new metric, "Number of Main Characters" (NºMC) which aims to count, besides the character with the highest degree, how many other central characters exist. This is done by counting any other node with a number of interactions (k_i) close to the highest degree (k_{max}), with a threshold of at least 80% and above a static lower bound of 10 interactions. This metric identifies both central and "secondary" main characters, reflecting narratives that feature multiple significant characters rather than focusing solely on one. The lower bound of 10 interactions ensures that characters considered as main have a meaningful level of engagement in the story. While metrics like maximum degree and single character centrality (SCC) highlight the most central character, they may overlook other important characters. The "Number of Main Characters" metric considers characters with substantial interactions, providing a more comprehensive view of the narrative's structure and recognizing the importance of key secondary characters alongside the primary one.

$$N^{\text{º}}MC = \sum_{\boldsymbol{i}} [k_i \geq 0.8\ k_{\boldsymbol{max}} \wedge k_i \geq 10] \quad \textbf{(1)}$$

Still in the topic of connectivity, density measures how closely the characters are connected in the network. This is possible by calculating the proportion between existing edges (E) in the network and all possible network connections (Silva et al., 2023; Labatut & Bost, 2020; Dekker et al., 2019). The possible connections value is the count of pair of nodes that can form an edge. The density value ranges from 0 to 1, where a value closer to 1 indicates a highly interconnected network, and a value closer to 0 suggests a sparsely connected network. For example, Ardanuy & Sporleder (2015) note that in the case of the *Harry Potter*[22] books, the earlier books are considerably denser than the later ones, as the community represented in them becomes broader and less tightly knit with every new book. This shows how the density

[22] https://www.goodreads.com/series/45175-harry-potter

metric can reflect difference in the structure and focus of a narrative, in this case in a series. The equation (2) used to calculate it was retrieved from Bhattacharya et al. (2023).

$$d = \frac{2E}{N(N-1)} \quad (\mathbf{2})$$

The average path length measures the average number of steps to achieve the shortest paths for all possible pairs of nodes in the network (Silva et al., 2023; Labatut & Bost, 2020; Dekker et al., 2019; Newman, 2010). This metric provides insight into how connected the nodes, characters, are within the network, reflecting the overall efficiency of information or interaction flow within the narrative. A lower average path length suggests a close community where any character can be reached from any other character through a few interactions. Conversely, a higher average path length indicates a more spread out or fragmented network. In our study, in the case of when we have a fragmented network, we have in consideration the average path length of the largest component since it's the best and fairest metric for that narrative. The real worlds concept of six degrees of separation, which suggests that any two people are, on average, six or fewer social connections apart (Milgram, 1967), can illustrate average path length usage in social networks. The defined equation (3) was retrieved from Newman (2010) and d_{ij} is the shortest distance between characters i and j.

$$APL = \frac{1}{N(N-1)} \sum_{i \neq j} d_{ij} \quad (\mathbf{3})$$

Now segueing into the centrality measures, which help identify the most important or influential nodes within the network, the single character centrality (SCC) measure for individual characters was proposed by Suen et al. (2013) to see how much a story is focused on a single character. This is achieved by measuring the disparity between the character with the highest weighted degree (s_i) and the second highest, normalized, and understanding the difference. While maximum degree provides a direct measure of a character's interactions, SCC quantifies the dominance or disparity in centrality within the network, indicating how much the story focuses on one character above all others. The SCC value ranges between 0 and 1, where a value closer to 1 indicates high dominance of a single character, and a value closer to 0 suggests a more evenly distributed character network

$$SCC = \frac{max_i(s_i) - next\_max_i(s_i)}{\sum_i s_i} \qquad (4)$$

Also proposed by Suen et al. (2013), single relationship centrality (SRC) measures the prominence of the most central relationship relative to others in the network. This metric quantifies how much a single relationship stands out in terms of interactions compared to the second most significant relationship. It helps identify the most dominant pair of characters, this is, the edge with more weight w_{ij}, and their influence on the narrative. While single SCC focuses on the dominance of a single character, single relationship SRC highlights the importance of a specific interaction between two characters following the same scale logic.

$$SRC = \frac{max_{i,j}(w_{i,j}) - next\_max_{i,j}(w_{i,j})}{\sum_{i,j} w_{i,j}} \qquad (5)$$

The betweenness centrality (BC) measures the extent to which a certain node is on the shortest paths between other nodes (Hansen et al., 2020; Labatut & Bost, 2020; Suen et al., 2013; Hettinger et al., 2015). In other words, it helps identify characters who play a "bridge" role in a network. First, calculate the betweenness centrality for each node by counting the number of times it lies on the shortest path between other nodes ($\sigma_{st}(i)$). Then, add up all the betweenness centrality values for the nodes in the network. Next, normalize each value by dividing it by the maximum possible value for any node in the network, ensuring all values fall between 0 and 1. Finally, compute the average betweenness centrality by dividing the total sum of the normalized values by the number of nodes in the network. By averaging it for the whole network, average betweenness centrality (ABC), we can capture the overall importance of intermediary nodes within the entire network. A value of 0 for average betweenness centrality means that, on average, nodes do not act as intermediaries; this could be due to a completely disconnected network or one where all the shortest paths are direct. In contrast, a high positive value means that many characters play a critical role in facilitating interactions or the flow of information between nodes.

$$ABC = \frac{1}{N}\sum_{i=1}^{N} \frac{2}{(N-1)*(N-2)} \sum_{s \neq i \neq j} \frac{\sigma_{st}(i)}{\sigma_{st}} \qquad (6)$$

The degree assortativity measures how correlated the degrees of connected vertices are, in other words, it measures the tendency of nodes to connect to similar nodes with respect to their degree (Silva et al., 2023; Labatut & Bost, 2020). This metric helps characterize the relationship between primary and minor characters. Being a coefficient, assortativity is calculated as the ratio of covariance to variance. To calculate the covariance, we look at the connections, degree k, of every node and then determine the joint degree distribution, e_{jk}. This joint degree distribution indicates how often nodes with a certain degree k are connected to nodes with degree j. In simpler terms, it measures the frequency of edges existing between nodes of different degrees. Then, this covariance is divided by the variance of the degree distribution, $\sigma^2 = \sum_{k} k^2 q_k - [\sum_k k^2 q_k]^2$. The result takes a value between 1, assortative, where nodes with similar degrees tend to connect with each other, and -1, disassortative, where nodes with differing degrees tend to connect, or 0 (neutral) in the case of non-correlation (Bhattacharya et al., 2023). In Silva et al. (2023), the networks were generally disassortative, reflecting the diversity of social interactions in the analysed works much like in real-world networks. The used formula was collected from Massey (2016).

$$r = \frac{\sum_{jk}(e_{jk} - q_j q_k)}{\sigma_q^2} \tag{8}$$

Lastly, the average clustering coefficient (ACC) measures the degree to which nodes in a network tend to cluster together, providing insight into the overall tendency of characters to form cohesive and closed groups (Newman, 2010; Suen et al., 2013; Ardanuy et al., 2015; Dekker et al., 2019; Labatut & Bost, 2020). This metric is calculated as the average of the local clustering coefficients of all nodes in the network. The local clustering coefficient is the probability, when randomly picking two neighbors of a node, k_i, that there is an edge between them m_i (Labatut & Bost, 2020) then averaging it by the number of nodes N. The ACC is useful for understanding the general clustering tendency across all nodes in the network. A higher ACC indicates a greater likelihood of strong community structures within the narrative. Dekker et al. (2019) provide a comparative analysis of clustering coefficients in social networks from novels and various well-known networks such as youtube and Flickr, demonstrating similar variation patterns across different network types. The used formula was collected from Massey (2016).

$$\bar{C} = \frac{1}{N} \sum_{i=1}^{N} \frac{2m_i}{k_i\,(k_i - 1)} \tag{9}$$

The measures were calculated individually for each novel using Python with the NetworkX package [23] and then averaged by genre.

Moreover, a Principal Component Analysis (PCA) was conducted with the metrics. PCA is a multivariate technique that extracts the most important information from the data by

[23] https://networkx.org

reducing its dimensionality thought the creation of new variables that are linear combinations of the original variables. This process simplifies the dataset and allows for an analysis of its structure (Abdi & Williams, 2010). The goal is to understand the key metrics that distinguish genres while also have a visual representations of their grouping.

4. RESULTS AND DISCUSSION

In this section, we discuss the results of our comparative analysis between literary character networks from different genres using graph measures. Since our research also aims to draw parallels between these networks and real-life networks, existing datasets available on popular websites, such as the Network Repository[24], were added to analysis with the book presented before. The networks chosen for inclusion in this work had to provide a meaningful comparison to the literary character networks, representing various types of social interactions with a comparable number of nodes, interaction contexts, and settings for the interactions.

The real-life networks added to our analysis include Zachary's karate club[25], which was extracted from a university-based karate club and collected over three years from activities in which club members attended, such as lessons and other social events (Zachary, 1977). Additionally, we included the Moreno High School dataset[26], which contains friendships between boys in a small high school in Illinois, and a contact network from a village in rural Malawi, collected through proximity sensors technology that measured social contacts (Ozella et al., 2021). The results for the graph comparison measures presented in the previous section were obtained from the analysis for each genre and real-life and are shown in table 2, with the standard deviation presented in parentheses below each value.

Starting with Fantasy, this genre distinguishes itself by having a single main character and the highest value for single character centrality, with both metrics showing low to no standard deviation. Fantasy also has a high maximum degree, although with significant standard deviations, highlighting a focus on a prominent central hero who interacts with many other characters. An infinite value in the average path length indicates that in at least one novel, there are characters who are disconnected and isolated. However, the average path length of its largest component is lower than other genres and real life, making this genre the one with the most complete connectivity within its main group. The low average betweenness centrality suggests a lesser reliance on intermediary characters reinforcing the focus on a main character. Moreover, Fantasy presents the lowest values for density and clustering coefficient among the genres, indicating a lower level of interconnectedness and a lesser tendency for its characters to form tightly knit groups compared to Mystery and Romance. These structural features align with the narrative expectation of a Fantasy novel, which often centers around a hero's journey interacting with a few key characters along the way.

[24] https://networkrepository.com
[25] https://networkrepository.com/karate.php
[26] https://networkrepository.com/moreno-highschool.php

Table 2 - Averaged network metrics by genre (with standard deviations)

Metric \ Genre	Fantasy	Mystery	Romance	Real-life networks
Network Size	42 (16.10)	30 (14.38)	38 (25.88)	63 (26.63)
Components	1.25 (0.05)	1.0 (00.00)	1.5 (00.58)	1.3 (00.58)
Density	0.1994 (0.08)	0.2279 (0.12)	0.2096 (0.14)	0.1158 (0.02)
Assortativity	-0.2762 (0.07)	-0.2869 (0.11)	-0.1727 (0.09)	-0.1188 (0.91)
Maximum Degree	263 (190.91)	127 (34.90)	108 (111.72)	32.7 (14.57)
Single Character Centrality	0.0786 (0.04)	0.0409 (0.02)	0.0689 (0.07)	0.0092 (0.00)
Single Relationship Centrality	0.0338 (0.05)	0.0665 (0.10)	0.0096 (0.01)	0.0018 (0.00)
Number of Main Characters	1.0 (0.00)	1.5 (0.58)	1.5 (0.58)	1.6 (0.58)
Average Path Length	inf (NaN)	3.3194 (0.25)	inf (NaN)	inf (NaN)
APL of largest Component	2.9622 (0.26)	3.3194 (0.51)	3.1008 (0.45)	3.6695 (1.81)
Average Betweenness Centrality	0.0456 (0.02)	0.0655 (0.02)	0.0644 (0.04)	0.0307 (0.02)
Clustering Coefficient	0.0447 (0.02)	0.0497 (0.01)	0.0623 (0.03)	0.4111 (0.15)

Regarding the Mystery genre, its networks are moderately sized yet still smaller than those of other genres, while simultaneously being denser. The longer average path lengths suggest that characters are more disconnected and spread out but, unlike the other genres and real-life, they form a connected network having a single component. This level of connectedness is complemented by a high average betweenness centrality, indicating the presence of bridging characters. The number of main characters is above one, which indicates that in some of the novels the focus is not on a single character. Furthermore, this genre has the highest single relationship centrality among all genres, suggesting that specific relationships between characters are particularly significant in these stories. Additionally, the clustering coefficient indicates fewer tightly knit groups within the networks. These structural features align with the narrative expectations of Mystery novels, which often revolve around complex plots and intricate character interactions.

Onto the Romance genre, these novels present large and variable network sizes, indicating a wide range of character networks from small to very large. These networks have moderate density and the least assortativity, which means there is a tendency for characters to connect with dissimilar others. Its level of assortativity is closest to that in real-world networks, suggesting it better represents societal interaction patterns. The Romance genre also shows the lowest maximum degree, indicating that characters are not as highly connected. The infinite average path length in two novels indicates isolated characters, making the network disconnected. Despite this, it has a high clustering coefficient, indicating the presence of close-knit groups. These characteristics, such as low assortativity, high clustering, and the presence of isolated characters, reflect the diverse and intricate subplots in Romance narratives. Moreover, the low single relationship centrality suggests a focus on multiple relationships rather than highlighting a single one, showcasing the diversity of character dynamics in this genre.

In the real-life networks, it's important to highlight the low values of Single Character Centrality and Single Relationship Centrality, since these are real-life networks where each person it's living their singular life in parallel rather than novels that specifically focus on a main character or several for the readers enjoyment. Regardless, the number of main characters still identifies that there is often a person with the highest degree not closely followed by others, highlighting the reality of a more "popular" person within each universe that is more connect to the rest.

In summary, by averaging the metrics for each genre, we can see that they exhibit distinct structural features in their character networks, even though the differences are not extremely pronounced. Fantasy novels often center around a single, highly central hero, with lower density and the lowest clustering coefficients, indicating a focus on central characters with less interconnected groups. Mystery novels have smaller but still moderately sized, denser networks that highlight significant relationships between specific characters and the presence of bridging characters, maintaining connectivity despite longer average path lengths. Romance novels display the most varied networks, with moderate density and the most assortativity, often including highly connected characters but sometimes isolated ones. This can lead to a disconnected network where the connected parts are dense and connect to nodes with the same degree, reflecting the degree diversity and complexity of relationships in Romance narratives.

Going forwards in the results, PCA was conducted to reduce the dimensionality of the dataset composed of the network metrics for each novel and to identify the main components that explain the variance in the network metrics. Based on tests, to simplify the analysis, reduce redundancy, and focus on broader structural properties of the character networks, we removed “Single Character Centrality” and “Single Relationship Centrality” for the PCA. The values of Average Path Lenth were also excluded due to its infinite results and were instead only used the Average Path Lenth for the largest component. The model loadings were trained exclusively on the novels and then the scores for the real-life networks were separately calculated and then the points added to the graph.

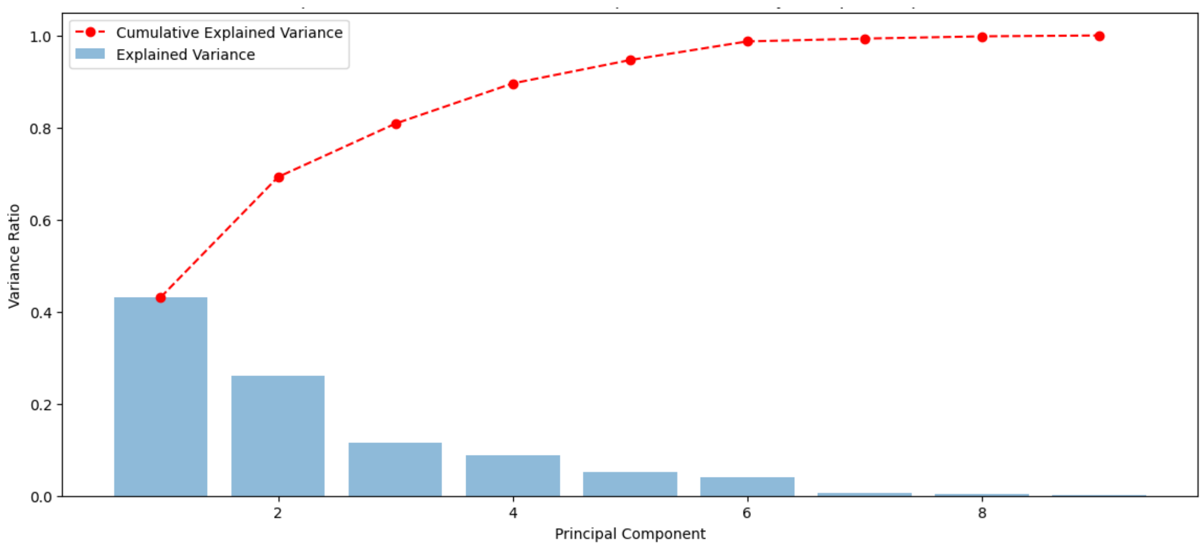

Figure 3 - Variance explained by each principal component and cumulative variance

Figure 3 presents the variance of the dataset explained by each principal component and the cumulative amount of variance in red. Two principal components were chosen to be analyzed because they each have an eigenvalue above 1 and together explain about 70% of the variance, with each PC explaining 43.12% and 26.12% of the variance, respectively. This total variance explained by the two principal components, represents a significant amount of the total variance, indicating that the selected principal components effectively capture the main structural differences in the character networks. The loadings, this is the correlation between a principal component and a variable, estimates the information they share understanding the contributions of each original network metric to each principal component. The loadings for each metric can be found in Table 3.

The first principal component (PC1) is distinguished by positive loadings for network size, number of components, maximum degree, and assortativity, but negative loadings for density, betweenness centrality, number of main characters, and clustering coefficient. This principal component, therefore, represents the contrast between large, sparse networks with few bridging characters and, in contrast, denser, smaller networks with many key connectors and cohesive subgroups. High PC1 values suggest networks with a larger number of components and higher maximum degree, indicating expansive and segmented networks. The a low value for PC1 suggest networks that are less interconnected and rely less on intermediary characters, indicating fewer tightly knit groups and therefore focus on character work rather than individual relationships. So, PC1 reflects the trade-off between overall network size and its connectivity, focusing on relationships between characters in a more expansive context. We called this component: “Network Size and Segmentation”.

Table 3 - Loading of each principal component

Index	PC1	PC2
Network Size	0.9851	-0.0282
Number of Components	0.6693	0.4388
Density	-0.7916	-0.4446
Assortativity	0.3853	0.6802
Maximum Degree	0.4324	-0.7840
Number of Main Characters	-0.3648	0.4413
Average Path Length of Largest Component	0.0257	0.7211
Average Betweenness Centrality	-0.9325	0.1346
Clustering Coefficient	-0.7048	0.3874

The second principal component (PC2) is characterized by high positive loadings for assortativity, number of main characters, average path length, but negative loadings for density and maximum degree. This component captures the structure of the network regarding character similarity and interconnectedness. High PC2 values indicate networks with more central characters, higher assortativity, and longer average paths, suggesting that characters are more spread out within the network but form cohesive subgroups interacting with characters that have the same degree. The positive loading for clustering coefficient supports this, indicating the presence of tightly knit groups. Conversely, negative loadings for maximum degree suggest fewer highly connected nodes and more spread networks. Therefore, PC2 differentiates networks with high assortativity and multiple main characters from those without, capturing the emphasis on main characters and their interconnections. We called this component: “Character Interconnectedness and Similarity” showcasing more a focus towards relationship building rather than character building represented in the other component.

The PCA scores for each novel and real-life network are the int table 4 and the first two Principal Components in a 2D Graph can be seen in figure 4.

Table 4 - PCA scores for each novel

Genre	Novel	PC1	PC2
Fantasy	Limbo	-0,7719	-0,2038
	Oblívio	1,6866	-2,7632
	A Batalha do Apocalipse	2,5381	0,6300
	A Última feiticeira	-0,4358	-1,6043
Mystery	A Célula Adormecida	0,9881	0,9814
	O bairro da Estrela Polar	-0,7824	-0,3656
	A Cidade do Medo	-0,2709	0,2707
	Dias perfeitos	-2,7239	-0,5932
Romance	O Filho de Mil Homens	-3,2480	-0,2761
	O teu rosto sera o último	0,6148	3,4162
	Equador	3,4064	-0,6993
	A Costa dos Murmurios	-1,0011	1,2071
Real-life	Zachary's karate club	-1.3885	-0.9743
	Moreno Highschool	0.1578	0.0844
	Malawi’s village	1.2307	0.8898

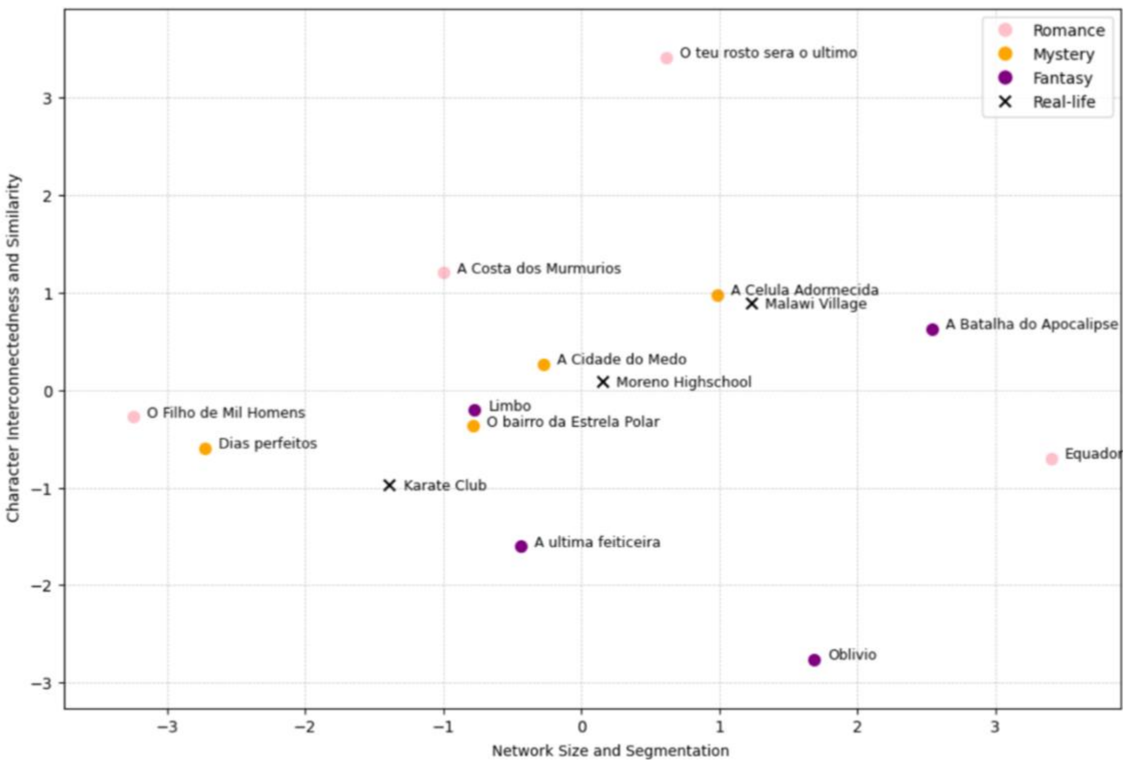

Figure 4 - Novels PC scores represented in 2D for 2 PC

Analyzing the results, it's clear that although each genre generally gravitates together, each genre has no clear and defined group within the PCA results, indicating diversity in narrative structures within every genre under study. This also means that while the genres naturally distinguish themselves by grouping together novels of the same category, they do not clearly separate from those of different genres, indicating differences but not complete separation. In fact, Mystery presents itself as a bridge having its cloud of novels located between Fantasy and Romance, as these last two are more separated. This could imply that the Mystery genre, despite the thematic differences, employs similar narrative strategies in terms of character networks as the other two genres.

Fantasy novels mostly present negative values of “Character Interconnectedness and Similarity” and have more prominent positive values than moderate negatives of “Network Size and Segmentation.” This translates into diverse structural features but a tendency to achieve networks that are more interconnected with fewer main characters interacting with a more diverse set of characters. Regarding network size its balanced, while some focus on larger, more segmented networks with many main characters, others lean towards smaller, denser networks, evidenced by the volatility in scores of “Network Size and Segmentation.” This suggests that this genre is more consensually focused on the creation of character work rather than relationships.

Mystery novels gravitate towards the center of the axis, with most novels showcasing moderate results on both PCs, bridging both Romance and Fantasy. This indicates a balance in the attention towards character work as well as relationships. The central positioning on the PCA plot suggests that Mystery novels maintain a moderate network size and density, with a balanced number of main characters and intermediary connecting characters. This balance is crucial for the genre, as it often involves intricate plots that require a network structure that supports both the development of individual characters and the relationships between them.

The Romance novels present the highest values of “Character Interconnectedness and Similarity”. This means that Romance novels focus on creating networks with multiple central characters, higher assortativity, and disconnected characters with a high average path length, indicating an emphasis on character relationships and interactions. Each Romance novel seems to have its unique value for “Network Size and Segmentation”, ranging the whole scale indicating a wide variability in network size and density. This variability, along with the presence of both highly connected and isolated characters, reflects the complexity and diversity of relationships typically explored in Romance narratives. This structure supports the genre's exploration of various subplots, providing a rich tapestry of character interactions.

In conclusion, the PCA analysis reveals that while each genre exhibits distinct structural characteristics, being Romance and Fantasy the most different in terms of their character network structures. Romance novels focus on creating intricate, interconnected networks with multiple central characters, while Fantasy novels are more variable, ranging from highly interconnected networks to larger, more segmented ones. Mystery novels, acting as a bridge and a compromise of the structure of Romance and Fantasy, showing balanced network structures that closely mimic real-life social networks.

5. CONCLUSIONS AND FUTURE WORKS

Genre provides an important frame of reference that helps readers identify, select, and interpret texts (Chandler, 1997). This work explored the structural differences among genres in Portuguese texts through their character networks. By combining averaged network metrics, Principal Component Analysis (PCA) and the inclusion of real-life networks, we can draw comprehensive conclusions about how these genres differ and relate to one another.

The averaged network metrics highlight some structural features within each genre. Fantasy novels are characterized by a single, highly central hero with high maximum degree and moderate levels of density and clustering, indicating a balanced interconnectedness. Mystery novels feature moderately sized, denser networks that highlight significant relationships between characters and the presence of bridging characters, maintaining connectivity. Romance novels display the most variable networks, with moderate density and high maximum degree, but also show tendencies for isolated characters, leading to disconnected networks. Lastly, real-life networks do not present any centrality on a particular character nor in a relationship.

The Principal Component Analysis results support these findings by showing no clear clustering of genres, suggesting that structural differences in character networks are not strongly tied to genre. The disperse in the PCA results within each genre, indicates diversity in narrative structures. While the genres naturally distinguish themselves by grouping together novels of the same category, they do not clearly separate from those of different genres, indicating differences but not complete separation. Fantasy novels show diverse structural features but tend to achieve networks that are more interconnected with fewer main characters interacting with a more diverse set of characters. Most Mystery novels showcase a moderate balance between character and relationship building, while Romance emphasizes relationship building within its narrative dynamics. Mystery presents itself as a bridge between Fantasy and Romance, suggesting that despite thematic differences, similar narrative strategies are employed in terms of character networks from both genres. The real-life networks stand near each other and near the literary networks sharing common structural characteristics while also exhibiting unique features that reflect their distinct contexts, being this group most similar to Mystery.

Our conclusion is that the similarities within genres and differences between genres are not prominent in graph-specific measures, even though each genre exhibits unique structural characteristics. Fantasy novels are characterized by a focus on a central hero and less interconnected groups, aligning with their narrative emphasis on individual journeys and less relationship building overall. Mystery novels exhibit moderately sized, denser networks with key bridging characters and a balance between relationships and character development, reflecting their intricate plots and are the most similar to real-life networks. Romance novels display the largest and most variable networks, having more than one central character and cohesive subgroups, indicative of their focus on diverse and intricate relationships.

These findings provide valuable insights into the narrative styles of Portuguese-language literature, showing how each genre employs distinct network structures to support their unique storytelling approaches. The lack of pronounced differences could be expected since, as Thelwall & Kousha (2017) point out, genre is a concept that readers, writers, and publishers may interpret uniquely. Booksellers may invent new genres to market books, libraries may categorize books in certain ways to attract readers, and genre types and typical structures can vary immensely in topics and internationally.

In interpreting the findings of this study, it is important to consider the limitations associated with the data and methodology. Firstly, all the interpretations relate to the pipeline and its existing limitations, such as the inability to process stories with multiple points of view and some degree of imprecision in the co-occurrence detection, which is not fully robust to non-Portuguese names. Furthermore, one significant concern is the use of averages to describe network metrics for each genre. While averages provide a useful summary, they can sometimes obscure important variations and nuances within individual works, this is avoided by completing the information with the standard deviation. Another limitation, this related to the data, is the underdevelopment of Portuguese themes in genres such as fantasy or other as SiFi that could be studied, with ebook versions that would allow for a larger dataset. In fact, all the small sample sizes could lead to results that are less reliable and harder to generalize to the entire genre. Future research should aim to include a larger and more diverse sample of novels to validate these findings.

BIBLIOGRAPHICAL REFERENCES

Abdi, H., & Williams, L. J. (2010). Principal component analysis. *WIREs Computational Statistics*, 2(4), 433–459. https://doi.org/10.1002/wics.101

Alberich, R., Miro-Julia, J., & Rossello, F. (2002). Marvel Universe looks almost like a real social network (arXiv:cond-mat/0202174). *arXiv*. https://doi.org/10.48550/arXiv.cond-mat/0202174

Amaral, D. O. F., Fonseca, E. B., Lopes, L., & Vieira, R. (n.d.). Comparative Analysis of Portuguese Named Entities Recognition Tools. Retrieved from https://repositorio.pucrs.br/dspace/bitstream/10923/14041/2/Comparative_Analysis_of_Portuguese_Named_Entities_Recognition_Tools.pdf

Ardanuy, M. C., & Sporleder, C. (2015). Clustering of Novels Represented as Social Networks. *Linguistic Issues in Language Technology*, 12. https://doi.org/10.33011/lilt.v12i.1379

Bento, F., Tagliabue, M., & Sandaker, I. (2020). Complex Systems and Social Behavior: Bridging Social Networks and Behavior Analysis *(pp. 67–91).* https://doi.org/10.1007/978-3-030-45421-0_4

Bhattacharya, S., Sinha, S., Dey, P., Saha, A., Chowdhury, C., & Roy, S. (2023). Chapter 5—Online social-network sensing models. In D. Das, A. K. Kolya, A. Basu, & S. Sarkar (Eds.), Computational Intelligence Applications for Text and Sentiment Data Analysis (pp. 113–140). *Academic Press*. https://doi.org/10.1016/B978-0-32-390535-0.00010-0

Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network Analysis in the Social Sciences. *Science*, 323(5916), 892–895. https://doi.org/10.1126/science.1165821

Borgatti, S. P., Everett, M. G., & Johnson, J. C. (2018). Analyzing Social Networks. *SAGE Publications*. https://books.google.pt/books?id=XD1ADwAAQBAJ

Cambria, E., & White, B. (2014). Jumping NLP Curves: A Review of Natural Language Processing Research [Review Article]. *IEEE Computational Intelligence Magazine*, 9(2), 48–57. https://doi.org/10.1109/MCI.2014.2307227

Canário,T., & Duarte, C. (2024). TAGGUS. *GitHub*. Retrieved July 3, 2024, from https://github.com/ticadu/taggus

Chandler, D. (1997). An introduction to genre theory. http://visual-memory.co.uk/daniel//Documents/intgenre/chandler_genre_theory.pdf

Dekker, N., Kuhn, T., & Erp, M. van. (2019). Evaluating named entity recognition tools for extracting social networks from novels. *PeerJ Computer Science*, 5, e189. https://doi.org/10.7717/peerj-cs.189

dos Reis, C. A. A. (1993). História crítica da literatura portuguesa. *Editorial Verbo*. https://books.google.pt/books?id=G8YQAQAAMAAJ

Earle, T. F., Parkinson, S., & Alonso, C. P. (Eds.). (2013). A companion to Portuguese literature (Vol. 282). *Boydell & Brewer* Ltd. https://books.google.com/books?hl=pt-PT&lr=&id=TNQTAgAAQBAJ&oi=fnd&pg=PP1&dq=portuguese+literature&ots=t812Ai5dnC&sig=1cJBKY0nl0i89YP2OJn0VI1EI7Q

Fan, C., & Li, Y. (2022). Network extraction and analysis of character relationships in Chinese literary works. *Computational Intelligence and Neuroscience*, 2022. https://doi.org/10.1155/2022/7295834

Fernández-Peña, R., Ovalle-Perandones, M.-A., Marqués-Sánchez, P., Ortego-Maté, C., & Serrano-Fuentes, N. (2022). The use of social network analysis in social support and care: A systematic scoping review protocol. *Systematic Reviews*, *11*(1), 9. https://doi.org/10.1186/s13643-021-01876-2

Fonseca, B. de P. F. e, Sampaio, R. B., Fonseca, M. V. de A., & Zicker, F. (2016). Co-authorship network analysis in health research: Method and potential use. *Health Research Policy and Systems*, 14(1), 34. https://doi.org/10.1186/s12961-016-0104-5

Foster, J. (n.d.). Examining mystery, science fiction, romance, and horror: A look at popular fiction genres. *Jamie Foster Science.* Retrieved June 19, 2024, from https://www.jamiefosterscience.com/mystery-science-fiction-romance-and-horror-are-all-considered/

Fowler, A. (1982). Kinds of literature: An introduction to the theory of genres and modes. *Cambridge*, Mass. : Harvard University Press. http://archive.org/details/kindsofliteratur0000fowl

Freeman, L. C. (2004). The Development of Social Network Analysis: A Study in the Sociology of Science. *Empirical Press*. https://books.google.pt/books?id=VcxqQgAACAAJ

Fronzetti Colladon, A., & Naldi, M. (2019). Predicting the performance of TV series through textual and network analysis: The case of *Big Bang Theory*. *PLOS ONE*, 14(11), e0225306. https://doi.org/10.1371/journal.pone.0225306

Gessey-Jones, T., Connaughton, C., Dunbar, R., Kenna, R., MacCarron, P., O'Conchobhair, C., & Yose, J. (2020). Narrative structure of A Song of Ice and Fire creates a fictional world with realistic measures of social complexity. *Proceedings of the National Academy of Sciences*, 117(46), 28582–28588. https://doi.org/10.1073/pnas.2006465117

Hansen, D. L., Shneiderman, B., Smith, M. A., & Himelboim, I. (2020). Chapter 6—Calculating and visualizing network metrics. In D. L. Hansen, B. Shneiderman, M. A. Smith, & I.

Himelboim (Eds.), *Analyzing Social Media Networks with* NodeXL (Second Edition) (pp. 79–94). Morgan Kaufmann. https://doi.org/10.1016/B978-0-12-817756-3.00006-6

Hettinger, L., Becker, M., Reger, I., Jannidis, F., & Hotho, A. (2015). Genre classification on German novels. *In 2015 26th International Workshop on Database and Expert Systems Applications (DEXA)* (pp. 249-253). IEEE. https://doi.org/10.1109/DEXA.2015.62

Isinkaye, F. O., Folajimi, Y. O., & Ojokoh, B. A. (2015*).* Recommendation systems: Principles, methods and evaluation*. Egyptian Informatics Journal*, 16(3), 261–273. https://doi.org/10.1016/j.eij.2015.06.005

Jarynowski, A., & Boland, S. (2016). Social Networks Analysis in Discovering the Narrative Structure of Literary Fiction. https://doi.org/10.48550/arXiv.1608.05982

Kim, K.-R., & Kim, J. (2012). Recommender system design using movie genre similarity and preferred genres in SmartPhone*. Multimedia Tools and Applications* - MTA, 61, 1–18. https://doi.org/10.1007/s11042-011-0728-y

Labatut, V., & Bost, X. (2020). Extraction and Analysis of Fictional Character Networks: A Surve*y. ACM Computing Surveys*, *52*(5), 1–40. https://doi.org/10.1145/3344548

Lourenço, R.F. (2014) Guia do Colaborador [v1.01], 16. https://projectoadamastor.org/wp-content/uploads/2014/09/Guia-do-Colaborador.docx

Mamede, N., & Chaleira, P. (2004). Character Identification in Children Stories. In J. L. Vicedo, P. Martínez-Barco, R. Muńoz, & M. Saiz Noeda (Eds.), Advances in Natural Language Processing (Vol. 3230, pp. 82–90). *Springer Berlin Heidelberg*. https://doi.org/10.1007/978-3-540-30228-5_8

Massey, S. E. (2016). Social network analysis of the biblical Moses. *Applied Network Science*, 1(1), Article 1. https://doi.org/10.1007/s41109-016-0012-1

Mata, A. S. da. (2020). Complex Networks: A Mini-review. *Brazilian Journal of Physics, 50(5), 658–672*. https://doi.org/10.1007/s13538-020-00772-9

Maulud, D. H., Zeebaree, S. R. M., Jacksi, K., Sadeeq, M. A. M., & Sharif, K. H. (2021). State of Art for Semantic Analysis of Natural Language Processing. *Qubahan Academic Journal, 1(2), Article 2*. https://doi.org/10.48161/qaj.v1n2a44

Middha, K., Munzir, M., Goja, S., & Choudhary, S. (2022). Entertainment content recommendation system using machine learning. *International Research Journal of Engineering and Technology* (IRJET), 10(1), 279-284. https://www.irjet.net/archives/V10/i1/IRJET-V10I148.pdf

Mourchid, Y., Renoust, B., Roupin, O., Văn, L., Cherifi, H., & Hassouni, M. E. (2019). Movienet: A movie multilayer network model using visual and textual semantic cues. *Applied Network Science*, 4(1), Article 1. https://doi.org/10.1007/s41109-019-0226-0

Network Density—An overview. *ScienceDirect Topics*. Retrieved June 20, 2024, from https://www.sciencedirect.com/topics/computer-science/network-density

Newman, M. (2010). Networks: An Introduction. *OUP Oxford*. https://books.google.pt/books?id=LrFaU4XCsUoC

Norman, M., Jones, C., Watson, K., & Previdelli, R. L. (2023). Social Network Analysis as a Tool in the Care and Wellbeing of Zoo Animals: A Case Study of a Family Group of Black Lemurs (Eulemur macaco). *Animals*, 13(22), Article 22. https://doi.org/10.3390/ani13223501

Otte, E., & Rousseau, R. (2002). Social network analysis: A powerful strategy, also for the information sciences. *Journal of Information Science*. 28 (6): 441–453. https://doi.org/10.1177/016555150202800601

Ozella, L., Paolotti, D., Lichand, G., Rodríguez, J. P., Haenni, S., Phuka, J., Leal-Neto, O. B., & Cattuto, C. (2021). Using wearable proximity sensors to characterize social contact patterns in a village of rural Malawi. *EPJ Data Science*, 10(1), Article 1. https://doi.org/10.1140/epjds/s13688-021-00302-w

Park, S.-B., Kim, Y.-W., Uddin, M. N., & Jo, G.-S. (2009). Character-Net: Character Network Analysis from Video. *2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology*, 1, 305–308. https://doi.org/10.1109/WI-IAT.2009.54

Park, G.-M., Kim, S.-H., & Cho, H.-G. (2013). Structural Analysis on Social Network Constructed from Characters in Literature Texts. *Journal of Computers, 8.* https://doi.org/10.4304/jcp.8.9.2442-2447

Rahul, Ayush, Agarwal, D., & Vijay, D. (2021). Genre Classification using Character Networks. *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 216–222. https://doi.org/10.1109/ICICCS51141.2021.9432303

Reddy, S. R., Nalluri, S., Kunisetti, S., Ashok, S., & Venkatesh, B. (2019). Content-based movie recommendation system using genre correlation. *In S. C. Satapathy, V. Bhateja, & H. S. Saini (*Eds.), Smart Intelligent Computing and Applications (Vol. 105, pp. 391-397). Springer. https://doi.org/10.1007/978-981-13-1927-3_42

Rocha, C., Jorge, A., Oliveira, M., Brito, P., Gama, J., & Pimenta, C. (2014). From entity extraction to network analysis: A method and an application to a Portuguese textual source. https://www.fep.up.pt/docentes/cpimenta/textos/pdf/ACM15_Rocha2.pdf

Rossi, R. A., & Ahmed, N. K. (2015). The Network Data Repository with interactive graph analytics and visualization. In *Proceedings of the AAAI Conference on Artificial Intelligence , 29(1).* https://doi.org/10.1609/aaai.v29i1.9277

Silva, M. O., Oliveira, G. P., & Moro, M. M. (2023). Analyzing Character Networks in Portuguese-language Literary Works. *Anais Do Brazilian Workshop on Social Network Analysis and Mining* (BraSNAM), 115–126. https://doi.org/10.5753/brasnam.2023.230585

Sudhahar, S., Cristianini, N. (2013). Automated analysis of narrative content for digital humanities. *International Journal of Advanced Computer Science*. 3. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=58e9de78dc9c5348339e83f774c689bbf7d9c45a

Sugishita, K., & Masuda, N. (2023*).* Social network analysis of manga: Similarities to real-world social networks and trends over decades. *Applied Network Science, 8(1), Article 1.* https://doi.org/10.1007/s41109-023-00604-0

Suen, C., Kuenzel, L., & Gil, S. (2013). Extraction and Analysis of Character Interaction Networks From Plays and Movies. *Digital Humanities Conference*. https://www.semanticscholar.org/paper/Extraction-and-Analysis-of-Character-Interaction-Suen-Kuenzel/a57bf7098ea7adccf626caba682f3b4350aa3955

Tamen, M., & Buescu, H. C. (2013). A revisionary history of Portuguese literature. *Routledge*. https://api.taylorfrancis.com/content/books/mono/download?identifierName=doi&identifierValue=10.4324/9780203726877&type=googlepdf

Tantardini, M., Ieva, F., Tajoli, L., & Piccardi, C. (2019). Comparing methods for comparing networks. *Scientific Reports*, 9(1), 17557. https://doi.org/10.1038/s41598-019-53708-y

Thelwall, M., & Kousha, K. (2017). Goodreads: A social network site for book readers*. Journal of the Association for Information Science and Technology*, 68(4), 972–983. https://doi.org/10.1002/asi.23733

Wellek, René., & Warren, A. (1956). Theory of literature. *Harcourt*, Brace & World.

Zachary, W. W. (1977). An Information Flow Model for Conflict and Fission in Small Groups. *Journal of Anthropological Research*, *33(*4), 452–473. https://doi.org/10.1086/jar.33.4.3629752

APPENDIX A

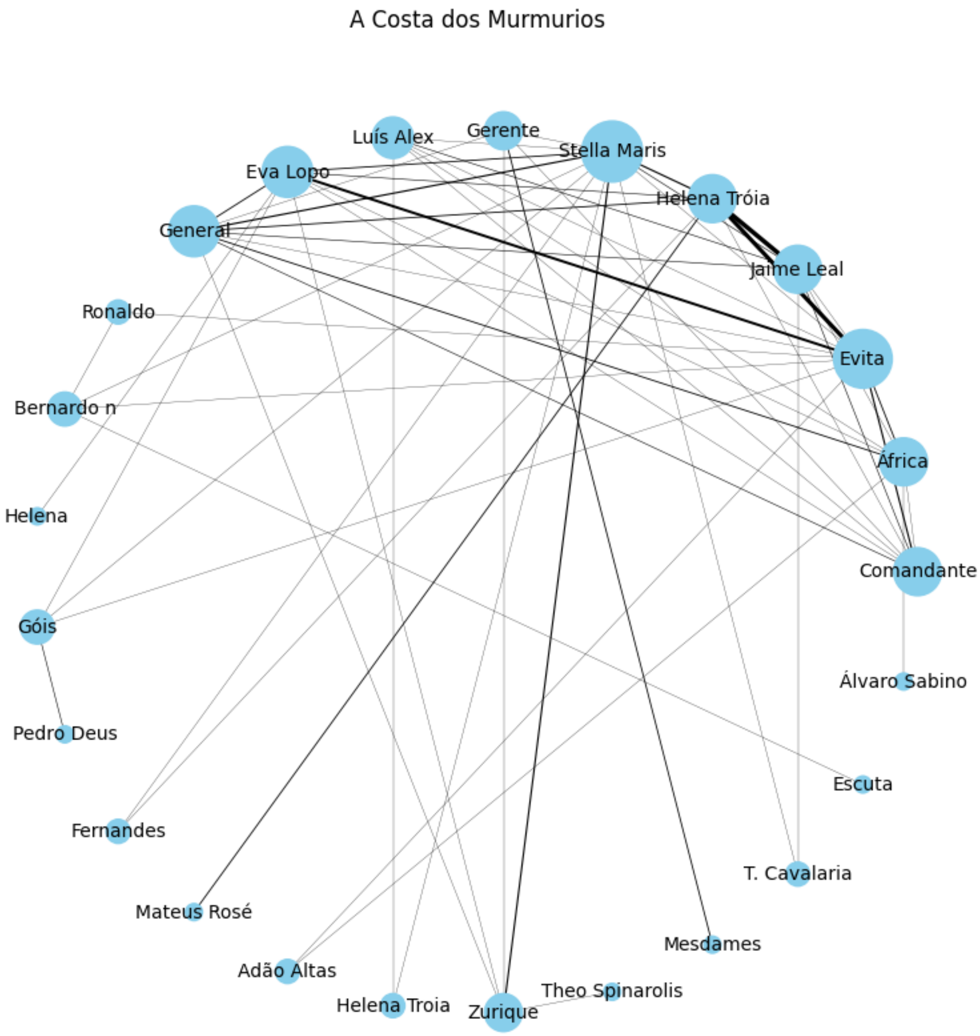

Figure 5 - Character Network for "A Costa dos Murmúrios"

Genre: Romance

Table 5 - Adjacency Matrix for "A Costa dos Murmúrios"

Genre: Romance

	Comandante	África	Evita	Jaime Leal	Helena Tróia	Stella Maris	Gerente	Luís Alex
Comandante	0	1	4	2	1	0	1	1
África	1	0	3	1	0	0	0	1
Evita	4	3	0	1	11	1	0	1
Jaime Leal	2	1	1	0	12	2	0	2
Helena Tróia	1	0	11	12	0	4	0	0
Stella Maris	0	0	1	2	4	0	1	1
Gerente	1	0	0	0	0	1	0	0
Luís Alex	1	1	1	2	0	1	0	0
Eva Lopo	1	1	7	0	2	3	0	0
General	2	3	1	2	3	4	1	0
Ronaldo	0	0	1	0	0	0	0	0
Bernardo n	0	0	1	0	0	1	0	0
Helena	0	0	0	0	0	0	0	0
Góis	0	0	1	0	0	1	0	0
Pedro Deus	0	0	0	0	0	0	0	0
Fernandes	0	0	0	0	1	1	0	0
Mateus Rosé	0	0	0	0	3	0	0	0
Adão Altas	0	1	1	0	0	0	0	0
Helena Troia	0	0	0	0	0	1	0	1
Zurique	0	0	0	0	0	4	1	0
Theo Spinarolis	0	0	0	0	0	0	0	0
Mesdames	0	0	0	0	0	0	3	0
T, Cavalaria	0	0	0	1	0	1	0	0
Escuta	0	0	0	0	0	0	0	0
Álvaro Sabino	0	1	0	0	0	0	0	0

	Eva Lopo	General	Ronaldo	Bernardo n	Helena	Góis	Pedro Deus	Fernandes
Comandante	1	2	0	0	0	0	0	0
África	1	3	0	0	0	0	0	0
Evita	7	1	1	1	0	1	0	0
Jaime Leal	0	2	0	0	0	0	0	0
Helena Tróia	2	3	0	0	0	0	0	1
Stella Maris	3	4	0	1	0	1	0	1
Gerente	0	1	0	0	0	0	0	0
Luís Alex	0	0	0	0	0	0	0	0
Eva Lopo	0	3	0	0	1	1	0	0
General	3	0	0	0	0	0	0	0
Ronaldo	0	0	0	1	0	0	0	0
Bernardo n	0	0	1	0	0	0	0	0
Helena	1	0	0	0	0	0	0	0
Góis	1	0	0	0	0	0	2	0
Pedro Deus	0	0	0	0	0	2	0	0
Fernandes	0	0	0	0	0	0	0	0
Mateus Rosé	0	0	0	0	0	0	0	0
Adão Altas	0	0	0	0	0	0	0	0
Helena Troia	0	0	0	0	0	0	0	0
Zurique	1	1	0	0	0	0	0	0
Theo Spinarolis	0	0	0	0	0	0	0	0
Mesdames	0	0	0	0	0	0	0	0
T, Cavalaria	0	0	0	0	0	0	0	0
Escuta	0	0	0	1	0	0	0	0
Álvaro Sabino	0	0	0	0	0	0	0	0

	Mateus Rosé	Adão Altas	Helena Troia	Zurique	Theo Spinarolis	Mesdames	T, Cavalaria
Comandante	0	0	0	0	0	0	0
África	0	1	0	0	0	0	0
Evita	0	1	0	0	0	0	0
Jaime Leal	0	0	0	0	0	0	1
Helena Tróia	3	0	0	0	0	0	0
Stella Maris	0	0	1	4	0	0	1
Gerente	0	0	0	1	0	3	0
Luís Alex	0	0	1	0	0	0	0
Eva Lopo	0	0	0	1	0	0	0
General	0	0	0	1	0	0	0
Ronaldo	0	0	0	0	0	0	0
Bernardo n	0	0	0	0	0	0	0
Helena	0	0	0	0	0	0	0
Góis	0	0	0	0	0	0	0
Pedro Deus	0	0	0	0	0	0	0
Fernandes	0	0	0	0	0	0	0
Mateus Rosé	0	0	0	0	0	0	0
Adão Altas	0	0	0	0	0	0	0
Helena Troia	0	0	0	0	0	0	0
Zurique	0	0	0	0	1	0	0
Theo Spinarolis	0	0	0	1	0	0	0
Mesdames	0	0	0	0	0	0	0
T, Cavalaria	0	0	0	0	0	0	0
Escuta	0	0	0	0	0	0	0
Álvaro Sabino	0	0	0	0	0	0	0

	Escuta	Álvaro Sabino
Comandante	0	0
África	0	1
Evita	0	0
Jaime Leal	0	0
Helena Tróia	0	0
Stella Maris	0	0
Gerente	0	0
Luís Alex	0	0
Eva Lopo	0	0
General	0	0
Ronaldo	0	0
Bernardo n	1	0
Helena	0	0
Góis	0	0
Pedro Deus	0	0
Fernandes	0	0
Mateus Rosé	0	0
Adão Altas	0	0
Helena Troia	0	0
Zurique	0	0
Theo Spinarolis	0	0
Mesdames	0	0
T, Cavalaria	0	0
Escuta	0	0
Álvaro Sabino	0	0

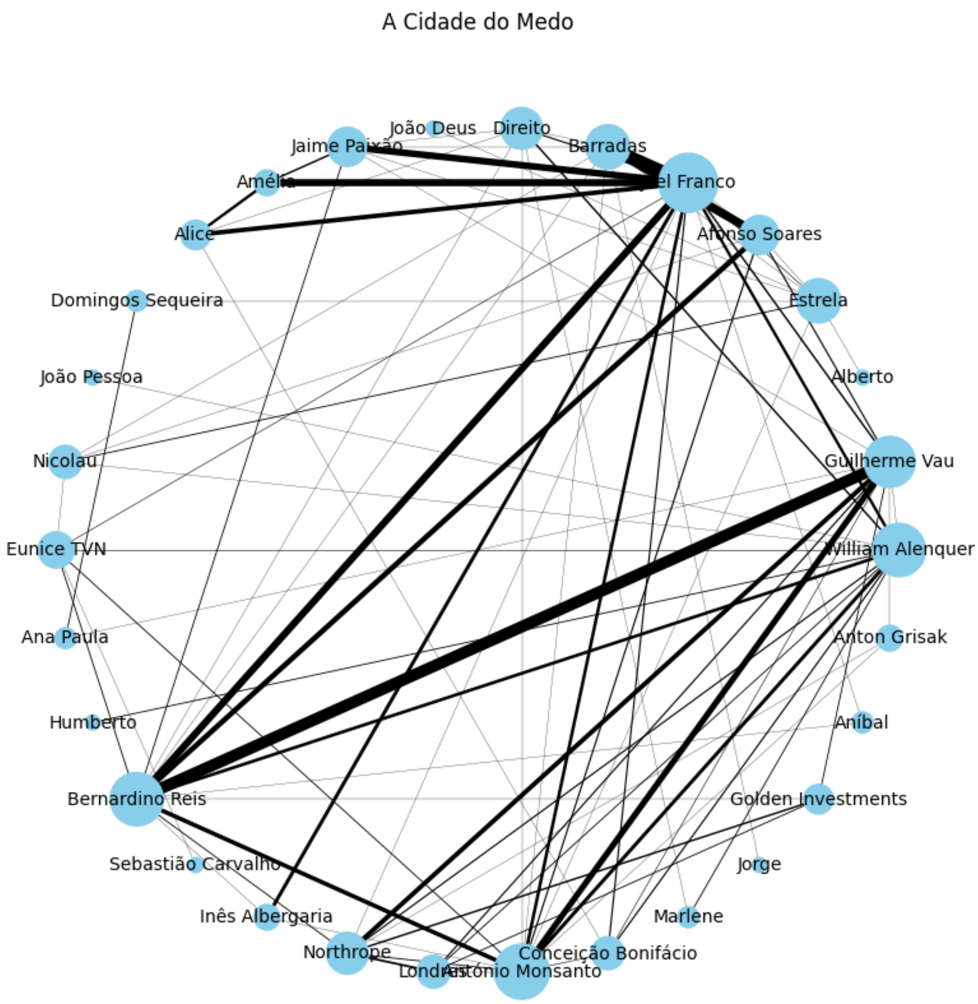

Figure 6 - Character Network for "A Cidade do Medo"

Genre: Mystery

Table 6 - Adjacency Matrix for "A Cidade do Medo"
Genre: Mystery

	William Alenquer	Guilherme Vau	Alberto	Estrela	Afonso Soares	Joel Franco	Barradas	Direito
William Alenquer	0	1	0	0	0	8	0	5
Guilherme Vau	1	0	0	0	4	5	0	0
Alberto	0	0	0	1	0	0	0	0
Estrela	0	0	1	0	1	2	1	0
Afonso Soares	0	4	0	1	0	1	24	0
Joel Franco	8	5	0	2	1	0	34	4
Barradas	0	0	0	1	24	34	0	1
Direito	5	0	0	0	0	4	1	0
João Deus	0	0	0	1	0	0	0	0
Jaime Paixão	0	1	0	1	0	19	1	1
Amélia	0	0	0	0	0	20	0	0
Alice	0	0	0	0	0	13	0	1
Domingos Sequeira	0	0	0	1	0	0	0	0
João Pessoa	1	0	0	0	0	0	0	0
Nicolau	1	0	0	3	1	0	1	0
Eunice TVN	2	0	0	0	0	2	0	0
Ana Paula	0	1	0	0	0	0	0	0
Humberto	3	0	0	0	0	0	0	0
Bernardino Reis	9	33	0	0	15	20	1	1
Sebastião Carvalho	0	0	0	0	0	0	0	0
Inês Albergaria	0	0	0	0	0	10	0	0
Northrope	4	13	0	0	0	1	0	0
Londres	3	4	0	0	0	0	0	0
António Monsanto	10	17	0	1	4	9	1	1
Conceição Bonifácio	4	2	0	0	0	4	0	0
Marlene	3	0	0	0	0	0	0	1
Jorge	0	0	0	0	0	0	1	0
Golden Investments	0	3	0	0	0	0	0	0
Aníbal	0	0	0	0	0	1	0	0
Anton Grisak	0	1	0	0	0	0	0	0

	João Deus	Jaime Paixão	Amélia	Alice	Domingos Sequeira	João Pessoa	Nicolau	Eunice TVN
William Alenquer	0	0	0	0	0	1	1	2
Guilherme Vau	0	1	0	0	0	0	0	0
Alberto	0	0	0	0	0	0	0	0
Estrela	1	1	0	0	1	0	3	0
Afonso Soares	0	0	0	0	0	0	1	0
Joel Franco	0	19	20	13	0	0	0	2
Barradas	0	1	0	0	0	0	1	0
Direito	0	1	0	1	0	0	0	0
João Deus	0	0	0	0	0	0	0	0
Jaime Paixão	0	0	5	0	0	0	0	0
Amélia	0	5	0	8	0	0	0	0
Alice	0	0	8	0	0	0	0	0
Domingos Sequeira	0	0	0	0	0	0	0	0
João Pessoa	0	0	0	0	0	0	0	0
Nicolau	0	0	0	0	0	0	0	1
Eunice TVN	0	0	0	0	0	0	1	0
Ana Paula	0	0	0	0	3	0	0	0
Humberto	0	0	0	0	0	0	0	0
Bernardino Reis	0	3	0	0	0	0	0	3
Sebastião Carvalho	0	0	0	0	0	0	0	1
Inês Albergaria	0	0	0	0	0	0	0	0
Northrope	0	0	0	0	0	0	0	0
Londres	0	0	0	0	0	0	0	0
António Monsanto	0	0	0	0	0	0	0	3
Conceição Bonifácio	0	0	0	1	0	0	0	0
Marlene	0	0	0	0	0	0	0	0
Jorge	0	0	0	0	0	0	0	0
Golden Investments	0	0	0	0	0	0	0	0
Aníbal	0	0	0	0	0	0	0	0
Anton Grisak	0	0	0	0	0	0	0	0

	Ana Paula	Humberto	Bernardino Reis	Sebastião Carvalho	Inês Albergaria	Northrope	Londres
William Alenquer	0	3	9	0	0	4	3
Guilherme Vau	1	0	33	0	0	13	4
Alberto	0	0	0	0	0	0	0
Estrela	0	0	0	0	0	0	0
Afonso Soares	0	0	15	0	0	0	0
Joel Franco	0	0	20	0	10	1	0
Barradas	0	0	1	0	0	0	0
Direito	0	0	1	0	0	0	0
João Deus	0	0	0	0	0	0	0
Jaime Paixão	0	0	3	0	0	0	0
Amélia	0	0	0	0	0	0	0
Alice	0	0	0	0	0	0	0
Domingos Sequeira	3	0	0	0	0	0	0
João Pessoa	0	0	0	0	0	0	0
Nicolau	0	0	0	0	0	0	0
Eunice TVN	0	0	3	1	0	0	0
Ana Paula	0	0	0	0	0	0	0
Humberto	0	0	0	0	0	0	0
Bernardino Reis	0	0	0	0	1	3	0
Sebastião Carvalho	0	0	0	0	0	0	0
Inês Albergaria	0	0	1	0	0	0	0
Northrope	0	0	3	0	0	0	6
Londres	0	0	0	0	0	6	0
António Monsanto	0	0	12	0	1	3	1
Conceição Bonifácio	0	0	0	0	0	0	0
Marlene	0	0	0	0	0	0	0
Jorge	0	0	0	0	0	0	0
Golden Investments	0	0	1	0	0	5	3
Aníbal	0	0	1	0	0	0	0
Anton Grisak	0	0	0	0	0	1	0

	António Monsanto	Conceição Bonifácio	Marlene	Jorge	Golden Investments	Aníbal	Anton Grisak
William Alenquer	10	4	3	0	0	0	0
Guilherme Vau	17	2	0	0	3	0	1
Alberto	0	0	0	0	0	0	0
Estrela	1	0	0	0	0	0	0
Afonso Soares	4	0	0	0	0	0	0
Joel Franco	9	4	0	0	0	1	0
Barradas	1	0	0	1	0	0	0
Direito	1	0	1	0	0	0	0
João Deus	0	0	0	0	0	0	0
Jaime Paixão	0	0	0	0	0	0	0
Amélia	0	0	0	0	0	0	0
Alice	0	1	0	0	0	0	0
Domingos Sequeira	0	0	0	0	0	0	0
João Pessoa	0	0	0	0	0	0	0
Nicolau	0	0	0	0	0	0	0
Eunice TVN	3	0	0	0	0	0	0
Ana Paula	0	0	0	0	0	0	0
Humberto	0	0	0	0	0	0	0
Bernardino Reis	12	0	0	0	1	1	0
Sebastião Carvalho	0	0	0	0	0	0	0
Inês Albergaria	1	0	0	0	0	0	0
Northrope	3	0	0	0	5	0	1
Londres	1	0	0	0	3	0	0
António Monsanto	0	2	0	0	0	0	1
Conceição Bonifácio	2	0	0	0	0	0	0
Marlene	0	0	0	0	0	0	0
Jorge	0	0	0	0	0	0	0
Golden Investments	0	0	0	0	0	0	0
Aníbal	0	0	0	0	0	0	0
Anton Grisak	1	0	0	0	0	0	0

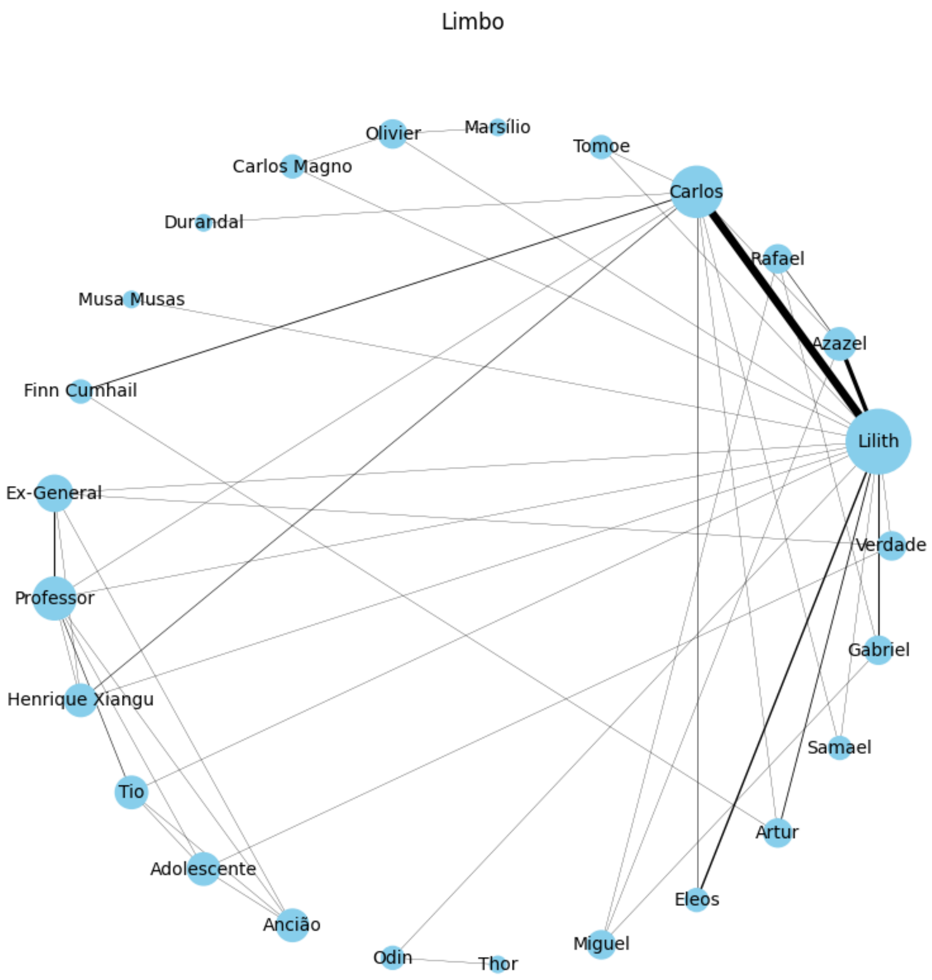

Figure 7 - Character Network for "Limbo"

Genre: Fantasy

Table 7 - Adjacency Matrix for "Limbo"

Genre: Fantasy

	Verdade	Lilith	Azazel	Rafae	Carlos	Tomoe	Marsílio	Olivier	Carlos Magno	Durandal
Verdade	0	1	0	0	0	0	0	0	0	0
Lilith	1	0	12	0	23	1	0	1	1	0
Azazel	0	12	0	2	1	0	0	0	0	0
Rafael	0	0	2	0	0	0	0	0	0	0
Carlos	0	23	1	0	0	1	0	0	0	1
Tomoe	0	1	0	0	1	0	0	0	0	0
Marsílio	0	0	0	0	0	0	0	1	0	0
Olivier	0	1	0	0	0	0	1	0	1	0
Carlos Magno	0	1	0	0	0	0	0	1	0	0
Durandal	0	0	0	0	1	0	0	0	0	0
Musa Musas	0	1	0	0	0	0	0	0	0	0
Finn Cumhail	0	0	0	0	3	0	0	0	0	0
Ex-General	1	1	0	0	0	0	0	0	0	0
Professor	0	1	0	0	1	0	0	0	0	0
Henrique Xiangu	0	1	0	0	2	0	0	0	0	0
Tio	0	1	0	0	0	0	0	0	0	0
Adolescente	1	0	0	0	0	0	0	0	0	0
Ancião	0	0	0	0	0	0	0	0	0	0
Odin	0	1	0	0	0	0	0	0	0	0
Thor	0	0	0	0	0	0	0	0	0	0
Miguel	0	0	1	1	0	0	0	0	0	0
Eleos	0	5	0	0	2	0	0	0	0	0
Artur	0	3	0	0	1	0	0	0	0	0
Samael	0	1	0	0	1	0	0	0	0	0
Gabriel	0	4	0	1	0	0	0	0	0	0

	Musa Musas	Finn Cumhail	Ex-General	Professor	Henrique Xiangu	Tio	Adolescente	Ancião
Verdade	0	0	1	0	0	0	1	0
Lilith	1	0	1	1	1	1	0	0
Azazel	0	0	0	0	0	0	0	0
Rafael	0	0	0	0	0	0	0	0
Carlos	0	3	0	1	2	0	0	0
Tomoe	0	0	0	0	0	0	0	0
Marsílio	0	0	0	0	0	0	0	0
Olivier	0	0	0	0	0	0	0	0
Carlos Magno	0	0	0	0	0	0	0	0
Durandal	0	0	0	0	0	0	0	0
Musa Musas	0	0	0	0	0	0	0	0
Finn Cumhail	0	0	0	0	0	0	0	0
Ex-General	0	0	0	4	1	0	0	1
Professor	0	0	4	0	1	2	1	1
Henrique Xiangu	0	0	1	1	0	0	0	0
Tio	0	0	0	2	0	0	1	1
Adolescente	0	0	0	1	0	1	0	1
Ancião	0	0	1	1	0	1	1	0
Odin	0	0	0	0	0	0	0	0
Thor	0	0	0	0	0	0	0	0
Miguel	0	0	0	0	0	0	0	0
Eleos	0	0	0	0	0	0	0	0
Artur	0	1	0	0	0	0	0	0
Samael	0	0	0	0	0	0	0	0
Gabriel	0	0	0	0	0	0	0	0

	Odin	Thor	Miguel	Eleos	Artur	Samael	Gabriel
Verdade	0	0	0	0	0	0	0
Lilith	1	0	0	5	3	1	4
Azazel	0	0	1	0	0	0	0
Rafael	0	0	1	0	0	0	1
Carlos	0	0	0	2	1	1	0
Tomoe	0	0	0	0	0	0	0
Marsílio	0	0	0	0	0	0	0
Olivier	0	0	0	0	0	0	0
Carlos Magno	0	0	0	0	0	0	0
Durandal	0	0	0	0	0	0	0
Musa Musas	0	0	0	0	0	0	0
Finn Cumhail	0	0	0	0	1	0	0
Ex-General	0	0	0	0	0	0	0
Professor	0	0	0	0	0	0	0
Henrique Xiangu	0	0	0	0	0	0	0
Tio	0	0	0	0	0	0	0
Adolescente	0	0	0	0	0	0	0
Ancião	0	0	0	0	0	0	0
Odin	0	1	0	0	0	0	0
Thor	1	0	0	0	0	0	0
Miguel	0	0	0	0	0	0	1
Eleos	0	0	0	0	0	0	0
Artur	0	0	0	0	0	0	0
Samael	0	0	0	0	0	0	0
Gabriel	0	0	1	0	0	0	0

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa