EXPLORING SEMANTIC SHIFTS OF BENEVOLENCE IN CONFUCIANISM WITH NLP TECHNIQUES

Weier Liu
EPFL
weier.liu@epfl.ch

Xi Lei
EPFL
xi.lei@epfl.ch

Yutaka Osaki
EPFL
yutaka.osaki@epfl.ch

ABSTRACT

This study investigates the semantic evolution of the concept of benevolence (仁, Ren) in Confucian philosophy, focusing on its transformations from the Pre-Qin period through the Western Han and Eastern Han dynasties (480 BCE - 192 CE). Utilizing data from the Chinese Text Project, we compiled a dataset of 26 significant Confucian texts, comprising 697 chapters, with 346 specifically discussing benevolence. Our project integrates quantitative distant reading methods from computational linguistics to study semantic changes and qualitative close reading methods to explore the relationship between language and discourse. By employing natural language processing (NLP) techniques, including word embedding, self-similarity analysis, and topic modeling, we traced the shifts in both the denotation and connotation of benevolence. Our findings reveal that while the core meaning of benevolence remained stable, notable variations occurred due to socio-political changes, particularly during the Western Han Dynasty when Confucianism was institutionalized. These results underscore the adaptability of Confucian thought and its enduring relevance through significant historical transitions. Detailed codes can be found in the associated GitHub repository [1] .

1. INTRODUCTION

As the essence and core of Confucian philosophical thought, the concept of 仁 (Ren), usually translated as "benevolence," has been a pivotal idea and virtue to be pursued throughout over two millennia of Chinese history, from the feudal era to the present day. Despite its origins in the tumultuous Warring States period of the Pre-Qin era, Confucian discussions emphasizing moral conscience and personal cultivation with benevolence as a guiding moral, have endured through centuries of scholarly discourse. This enduring relevance highlights the concept's continued capacity to provide meaningful guidance for modern life.

The origins and evolution of benevolence have been extensively discussed and studied by scholars. However, these research often concentrate on the works of individual philosophers such as Confucius, Mencius, and Xunzi, focusing on their specific attitudes towards benevolence.

There has been limited exploration of how the concept of benevolence has transformed over extended historical periods.

To address this gap, our study aims to identify and explore the shifts in the concept of benevolence during its formative stages. This investigation employs computational methods to analyze chapters mentioning benevolence in Confucian texts from the Pre-Qin, Western Han, and Eastern Han periods. By examining both the denotation and connotation of benevolence, we seek to understand its semantic evolution over time.

Our study utilizes data from the Chinese Text Project, accessible online at https://ctext.org/, an open-access digital library of pre-modern Chinese texts. We created a dataset of 26 significant Confucian texts from the Pre-Qin, Western Han, and Eastern Han dynasties (480 BCE - 192 CE), representing the core of Confucian thought. The dataset includes 697 chapters, with 346 chapters directly discussing benevolence, and is categorized by dynasty. This classification allows for a comprehensive analysis of the semantic shifts in the concept of benevolence across different historical periods.

We employed natural language processing (NLP) techniques to study our corpus, aiming to integrate computational distant reading methods with social and historical context. Utilizing self-similarity analysis, topic modeling, co-occurrence analysis, and keywords distance calculation, we obtained digital representations of the semantic information of benevolence, enabling us to analyze its changes over time. Our approaches allowed us to correlate semantic changes with specific texts and historical events, considering factors such as political shifts and academic developments, thereby enriching our research context.

However, several limitations must be acknowledged. Our dataset is limited to classical Confucian texts, not fully capturing the broader discourse of the periods studied. Expanding the dataset could enhance our analysis. Additionally, our findings are constrained by our current philosophical and historical knowledge; collaboration with experts could provide a more comprehensive understanding. Lastly, variations in random seeds and the significant computational resources required may affect result consistency.

2. PREVIOUS WORK

2.1 Study of Benevolence

仁 (Ren), often translated as benevolence, humanity, or hu-

---

[1] GitHub repository link: https://github.com/WayerLiu/DH412_History-and-Digital

maneness, is the foremost of the five virtues in Confucian culture: Ren (benevolence), Yi (righteousness), Li (propriety), Zhi (wisdom), and Xin (trustworthiness). It is considered the thread that runs through all other virtues. The various translations of benevolence highlight its rich content and wide extension as a crucial concept in Confucianism [1]. At its core, benevolence deals with interpersonal relationships, encompassing love and assistance to others. In practical applications within daily life, benevolence extends further, exemplified in different roles: as a courtier, it manifests as loyalty to the emperor; as a son, it is expressed as filial piety towards parents; and as an individual, it involves magnanimity towards others [2].

After Confucius, philosophers within the Confucian school, including Mencius and Xunzi [3], began exploring the political implications of benevolence. Their discussions filled benevolence with a richer political significance. This trend became particularly pronounced during the Western Han Dynasty when the government adopted the policy of "rejecting the other schools of thought and respecting only Confucianism," thereby establishing Confucianism as the official state ideology [4].

During this period, Dong Zhongshu's interpretation of benevolence underscored its political dimension. Dong posited that benevolence is an essential attribute of heaven, and as such, it possesses an external, transcendent origin. According to his doctrine of the "divine rights of emperors," rulers and social elites were endowed with extraordinary wisdom from birth, serving as the representatives and embodiment of heaven [5]. This theory not only provided a metaphysical foundation for the legitimacy of ancient monarchies but also emphasized the political utility of benevolence in maintaining governance and social order.

Current research on the philosophy of benevolence predominantly focuses on specific philosophers such as Confucius, Mencius, and Xunzi, and their perspectives on benevolence [6]. However, there is a notable research gap in the comprehensive analysis of the evolution of the concept of benevolence across different historical periods. Existing studies often isolate individual philosophers without considering the broader historical and cultural transformations of benevolence.

2.2 Investigating Semantic Change Using NLP Techniques in the Humanities

The meanings and usages of words can evolve due to sociocultural forces [7]. Beyond traditional theoretical research, there is a growing trend of employing computational methods to analyze semantic change. For example, researchers may train word embeddings [8] on various corpora or consider the neighbors of each word to detect changes in meaning over time [9]. However, while the emergence of vast amounts of digital data has spurred an increase in natural language processing (NLP) studies focused on language and meaning, many of these studies tend to emphasize demonstrating that changes have occurred rather than uncovering the underlying reasons for these changes [10].

To address this gap, scholars have introduced the discourse-historical approach (DHA) [11] into their research, particularly when using diachronic corpora. By focusing on discourse, it adds sociohistorical and sociolinguistic depth to computational analyses [12]. In [?], the authors emphasize the importance of considering both the intension and extension of a concept in their study on semantic shift. Integrating this perspective with the DHA allows for a more comprehensive understanding of how concepts evolve over time.

In our study, we aim to investigate the evolution of ancient Chinese philosophical concept benevolence by integrating the quantitative findings with a historical interpretation.

In our study, we aim to leverage various NLP methods to investigate the evolution of the ancient Chinese philosophical concept of benevolence, integrating the quantitative findings with historical interpretation. We will examine both the literal meaning (denotation) and the extension meaning (connotation) of benevolence over time, providing a comprehensive understanding of its semantic and cultural evolution.

3. DATA

This study utilizes data sourced from the Chinese Text Project, an online open-access digital library containing pre-modern Chinese texts [2]. We used the Python API provided by the Chinese Text Project to retrieve and create our own dataset, which comprises 26 significant Confucian texts from the Pre-Qin, Western Han, and Eastern Han dynasties (480 BCE - 192 CE). These classics are the crystallized theories of the most important and influential Confucian scholars of these centuries, representing Confucian thought of that period. We used chapters, rather than whole books or individual sentences, as the unit of analysis. Overall, the dataset includes 697 chapters, of which 346 chapters discuss benevolence.

The dataset provides metadata for each book, including the book title, approximate date of composition, and author. We calculated the average of the composition date range to determine the book's composition date and subsequently classified the books by dynasty:

- **1st Period: Pre-Qin** (202 BCE and before): 6 books, including 139 chapters.

- **2nd Period: Western Han Dynasty** (202 BCE - 25 CE): 9 books, including 293 chapters.

- **3rd Period: Eastern Han Dynasty** (25 CE and after): 11 books, including 265 chapters.

It can be seen that although the number of books as a whole increases with the dynasty, the Pre-Qin period only includes the most famous texts, such as the Analects, Mencius, and the Book of Rites. When divided into chapters, the number of chapters during the Western Han and Eastern Han periods is similar.

4. METHODS

In this section, we describe the methods we used to process and analyse our data.

4.1 Pre-Processing

4.1.1 Word Segmentation

In processing Chinese texts, word segmentation is an essential step due to the nature of the Chinese language, where the basic unit is the character rather than the word, and the boundaries between words are not always clear. Therefore, accurate word segmentation is crucial for further analysis. For our study, we utilized a model based on Roberta [3], pre-trained on a large corpus of Classical Chinese texts. This model is capable of precisely parsing the text and performing accurate word segmentation [13].

4.1.2 Word Embedding

Word embedding is a fundamental technique in NLP that maps words or phrases from a vocabulary to vectors of real numbers [14]. By capturing semantic relationships, word embeddings allow us to analyze and compare the meanings of words in a mathematically precise way. In our study, we utilized a model [4] pre-trained on a large corpus of Classical Chinese texts to generate embedding [15]. It is particularly well-suited for our analysis of ancient Chinese philosophical concepts.

We processed the chapters from the three historical periods (Pre-Qin, Western Han, and Eastern Han) by treating them as separate sub-corpora. Each sub-corpus was input into the model to obtain embeddings for each chapter, generating dense vector representations that encapsulate the semantic content. From these chapter embeddings, we derived the embeddings for individual words, enabling us to analyze the semantic shifts of specific terms across different periods.

Additionally, we processed the entire corpus as a whole to obtain embeddings from an overall aspect. Comparing these with the sub-corpus embeddings revealed significant differences in contextual meanings. This shows that the broader context provided by the entire dataset influences the semantic representation of words within each chapter.

4.2 Self-Similarity Analysis

To examine the denotation of benevolence, we utilize the word embedding vector in each chapter, with method in section 4.1.2, to represent its semantic value within the word space and employ self-similarity analysis. This allows us to analyze shifts in denotation by observing changes in the word embedding vectors of benevolence over time.

The self-similarity matrix is constructed to measure the similarity between word embeddings of benevolence across different chapters. For a set of word embedding vectors $\{\nu_i\}$, where i indexes the chapters in temporal order, the similarity between any two vectors ν_i and ν_j is computed using the cosine similarity formula:

$$\mathrm{sim}(\nu_i, \nu_j) = \frac{\nu_i \cdot \nu_j}{\|\nu_i\|\|\nu_j\|} \tag{1}$$

where $\nu_i \cdot \nu_j$ denotes the dot product of the vectors ν_i and ν_j, and $\|\nu_i\|$ represents the Euclidean norm of vector ν_i. This similarity measure yields values in the range [-1, 1], where 1 indicates maximum similarity, -1 indicates maximum dissimilarity, and 0 indicates orthogonality.

We computed the self-similarity matrix for two types of corpora. First, we constructed a comprehensive self-similarity matrix using the word embedding vectors of benevolence from all chapters. This matrix provides an overall view of the semantic consistency or change throughout the entire corpus.

Second, we divided the chapters according to historical periods—Pre-Qin, Western Han, and Eastern Han—and computed separate self-similarity matrices for each period. By doing so, we compare the internal semantic structures of benevolence within these distinct temporal segments.

The reason for analyzing different corpora lies in the temporal and contextual differences that could influence word embeddings. Variations in language use, cultural context, and historical events across different periods may lead to differences in word embeddings.

4.3 Study of Connotation Shift

To study the connotation shift of benevolence, we employ several analytical methods. Topic modeling allows us to understand the primary themes within our corpus, while co-occurrence analysis helps reveal trends and popular topics associated with benevolence. Additionally, calculating the cosine similarity between keywords and benevolence provides a more mathematical approach to examining the relationships.

4.3.1 Topic modeling

To get a general understanding of what aspects of concepts are involved in the connotation of benevolence, we employed topic modeling to analyze our corpus. BERTopic [16] is a topic modeling method based on embeddings and clustering. It uses pre-trained language models like BERT [17] to transform texts into high-dimensional vector embeddings and then employs clustering algorithms such as HDBSCAN [18] to group these embeddings into topics. Our topic model is based on the word segmentation result mentioned in 4.1.1 and word embedding result mentioned in 4.1.2 to better capture Chinese language characteristics.

Unlike traditional topic modeling methods such as LDA (Latent Dirichlet Allocation) [19], BERTopic does not require a preset number of topics nor pre-set stop words, and can automatically determine the optimal number of topics as well as topic words. Its ability to handle high-dimensional and complex semantic information, makes it more accurate and flexible.

4.3.2 Co-occurrence

Co-occurrence analysis is a method used to investigate the frequency with which specific words or phrases appear together in a text dataset. This analysis helps in understanding the relationships between words and shifts in their meanings. Specifically, it is useful for revealing how words associated with certain topics are used together, thereby uncovering hidden patterns in the text.

To conduct this analysis, we first obtain combinations that include keywords identified through topic modeling. Then, we perform co-occurrence analysis across different time periods to understand the changes in the relationships of keywords around benevolence. By doing so, we can grasp how benevolence is used in different contexts and how these contexts change over time.

4.3.3 Similarity Between Topic Keywords and Benevolence

We determine the similarity between topic keywords and benevolence in three different historical periods: Pre-Qin, Western Han, and Eastern Han, which allows us to trace changes of connotation over periods of historical time.

For each period, we employ keywords identified in topic modeling section 5.2, and obtain their corresponding word embedding vectors. Let $\{\kappa_i\}$ denote the set of keyword embedding vectors for a given period, and let $\nu_{\text{benevolence}}$ represent the word embedding vector of benevolence for the same period. We compute the cosine similarity using the following formula:

$$\text{sim}(\kappa_i, \nu_{\text{benevolence}}) = \frac{\kappa_i \cdot \nu_{\text{benevolence}}}{\|\kappa_i\| \|\nu_{\text{benevolence}}\|} \quad (2)$$

Here, $\kappa_i \cdot \nu_{\text{benevolence}}$ denotes the dot product of the keyword vector κ_i and the benevolence vector $\nu_{\text{benevolence}}$, and $\|\kappa_i\|$ and $\|\nu_{\text{benevolence}}\|$ represent the Euclidean norms of the respective vectors.

By calculating the cosine similarity for each keyword embedding vector relative to the benevolence vector, we can assess the semantic proximity of topic keywords to benevolence within each dynasty. This analysis helps us understand how closely related the topics discussed in each period are to benevolence, revealing shifts in thematic emphasis and semantic associations over time.

5. RESULTS

5.1 Self-Similarity

Figure 1 shows that the similarity scores between word embedding vectors of benevolence are generally high, predominantly exceeds 90%.

Despite the general stability, distinct patterns and areas of change in the concept of benevolence are noticeable. The discourse on benevolence over the span of six centuries can be divided into three primary blocks. Each block concludes with sections of lower similarity, indicated by clustered red regions. We found that the sections marking the end of each block correspond to the terminal phases of different dynasties. After close reading of these sections,

we identified their corresponding texts and historical periods, as annotated in Figure 2.
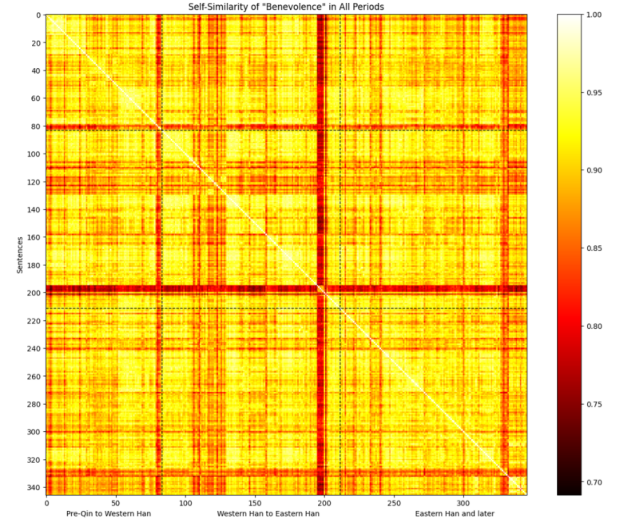

Figure 1. Self-Similarity result for benevolence vector in 3 dynasties.

Given the shared social, cultural, and historical contexts within the same dynasty, we hypothesize that the denotation of benevolence will be more stable compared to the entire corpus. To conduct a more refined analysis, we segmented the corpus by dynasty and utilized more detailed embedding results, focusing on the denotation of benevolence within each dynasty and significant philosophical shifts of benevolence. The results are as follows:

1. After segmenting the corpus, we found that the shift in the denotation of benevolence within each dynasty is more prominently displayed, and the changes appear more significant. Although the overall similarity remains relatively stable, the self-similarity score of benevolence noticeably decreased towards the end of each dynasty.

2. We observed that the similarity score of benevolence also varied in smaller patterns across different books and even within the same book. By filtering out chapters with a similarity score of less than 0.9 in Figures 1 and 2 across different dynasties and examining them closely, we found that many well-known sentences were located in these red-marked areas. Through further close reading of these sentences, we discovered that these chapters often indicate significant shifts in the concept of benevolence throughout the history of Confucianism or represent transitions between different authors. We will elaborate further with concrete examples in the section 6.

5.2 Topic modeling

As shown in the Figure 3, our topic modeling results in three distinct clusters. Table1 presents unique keywords within each cluster along with their English translations. Based on these keywords, we identified and named the themes reflected by the clusters as cluster name.

Since BERTopic primarily relies on word frequency and vector representations to select keywords, it may over-
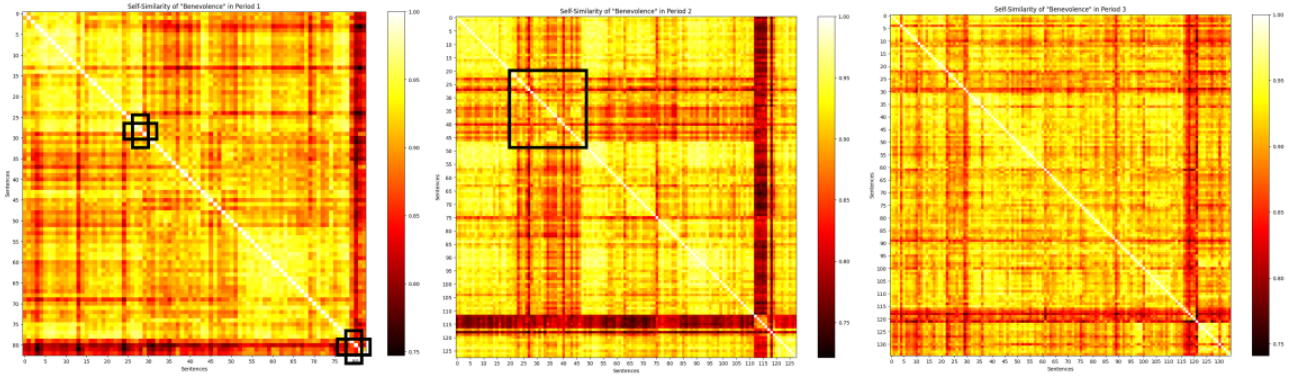
Figure 2. Self-similarity results in each dynasty, namely Pre-Qin, Western Han and Eastern Han.

Cluster index	Cluster name	Keywords	English translation
1	Human ethics and relationship	仁 礼 和	benevolence, rites, harmony
2	Political governance	天 王 命	heaven, king, fate
3	Human life	知 人 有	knowledge, human, be/have

Table 1. Unique keywords in each cluster.

Figure 3. Visualization of BERTopic modeling.

look some infrequent but significant words. Therefore, we referred to a word frequency analysis of Confucian and Daoist classics [20] to supplement the keywords for each cluster. The expanded keyword vocabulary is presented in table 2.

5.3 Co-occurrence

The following co-occurrence results examine the relationship between the keyword benevolence and three topics derived from the topic modeling results in section 5.2.

In Figure 4, the co-occurrence relationship between benevolence and the human ethics and relationship cluster is presented. Before 202 BC, the keywords most frequently co-occurring with benevolence were "rites" and "familial." However, after 25 AD, "sage" emerged as the most co-occurring keyword, which is often representing Confucian in ancient Chinese context.

Figure 5 illustrates the co-occurrence relationship between benevolence and the political governance cluster. The visual representation shows that "politics" and "virtue" were weakly associated with benevolence before 202 BC but became more strongly connected from 202 BC to 25 AD. The term "heaven" consistently maintains a strong association with benevolence across all periods. Conversely, "ruler" was closely linked with benevolence before 202 BC and from 202 BC to 25 AD, but this link weakened after 25 AD. The overall connection between benevolence and this cluster is increasing, and shows strongest connection in West Han dynasty.

lIn Figure 6, the co-occurrence relationship between benevolence and the human life keyword group is presented. The analysis indicates that after 25 AD, the overall color of the paths becomes lighter, yet the co-occurrence is still more frequent than previous two topics. Additionally, the relationship between "righteousness" and benevolence
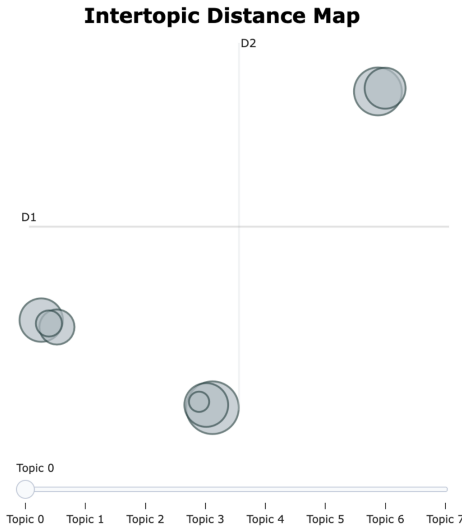
Cluster	Keywords	English translation
1: Human ethics and relationship	贤, 礼, 中, 和, 小, 经, 常, 伦, 孝, 正, 圣, 尊, 亲	virtuous, rites, moderation, harmony, small, classic, constant, ethics, filial piety, righteous, sage, respect, familial
2: Political governance	君, 臣, 天, 命, 道, 德, 父, 子, 妻, 家, 国, 政, 治, 法	ruler, minister, heaven, fate, way, virtue, father, son, wife, family, country, governance, politics, law
3: Human life	善, 恶, 义, 知, 信, 性, 欲, 有, 无, 修, 齐, 贵, 贱, 人, 生	goodness, evil, righteousness, knowledge, trust, nature, desire, have, none, cultivate, regulate, noble, humble, human, life

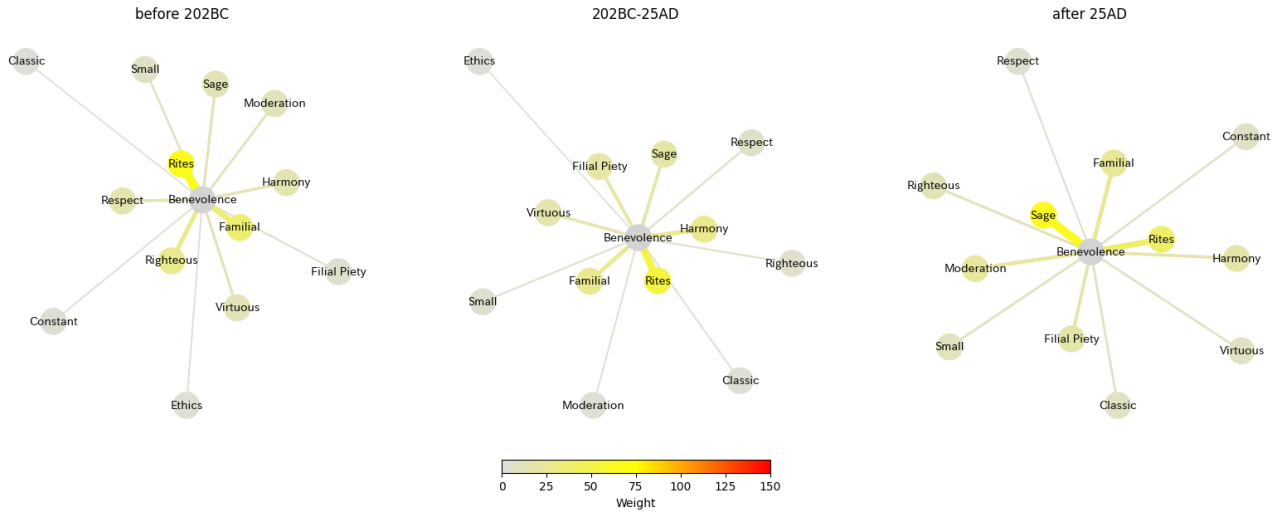**Table 2**. Expanded keyword vocabulary for topic clusters.

Figure 4. Co-occurrence between benevolence and keywords in human ethics cluster.

was strong before 202 BC and from 202 BC to 25 AD, with co-occurrence exceeding 200. However, after 25 AD, this relationship appears to have diminished.

5.4 Similarity Between Topic Keywords and Benevolence

In the process of using cosine similarity to explore the relationship between benevolence and three aspects: human ethics and relationships, political governance, and human life, we obtained the following results:

Human ethics and relationships: Figure 7 indicates that the period from the pre-Qin Dynasty to the Western Han Dynasty saw the highest relevance between keywords related to human ethics and benevolence. Specifically, the keyword "filial piety" had a correlation of 0.619 with benevolence during this period, marking the most significant decrease over time.

Political governance: Figure 8 shows that keywords related to political governance were most frequently discussed during the Western Han and Eastern Han periods. While the degree of discussion slightly decreased during the Eastern Han period, it was still higher than in the Pre-Qin period. Notably, the words "government", "country",

"fate", and "law", which pertain to governing the country, saw a significant rise in discussion during the Western Han Dynasty.

Human life: As shown in Figure **??**, the degree of discussion and the correlation with the core concepts of life have gradually increased. The cosine similarity remains around 0.6. The similarity between terms related to personal development and attitudes towards life, as aligned with Confucian principles, has been increasing.

Overall, among these three topic clusters, the relevance of human ethics and relationships has slightly decreased, while political governance and human life have shown a slight increase. The degree of similarity hovers around 0.6. Employing the bootstrap method to sample and calculate the confidence intervals of cosine similarity further supports this observation, shown in Figure 10.

6. DISCUSSION

6.1 Denotation Shift

The analysis of denotation shifts in the concept of benevolence, as revealed by the self-similarity matrices, provides insights into the evolution of the semantic meaning across
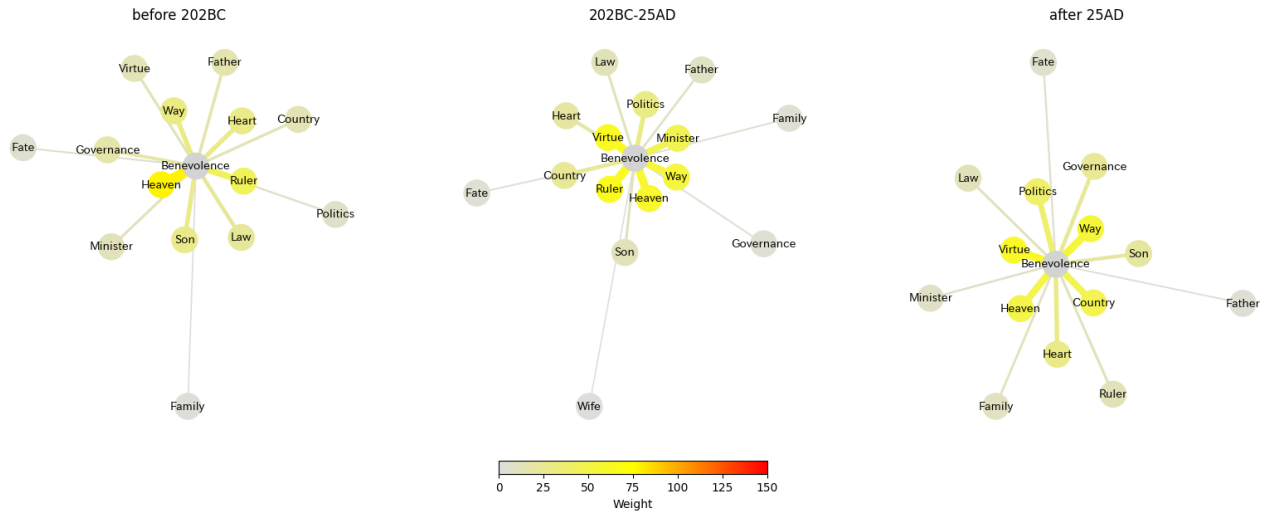
Figure 5. Co-occurrence between benevolence and keywords in political governance cluster.
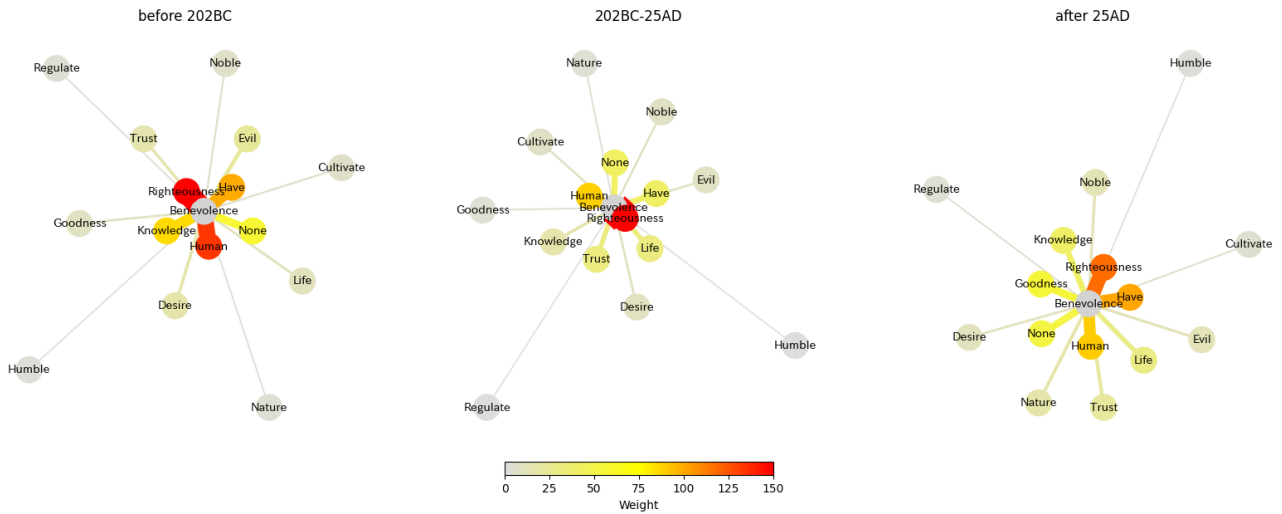

Figure 6. Co-occurrence between benevolence and keywords in human life cluster.

different historical periods. It is important to note that our self-similarity analysis is not aimed at identifying repeating patterns or segments, as one might do in musicology to find choruses or bridges. In the context of our corpus, each column in the self-similarity matrix represents the semantic vector of benevolence, and our objective is to investigate chapters of significant semantic change rather than recurrent thematic elements.

The results of our self-similarity analysis indicate that the denotation of benevolence remains generally consistent across the entire corpus, with similarity scores mostly exceeding 0.9. This high level of similarity suggests a stable semantic interpretation of benevolence throughout the studied periods. However, we observe occasional dips in similarity scores, particularly towards the end of each historical period, where scores drop below 0.8. These drops are indicative of political upheavals or transitions, which likely influenced philosophers' interpretations and discussions of benevolence.

One intriguing section in Figure 2 is marked by the black block in the lower right corner of the first sub-image, highlighting chapters from "Su Shu". Authored by Huang Shigong in the late Qin period, "Su Shu" was esteemed as a sacred text during the Han dynasty. The prominent Han dynasty chancellor Zhang Liang used this book to aid Emperor Liu Bang, the first emperor of the Han dynasty, in defeating his adversary Xiang Yu, thereby laying the foundation of the Han dynasty.

Unlike other Confucian texts, "Su Shu" was not intended for the general populace. Huang Shigong did not aim to educate the common people with this book's content. Instead, it was directed towards rulers and their officials, serving as a guide on how to govern as a ruler, fulfill duties as a minister, manage the state, administer politics, enhance personal cultivation, and ultimately ensure that the populace could live prosperous and content lives.

Consequently, the doctrines advocated and audience in "Su Shu" are markedly different from those in other Confucian classics, which explains why the concept of benevolence in this text diverges significantly from that in other
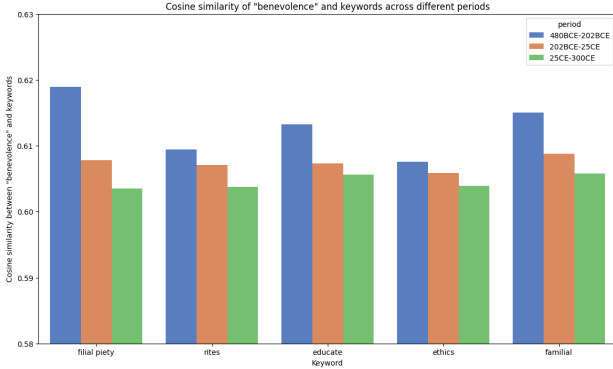
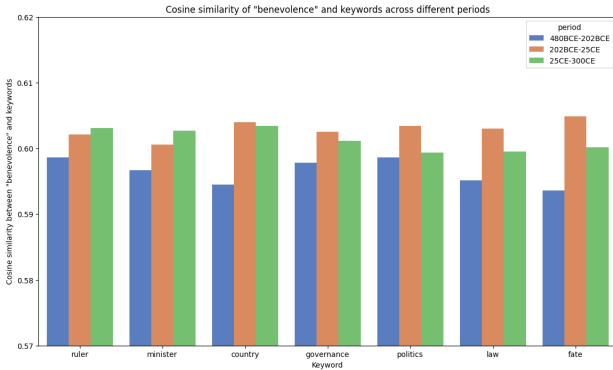**Figure 7**. Similarity score between benevolence and keywords in human ethics and relationship cluster.

Figure 8. Similarity score between benevolence and keywords in political governance cluster.

Figure 9. Similarity score between benevolence and keywords in human life cluster.

Figure 10. Mean similarity score of with error-bar in three periods for each cluster, namely human ethics and relationship, political governance and human life.

works.

When conducting individual studies for each period, we further discovered that the previously mentioned red regions often coincide with significant shifts in the denotation of benevolence in Confucianism or transitions between different authors' works.

For instance, in the self-similarity matrix in the Pre-Qin period in Figure 2, marked by the black block in the upper left corner of the first sub-image corresponds to the text: "Only the utmost sincerity in the world can manage the great principles of the world, establish the great foundation of the world, and understand the transformation and nurturing of heaven and earth." A close reading reveals that this passage is from the "Zhong Yong", one of the Four Books and Five Classics of Confucianism. It is in this chapter that the association between benevolence and sincerity is introduced for the first time within the Confucian framework.

In the self-similarity matrix from the Western Han to the Eastern Han period in Figure 2 marked by the black block, the text corresponds to a sentence from Dong Zhongshu's book "Chun Qiu Fan Lu." In this work, Dong Zhongshu promoted Confucianism as the state ideology during the Han dynasty, embedding it deeply into the political and social fabric of China. The significant changes he introduced are also evident in the text.

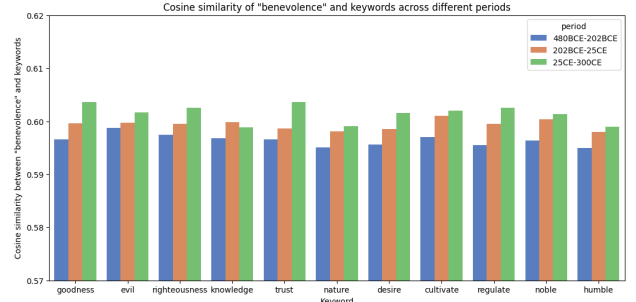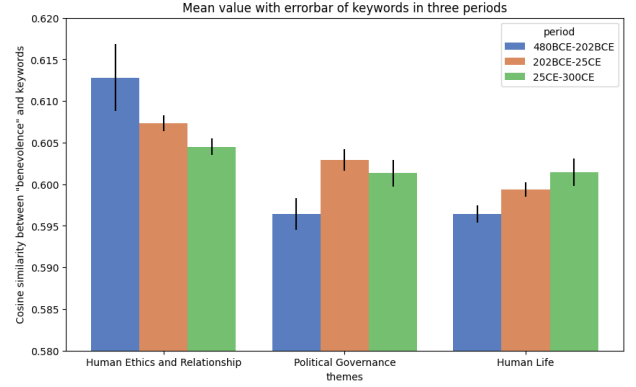Furthermore, the discrepancy between the self-similarity scores in the overall corpus and the individual

Eastern Han corpus requires theoretical explanation. One plausible reason is that when the Eastern Han is analyzed within the broader context, the embedding vectors may reflect an amalgamation of influences from preceding periods, thereby smoothing out some of the variations seen when it is examined in isolation. In contrast, the isolated analysis of one period highlights the internal semantic diversity without the mitigating effect of cross-period influences.

6.2 Topic Cluster Interpretation

In this section, we discuss the philosophical meaning of the topic clusters in table 2.

Cluster 1: Human ethics and relationships. Central to Confucian thought, the theme of Human Ethics encapsulates the ideal of loving and helping others, forming the foundation of a harmonious society. This cluster includes concepts like benevolence, rites, and harmony. Confucian philosophy emphasizes the cultivation of personal virtues and the importance of familial and societal relationships. For instance, Confucius stated, "A benevolent person loves others" and "Filial piety and brotherly respect are the root of benevolence". These principles underscore the moral duties and ethical conduct expected in personal and communal interactions, promoting a cohesive and morally grounded society.

Cluster 2: Political governance. This cluster reflects

the political significance of the concept of benevolence. In Confucian philosophy, benevolence is a transcendent concept, the highest virtue and an aspirational goal. It is more than a moral concept, but also possesses strong political implications. Confucius introduced benevolence with the hope of "taking benevolence as one's responsibility" (*The Analects*), positioning it as the highest ideal, inspiring scholars to undertake the responsibility of governing the world and rebuilding political order.

Furthermore, during the Western Han dynasty's policy of dismissing other schools of thought in favor of Confucianism, the core Confucian value of benevolence became a mandatory moral principle for everyone from the emperor to local officials. This development imbued the concept of benevolence with political significance beyond its philosophical roots.

Cluster 3: Human life. The theme of Human Life promotes the cultivation of personal virtues, the balance of desires, and continuous self-improvement. This cluster includes terms like goodness, knowledge, and life, reflecting the focus on personal development and ethical living. The philosophy encourages individuals to strive for moral excellence, maintain a balance between personal desires and ethical responsibilities, and engage in lifelong learning and self-cultivation.

These topics effectively encapsulate the core values of Confucian philosophy, which ensures that our subsequent analyses are comprehensive and able to reflective of the foundational principles of Confucian thought. By analyzing these themes, we gain structured insights into the essential aspects of Confucian philosophy, providing a solid basis for further exploration of how these values are represented and interconnected with benevolence concept.

6.3 Connotation Shift

The concept benevolence, has undergone significant shifts in connotation across different historical periods, reflecting the changing socio-political and cultural contexts of ancient China.

During the Pre-Qin period, Confucian thought emphasized the role of benevolence in personal ethics and familial relationships. This era was characterized by political fragmentation and social upheaval, leading philosophers like Confucius and Mencius to advocate for moral cultivation and ethical behavior as foundational for social harmony. Benevolence was primarily viewed as a guiding principle for interpersonal conduct and familial duties, stressing virtues such as filial piety and righteousness.

In the Western Han Dynasty, significant changes occurred as Confucianism was adopted as the state ideology. Emperor Wu's policy of "abolishing the hundred schools of thought and exclusively respecting Confucianism" institutionalized Confucian principles within the administrative and educational systems. This period saw an expansion of benevolence to encompass broader ethical responsibilities, including political governance. Benevolence became integral to the concept of virtuous rule, aligning with the idea that a benevolent ruler would ensure the well-being of

the state and its people. This shift reflects the merging of moral philosophy with practical governance, where ethical statecraft was seen as essential for political stability and legitimacy.

The Eastern Han Dynasty continued to build on these developments, but with a notable diversification in the application of benevolence. As Confucianism became deeply entrenched in statecraft and education, its teachings also began to address broader aspects of human life. Philosophers and scholars of this period expanded the discourse on benevolence to include personal development and public morality, reflecting the complex socio-political landscape of the time.

Overall, the evolution of benevolence from a principle of personal and familial ethics to a cornerstone of political ideology and broader human values illustrates the adaptability of Confucian thought. This adaptability allowed Confucianism to remain relevant through significant historical transitions.

7. LIMITATIONS

Despite the insights gained from our study, several limitations must be acknowledged.

First, the size and source of our dataset are limited. Our dataset comprises only the classical Confucian texts, which do not fully capture the discourse of the three periods under study. To ensure the reliability and extend the breadth and depth of our analysis, it would be necessary to include a wider range of discussions and debates by Confucian scholars from these periods.

Second, our discussion is constrained by our current level of philosophical understanding and historical knowledge. While our focus on sections with significant differences, there may be intriguing aspects of the results that remain unanalyzed. Further research could benefit from collaboration with experts in Confucian philosophy and history to combine the analysis with historical facts and events, providing a more comprehensive understanding.

Lastly, different random seeds can yield different results, which may affect the consistency of our findings. Additionally, the production of our results requires substantial computational resources, making it challenging to generate multiple results and compute averages to reduce randomness.

8. CONCLUSION

Our study demonstrates the dynamic nature of the concept of benevolence in Confucian philosophy across different historical periods. The integration of NLP techniques with historical and philosophical analysis enabled us to capture subtle semantic shifts in the concept of benevolence, reflecting broader socio-political changes. The Western Han Dynasty, in particular, marked a significant expansion of the concept to encompass broader ethical responsibilities and political governance, showing the results of institutionalization of Confucian principles. Despite these shifts, the core meaning of benevolence as a guiding moral principle

remained remarkably stable, illustrating the enduring relevance and adaptability of Confucian thought.

Our research, focusing on both the denotation and connotation of benevolence, provides a thorough and comprehensive understanding of its evolution. It is important to note that the NLP techniques employed in our study should not be seen as replacements for close reading strategies but rather as complementary tools. Our methods allow researchers to adopt a macroscopic perspective when dealing with corpora spanning extensive time periods. By combining NLP with the discourse-historical approach, our project demonstrates how close and distant reading methods can be integrated to provide a deeper and more complete understanding of the evolution of key philosophical concepts. Future research could benefit from expanding the dataset and incorporating insights from Confucian philosophy experts to further explore these transformations.

9. REFERENCES

[1] X. Yao, *An introduction to Confucianism*. Cambridge University Press, 2000.

[2] C. Wang and N. Madson, *Inside China's legal system*. Chandos Publishing, 2013.

[3] X. Jin, "The two poles of confucianism: A comparison of the inter-state political philosophies of mencius and xunzi," *Ancient Chinese thought, modern Chinese power*, vol. 172, 2011.

[4] Z. Xiangcai, "Dong zhongshu and confucian economic thought as state ideology in the western han dynasty," in *The Political Economy of the Han Dynasty and its Legacy*. Routledge, 2019, pp. 51–66.

[5] J. Wang, "Dong zhongshu's innovation and development of confucianism," *Journal of Jinan University: Social Sciences*, vol. 19, no. 6, pp. 9–12, 2009.

[6] K. Nielsen, "Background and history of confucianism-a literature review," 2015.

[7] A. Blank and P. Koch, *Historical semantics and cognition*. Walter de Gruyter, 2013, vol. 13.

[8] V. Kulkarni, R. Al-Rfou, B. Perozzi, and S. Skiena, "Statistically significant detection of linguistic change," in *Proceedings of the 24th international conference on world wide web*, 2015, pp. 625–635.

[9] H. Gonen, G. Jawahar, D. Seddah, and Y. Goldberg, "Simple, interpretable and stable method for detecting words with usage change across corpora," *arXiv preprint arXiv:2112.14330*, 2021.

[10] A. Kutuzov, L. Øvrelid, T. Szymanski, and E. Velldal, "Diachronic word embeddings and semantic shifts: a survey," *arXiv preprint arXiv:1806.03537*, 2018.

[11] R. Wodak, "Critical discourse analysis, discourse-historical approach," *The international encyclopedia of language and social interaction*, pp. 1–14, 2015.

[12] L. Viola and J. Verheul, "One hundred years of migration discourse in the times: A discourse-historical word vector space approach to the construction of meaning," *Frontiers in Artificial Intelligence*, vol. 3, p. 64, 2020.

[13] K. Yasuoka, "Universal dependencies treebank of the four books in classical chinese," *International Conference of Digital Archives and Digital Humanities*, pp. 20–28, 2019.

[14] O. Levy and Y. Goldberg, "Neural word embedding as implicit matrix factorization," *Advances in neural information processing systems*, vol. 27, 2014.

[15] P. Wang and Z. Ren, "The uncertainty-based retrieval framework for ancient chinese cws and pos," in *Proceedings of the Second Workshop on Language Technologies for Historical and Ancient Languages*, 2022, pp. 164–168.

[16] M. Grootendorst, "Bertopic: Neural topic modeling with a class-based tf-idf procedure," *arXiv preprint arXiv:2203.05794*, 2022.

[17] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[18] L. McInnes, J. Healy, S. Astels *et al.*, "hdbscan: Hierarchical density based clustering." *J. Open Source Softw.*, vol. 2, no. 11, p. 205, 2017.

[19] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, vol. 3, no. Jan, pp. 993–1022, 2003.

[20] S. Mo, "A comparative study of thoughts in "the analects" and "tao te ching" based on word frequency analysis [d]," Ph.D. dissertation, 2014.