

Indian Institute Of Technology(ISM), Dhanbad

Final Semester Assignment -Winter

Text Mining

Humanities and Social Sciences

Course Guide : Prof. Shanmugapriya T

Submitted by: Pooja Kumari (24MA0014)

A Distant Reading Approach Using “Text Mining Methods”

Introduction

The Ramayana is one of the most respected and important ancient Indian epics. For many years, it has been studied for its itinerary, religious and cultural values. Written by the sage Valmiki. And the Ramayana which I have taken for my study is translated into English by C. Rajagopalachari. The story talks about big ideas like duty, honor, exile, love, and sacrifice. Women characters play very important roles in the story. Figures like Sita, Kaikeyi, Mandodari and Shabari shape the events of the Ramayana and show different sides of womanhood like devotion, strength, independence and resistance. This research explores how women are represented in the Ramayana and what their portrayal tells us about gender roles, ideals of womanhood and social norms in ancient india. Most studies of the Ramayana so far have focused on reading the text closely and interpreting it using mythology or gender theory. This study uses text mining and digital humanities methods such as keywords in context analysis, topic modeling, word clustering and character network analysis to examine patterns in the text. These techniques help us look at the epic from a broader angle, often called “distant reading”, where we focus on trends instead of only close reading specific passages. Through this research, we aim to highlight both the diversity and limitations in how women are portrayed in the Ramayana. By doing so, we can understand how this ancient text has shaped ideas about women in Indian society.

Literature Review

The Ramayana, attributed to the sage Valmiki, has been widely studied across literature, mythology, cultural studies, and gender studies. Central to many analyses is the portrayal of women, particularly iconic characters like Sita. Scholars such as Dalmia (2001) and Pattanaik (2010) have explored Sita’s trials as reflections of patriarchal expectations, focusing on her purity, sacrifice, and obedience as central to her moral value. These

studies highlight how Sita's narrative aligns with societal constructs surrounding gender roles in ancient India.

Traditional close readings have provided valuable insights, the advent of digital humanities has introduced new methods for analyzing ancient texts. Franco Moretti's concept of "distant reading" (2013) encourages the examination of large-scale patterns and structures in literature.

Recent digital studies have begun applying computational methods to Indian texts. For example, Arora (2020) used topic modeling to analyze gendered language in the Mahabharata, revealing patterns of gender and power. Similarly, Bose (2022) utilized keyword analysis to examine female agency and devotion in various Ramayana retellings. These studies show the potential of computational methods to uncover hidden patterns and challenge traditional gender interpretations.

Recent research uses AI to summarize the *Ramayana* by identifying key events, characters, and locations from the Hindi text. With 70% of users understanding the summaries, the study shows AI's usefulness in making ancient texts more accessible. Though not focused on gender, the methods used—like topic modeling and character mapping—offer a strong base for exploring how women are portrayed in the epic. This approach can support further gender-based analysis of female characters like Sita, Kaikeyi, and Surpanakha.

Despite these advancements, there is a gap in studies using techniques like keyword-in-context (KWIC), clustering, network analysis, and topic modeling to examine female representation in the Ramayana. This research aims to bridge this gap by using a multi-method distant reading approach to analyze the roles and narrative positions of female characters in the epic, offering a new perspective on gender and societal norms in ancient Indian literature.

Research Question : How the representation of women in Ramayana can be studied through “Distant Reading” using text mining techniques like Topic Modeling, Keywords in context, character network analysis and word clustering.

Objectives

- To identify and analyze the representation of women in the Ramayana using distant reading methods.
- To visualize the relationships and clusters that emerge from the analysis of the women characters in the text.

Methodology

1. Keywords in context

- In Ramayana, Sita, Kausalya and Sumitra represent the ideal qualities of a good woman like motherhood and patience. They are praised for their loyalty, strength in suffering and moral values. Sita's hardship in the forest , Kausalya's silent acceptance and Sumitra's calm support show that a woman's goodness is seen in her sacrifice and quiet strength. These characters are respected not because they take action, but they follow their duties as wives and mothers. The story promotes the idea that a woman's value lies in her obedience, devotion to family and ability to control her emotions. Because when Rama left the home then sita also left home and this is not because she wants this is because she follows the duty.
- Urmila and Shabri show how women's sacrifices are often hidden or turned into spiritual acts. They are not in the main story, but their quiet loyalty and devotion are very important. Urmila gives us her Lifewith Lakshman without any complaint and Shabri waits for Rama with deep faith. These women show how female suffering is made into something holy or noble, even though it is often ignored. Their stories remind us that many women in the Ramayana support the main events in silent, unseen ways.
- Kaikeyi and Surpanakha are bold women in the Ramayana who act on their own wishes. Kaikeyi demands boons and Surpanakha expresses desire for Rama. but both are shown in a negative way, as causing problems or breaking moral rules. Their actions lead to blame or punishment. This shows that when women in the epic use their own power, they are often

seen as warnings as examples of what happens when women step outside traditional roles.

- Mandodari and Tara are shown as smart and emotionally strong women in the Ramayana. They understand what is happening around them, especially in politics and family matters. Their sadness and wise words show that they do not fully agree with the violence caused by men. They try to bring peace and think carefully about what is right and wrong. However, their stories are not the main focus of the epic. Even though they are important, they remain in the background, showing how thoughtful and brave women are often not given the main role in the story.
- Shabari, a woman from a lower caste, shows deep love and devotion to Rama in the Ramayana. Her story shows that true devotion or bhakti is more important than social status. Even though she is not from a royal or high family, she is respected for her faith and kindness. This shows that the Ramayana values spiritual love from all people, not just those from that powerful background. Shabari's character gives a message that anyone, no matter their caste or position, can be close to God through sincere devotion.

Women in Ramayana are depicted as embodying loyalty, sacrifice and moral strength, reinforcing traditional gender roles that shape societal expectations of women's behavior and values.

2. Topic modeling

When I applied topic modeling to the *Ramayana*, most of the topics that emerged were centered around men. This shows that the narrative of the *Ramayana* focuses primarily on male characters. Even when I applied topic modeling specifically to the chapters where women are mentioned, many topics emerged, but only 3 to 4 were directly related to women.

Here the result of topic modeling :

Topic 1:

Keywords: *parrying | habitation | frighten | instill | slumbering | souvenir | damage | cracking | unaccountable | bodyguard*

Interpretation : This seems to capture violence, fear, protection and possible threats to space or people. Though not directly about women, it may represent moments where female characters are protected or in danger potentially Sita's abduction, guarded spaces where she lived in the jungle (ashoka vatika) or metaphorical violation of domestic space. Keywords like bodyguard and frighten suggest women as either vulnerable figures or triggers for conflict.

Topic 2:

Keywords: *Dadhimukha | parayana | furnishing | fatiguing | sheltered | keeper | recrossing | assaulted | grateful | declined*

Interpretation : The word sheltered and keeper might suggest domestic or protective roles, possibly relating to female virtue. Assault could indicate threat, again tying back to episodes like Sita's captivity or other gendered violence. There may be subtle hints of women as dependents, or protected within social structure.

Topic 3:

Keywords: *Ravana | hanuman | forest | sugriva | bharata | shall | brother | army | lanka | word*

Interpretation : This is clearly about the conflict between Rama's allies and Ravana, but this is Sita, whose abduction by Ravana triggers the war. She is the symbolic core of this topic even if not mentioned, the actions revolve around her, representing how women are central but often backgrounded in heroic masculine narratives.

Topic 4:

Keywords: *bhagiratha | amsuman | bower | patala | sumati | kapila | rod | whirling | aksha | habitation*

Interpretation : The word bower means a woman's private resting place and Sumati is a woman's name. This topic may be about family history and how important women are in passing on family lines, even though the main story mostly talks about men.

Topic 5:

Keywords: *Parasurama | bamboo | subsist | beehive | vie | abstemiously | unspoilt | twig | topographical | interference*

Interpretation : More abstract, with keywords related to asceticism, nature and conflict. The presence of Parasurama suggests scenes involving confrontation, possibly with Sita's swayamvar. This may reflect how women are won or contested, emphasizing objectification or their symbolic value in masculine conflict.

Topic 6:

Keywords: *Ahalya | purana | lapse | entangled | untenanted | solicitation | reveled | frequent | evoke | eunuch*

Interpretation : This is very relevant. Ahalya is a central figure here, with her story of curse, seduction and redemption. Words like lapse, entangled, solicitation point to themes of female transgression, sexuality and moral judgement. This topic reflects society's scrutiny of women's behaviour and the possibility of forgiveness or spiritual awakening.

Topic 7:

Keywords: *uttara | appreciated | legend | seed | gandhiji | parasurama | tyagaraja | behaves | rebel | drama*

Interpretation : This topic reflects how women are appreciated when they conform and criticized when they rebel. Characters like Kaikeyi and Surpanakha face judgment for

breaking norms. Keywords like “behaves” and “rebel” show how female actions are moralized, reinforcing obedience as virtue and portraying agency as a threat to order.

Justification - The topic modeling analysis of the *Ramayana* reveals a predominantly male-centered narrative, where most themes revolve around male characters, conflicts, and actions. Even when focusing on chapters featuring women, only a few topics directly concern them, and often only in relation to male figures. Women such as Sita, Ahalya, Kaikeyi, and Surpanakha appear primarily in contexts of protection, judgment, captivity, or conflict, not as autonomous agents. Themes related to women emphasize vulnerability, moral scrutiny, and symbolic importance such as Sita’s abduction triggering war—while their voices remain sidelined. This suggests that while women are crucial to the emotional and moral fabric of the epic, their roles are largely passive or symbolic. Topic modeling thus highlights the gendered structure of the text, where women are defined by societal expectations and male action. It offers valuable insight into how traditional narratives reinforce cultural norms and limit female subjectivity within epic storytelling.

3. Clustering method

- **Cluster 3 (Blue/Light) :** *mother*

In this cluster only the word “Mother” came. This suggests it has a special meaning, maybe more spiritual, emotional or respected than just woman or wife. In the *Ramayana*, mothers like Kaushalya, Kaikeyi and sumitra are important and often treated with honor.

Overall, the analysis shows that women in the *Ramayana* are represented in many meaningful ways. They are often described through their beauty and emotions, showing how they are seen as gentle and caring. At the same time, they play important roles in the story’s major events, such as journeys and moments of moral choice—especially in the case of Sita. Women are also shown as central to family life, as mothers, wives, and caretakers. The special focus on the word “mother” highlights the deep respect for motherhood in the story. Together, these clusters show that women in the *Ramayana* are not just background characters—they are emotionally strong, spiritually connected, and play key roles in the story’s values and events.

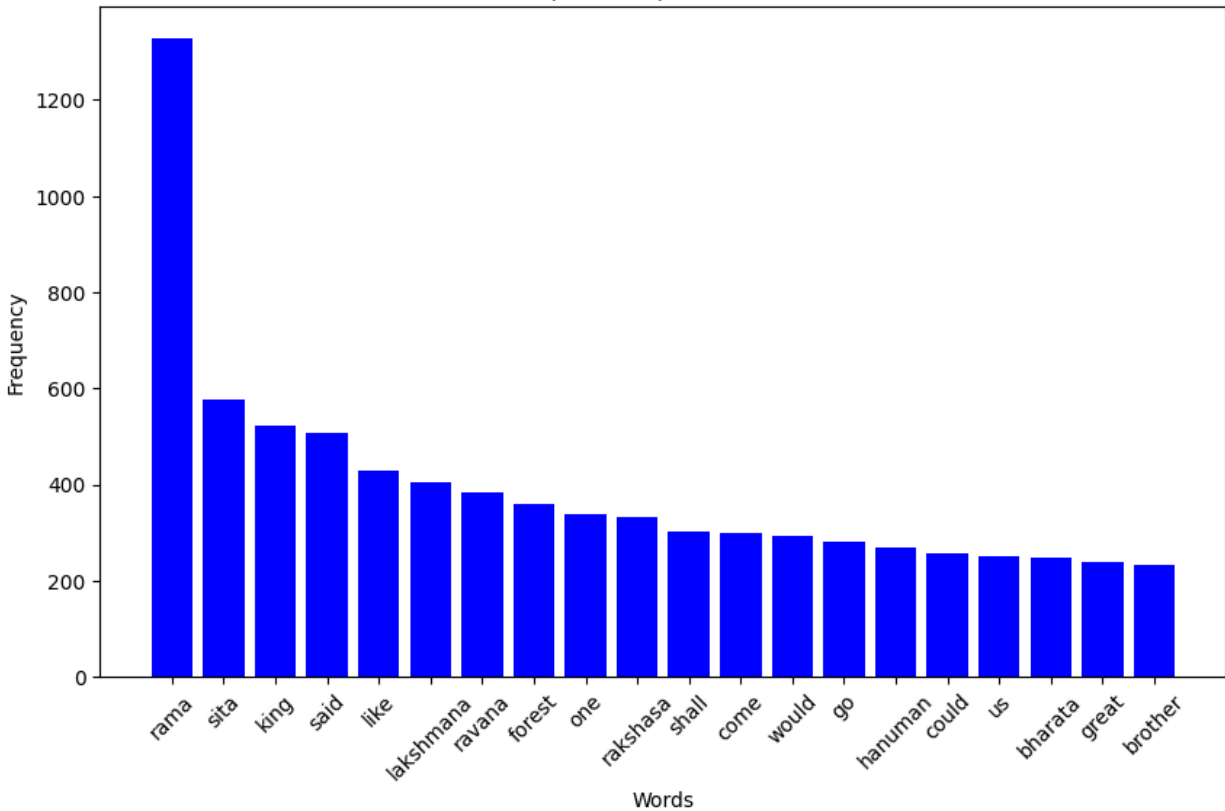
- **Cluster 4 (pink)** : *woman, women, feet, beauty, hand, heart, forest, return, chariot, sage, dharma, people, beautiful, strength, etc.*

Words like beauty, feet, heart, beautiful show that women are often described by how they look or their feelings. The words like return, entered, forest, chariot suggest their movement or journey this could point to Sita's story, her exile, abduction and return. Also words like dharma, sage, rishis connect women with morality and spiritual life. Overall women are shown as emotional, beautiful and part of important journeys and moral lessons in the story.

- **Cluster 5 (Green)** : *brother, child, princes, happy, delay, etc.*

Words like child and princes may relate to women's roles in the family, like being mothers or caretakers. These women's roles connected to family life and relationships.

Top 20 Frequent Words



Word Cloud



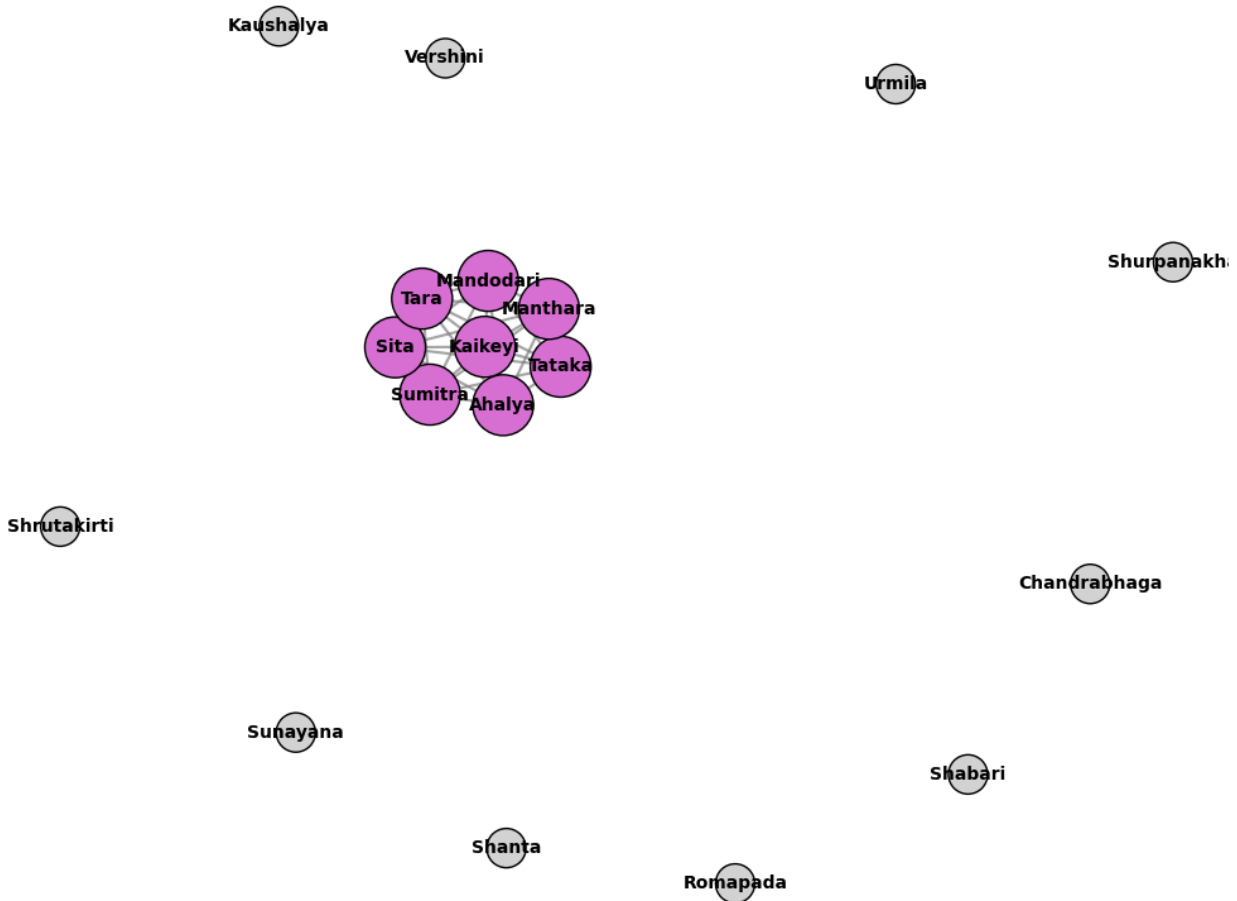
In the word cloud visualization, the prominence of words like “Sita,” “Rama,” “Rakshasa,” “King,” “Brother,” and “Come” highlights the central characters and themes in the Ramayana. Notably, “Sita” appears as one of the largest words, indicating her high frequency and narrative importance. This visual prominence shows that Sita holds significant symbolic and emotional value in the story. Her presence is central to many moral and dramatic tensions, such as her abduction, her role as a devoted wife, and her trials of purity and loyalty. However, when we look at the representation of other female characters like Kaikeyi, Mandodari, and Tara, their names appear much smaller, indicating a lower frequency and reduced narrative presence.

Additionally, the presence of strong action and emotion words such as “kill,” “go,” “destroy,” “fear,” “sorrow,” and “grief” reflects a world shaped by conflict, power, and male-driven actions. In this context, women are often portrayed as passive figures or symbolic representations of ideals like virtue and suffering, rather than as active participants in the plot. Overall, while Sita’s character is given a strong presence, the broader narrative focuses more on male characters and actions, with most other women playing secondary, supportive, or symbolic roles in the story.

Justification : Sita is highly visible in the Ramayana, her prominence is symbolic and emotional rather than action-driven. The word cloud highlights that other women like Kaikeyi, Mandodari, and Tara have much less presence, indicating limited roles. Meanwhile, the frequent use of action words like “kill” and “destroy” points to a male-dominated narrative, where men drive the plot and women are often passive or symbolic. This imbalance in word frequency and thematic focus visually and semantically proves that women in the Ramayana are represented in a selective, secondary, and gendered manner.

5. Character network analysis

Ramayana: Female Character Network (Isolated & Connected)

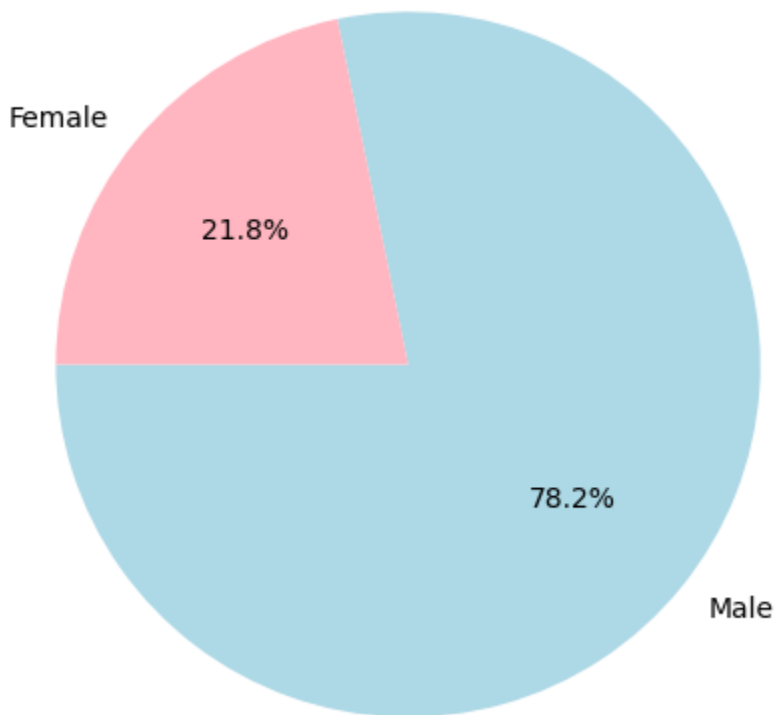


Justification : The female character network in the Ramayana shows that only a few women like Sita, Kaikeyi, and Mandodari are closely connected and important to the story, while many others like Urmila, Shabari, and Shurpanakha are isolated and have little interaction. This means that women are not equally represented — only a small group of royal women are given space and strong roles, while most others stay in the background without much voice or connection. Their roles often depend on men, and they don't act together as a group. This shows that female characters are shown in a limited and uneven way, supporting the idea that the

Ramayana represents women more as individuals tied to male characters than as independent or united figures.

6. Name Entity recognition (NER)

Gender Distribution of Person Names in Ramayana



Through NER , the result is the number of women in ramayana is very less compared to male and also the organisations is mainly focused on male like almost organisations are related to male so , This shows :

- **Fewer women character** : The lower number of female characters in the *Ramayana* reflects the gender norms of the time. Women are often shown in secondary roles, like wives or mothers, while men are the main figures driving the action and events.

- **No Female-led Organizations** : Men are often associated with powerful groups, like kings or warriors leading armies, while women are not linked to any organizations. This shows that women were not seen as leaders or agents of change in the story, reinforcing traditional gender roles.

The story is centered around the male hero, Rama, and his journey, with female characters like Sita playing supportive roles rather than leading or creating change. This focus on men shows the narrative's male-driven structure.

Limitations

While analyzing the *Ramayana* using NER, many character names were not recognized or marked as "unknown" because the model wasn't trained for ancient or literary texts. This made it hard to study gender roles and character connections. In the future, the results can be improved by adding known names, correcting the data by hand, or using a model trained on similar texts.

Some technical challenges while coding, in coding I face some problems but the result is mostly words that are most important but not in the main part.

Future Work

In future work, I want to comparative study how women are shown in the *Ramayana* and the *Mahabharata* using the same text mining methods. This will help us to see the similarities and differences in the roles, speech, and importance of women across both epics. Also want to study different versions of the *Ramayana* like Tulsidas' *Ramcharitmanas*, and the Hindi version of *Ramayana* to understand how the image of women changes based on language, region, or time. This comparison can give deeper insights into cultural, historical, and literary shifts in how women were seen in ancient India.

In the future, I can compare how the environment is shown in other Hindi novels or in literature from different regions or time periods.

AI Help : I took 40% of chat Gpt for coding to understand the coding meaning and To generate the code. And 40 % from the given code by Ma'am .

Conclusion

This research explored how women are represented in the *Ramayana* using digital methods like topic modeling, clustering, word frequency analysis, KWIC, and character network analysis. These tools helped us look at the epic from a broader angle, showing not just individual stories, but larger patterns in how women are talked about and what roles they play.

The results show that the *Ramayana* mostly focuses on male characters, with women appearing less frequently and in limited roles. Sita is the most prominent female figure, but even her presence is often centered around her suffering, devotion, and loyalty to Rama. Other women like Kaikeyi, Mandodari, Tara, and Shabari also play meaningful roles, but they are mentioned much less and have limited impact on the main storyline. Many women characters are shown as emotional, beautiful, or spiritual, and are often praised for being patient, obedient, and supportive of male characters.

Topic modeling and clustering revealed that women are mostly connected with themes like sacrifice, motherhood, moral values, and suffering. When women act with independence, like Kaikeyi or Surpanakha, they are usually judged or blamed. The character network analysis also shows that only a few women are central, while most are isolated or appear briefly.

In short, the *Ramayana* presents women in ways that reflect traditional gender roles of ancient Indian society. Women are valued for their loyalty, devotion, and moral strength, but they are not often shown as independent or active decision-makers. This study helps us understand how an important cultural text like the *Ramayana* shapes ideas about women and their roles. It also shows the power of digital humanities tools in uncovering deeper insights from classical literature that might be missed through close reading alone.

Reference

<https://ceur-ws.org/Vol-2823/Paper1.pdf>

https://www.academia.edu/30007533/Book_Review_Moretti_Franco_2013_Distant_Reading

<https://www.worldscientific.com/doi/full/10.1142/S2196888825500010>

https://colab.research.google.com/drive/1S8GqGrE3dNW1CGn-DvMfBT33B2cJv_RM#scrollTo=j6qWRtBLX1Km