Research Article

Alessandra Luce, Paul Robertson*

Comparing and Assessing Statistical Distance Metrics within the Christian Apostle Paul's Letters

Abstract: This study applies and compares different metrics and statistical methods to Paul's letters and a variety of roughly contemporary texts also written in Greek. The primary objective is to elucidate the distinctions between several notable metrics and statistical methods of comparison to discern the relative impact of their application on our understanding of Paul's letters *vis-à-vis* our comparanda. The secondary objective is to compare the findings from different statistical metrics in order to assess their relative merits as statistical tools for literary analysis. Such analysis extends to determining the statistical significance of different methods' relative findings and extracting meaningful insights about each metric and method. We aim to determine which metrics are most accurate and provide the most useful information in comparing Paul's letters to other types of literature such as Greco-Roman philosophy, formal oratory, and so-called Jewish apocalyptic literature. Such findings are especially salient when comparing these quantitative methods to qualitative conclusions from the biblical commentary traditions of literary analysis and comparison.

Keywords: Paul's letters, digital humanities, comparative literature, computational methods

1 Introduction

This study applies and compares different metrics and statistical methods to Paul's letters and a variety of roughly contemporary texts also written in Greek. The primary objective is to elucidate the distinctions between several notable metrics and statistical methods of comparison to discern the relative impact of their application on our understanding of Paul's letters *vis-à-vis* our comparanda. The secondary objective is to compare the findings from different statistical metrics in order to assess their relative merits as statistical tools for literary analysis. Such analysis extends to determining the statistical significance of different methods' relative findings and extracting meaningful insights about each metric and method. We aim to determine which metrics are most accurate and provide the most useful information in comparing Paul's letters to other types of literature such as Greco-Roman philosophy, formal oratory, and so-called Jewish apocalyptic literature. Such findings are especially salient when comparing these quantitative methods to qualitative conclusions from the biblical commentary traditions of literary analysis and comparison.

By comparing computational methods and statistical techniques, this research enhances our understanding of stylometric analysis – a key approach in biblical and literary studies interested in empirical approaches – by suggesting the most effective applications for particular types of data and research questions,

---

* Corresponding author: Paul Robertson, Department of Classics, Humanities and Italian Studies, University of New Hampshire, Durham, United States; Department of New Testament and Related Literature, University of Pretoria, Pretoria, South Africa, e-mail: Paul.Robertson@unh.edu

Alessandra Luce: Department of Computer Science, University of New Hampshire, Durham, United States

thereby yielding more accurate outcomes.[1] Its interdisciplinary approach bridges computational approaches with traditional biblical and literary scholarship, offering valuable insights into methodological strengths and limitations. To our knowledge, there is no such paper that engages such a mathematical comparison of methods. Appendices include a concise primer on our chosen methods as well as full results from each, for both the methodological novice and those seeking greater detail on our findings.

For these comparative analyses, this article uses the set of data features chosen and explicated in a 2016 monograph, *Paul's Letters and Contemporary Greco-Roman Literature: Theorizing a New Taxonomy*.[2] In that monograph, Robertson makes the case for what are termed "top-down" features, namely features that are mostly second-order categories selected by the analyst trained in this area and hand-coded into Excel spreadsheets. This method is distinct from and contrasted to the so-called "bottom-up" feature selection comprised of vocabulary words and grammatical position, which can be machine-read and require no particular training or expertise in the text or culture in question. The latter (bottom-up) forms the foundation of traditional stylometry and related methods; the former (top-down) is mostly Robertson's unique contribution to the field. Interested readers can find further discussion on theorizing second-order features in that monograph, where Robertson makes the case that top-down, second-order features both better capture an author's distinct style and reveal interesting conclusions around comparing Paul's letters with various contemporary texts.

The feature list is reproduced here, with all subsequent analyses using the hand-coded spreadsheets in Robertson 2016 that indicate the position of every feature occurrence in each text.

Appeals to Authority
Conversation
Prosópopoiia/Éthopoiia
Rhetorical Questions
Metaphors or Analogies
Anecdotes or Examples
Imperatives
Exhortation
Caustic Injunctions
Pathos
Irony or Satire
Hyperbole
Oppositions or Choices
Figurations of Groupness
Plural Inclusive versus Second Person Address
First Person Reflection
Analysis of Potential Questions or Objections
Systematic Argument

2 Methods Introduction

This section introduces each method, describing its function, relevant mathematical information, typical applications, and utility for our research questions, including strengths and limitations.

---

1 Al-Anazi, "Finding Similar Documents using Different Clustering Techniques," 28–34.

2 Robertson, *Paul's Letters and Contemporary Greco-Roman Literature.*

2.1 Euclidean Distance

A key method in author attribution, Euclidean distance aids stylometry by quantifying stylistic differences between texts and revealing textual patterns.[3] Euclidean distance calculates the separation between texts in a high-dimensional space, considering attributes like word frequency and sentence structure, both key stylometric criteria. Euclidean distance is defined by the following:

$$d(p, q) = \sqrt{\sum_{i=1}^{n}(q_i - p_i)^2}\,.$$

A simple example to illustrate: imagine you are standing at one corner of a room, and you want to get to the opposite corner. The Euclidean distance is the straight line that takes you directly to that corner. In a two-dimensional space, the formula to compute this straight-line distance between two points (say, Point 1(x_1, y_1) and Point 2(x_2, y_2)) is:

$$\sqrt{((x_2 - x_1)^2 + (y_2 - y_1)^2)}\,,$$

the Euclidean distance between two points $P1(x_1, y_1)$ and $P2(x_2, y_2)$ in a 3D space is given by the following formula:

$$d(P1, P2) = \sqrt{((x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2)}\,.$$

So, the distance will be calculated as the square root of the sum of the squares of differences in both x and y coordinates.
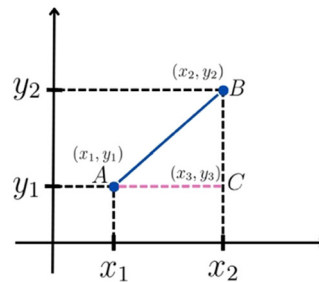
Assuming:

$AC = 2\text{m}$ and $BC = 1\text{m}$

Euclidean distance can be calculated as (for 2D):

$AB^2 = AC^2 + BC^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2$

Therefore,

$AB = \sqrt{AC^2 + BC^2} = \sqrt{2^2 + 1^2} = \sqrt{5}$

$AB \approx 2.236\text{m}$

In the context of stylometric analysis, Euclidean distance serves as a crucial tool for measuring the similarity or dissimilarity between texts based on their stylometric features.

Euclidean space is a geometric construct that provides a framework for understanding the relationship between texts in terms of their stylometric attributes. Stylometric features encompass various linguistic attributes such as word frequency, length, syntactic structures, and vocabulary richness. Each text can be represented as a point in a high-dimensional space, with each coordinate corresponding to the values of these stylometric features. Euclidean distance calculates the separation between these points, indicating the similarity or dissimilarity between texts.

Beyond stylometric analysis, Euclidean distance finds applications in various fields, including machine learning, data science, computer vision, image processing, GIS, and spatial analytics. It is used in algorithms like K-Nearest Neighbors, K-Means, and Principal Component Analysis for classification, clustering, and dimensionality reduction tasks. In computer vision, Euclidean distance aids in pattern recognition, similarity checks, and object detection. In GIS, it helps calculate geographical distances and perform spatial operations.

One of the most significant advantages of Euclidean distance is its simplicity and intuitive calculation method. It can be applied across different dimensions, making it highly versatile for multi-dimensional data analysis tasks,

---

3 Liberti et al., "Euclidean Distance Geometry and Applications."

such as assessing the proximity of texts that are coded according to many different features. Additionally, it represents the most direct path between two points, facilitating easy interpretation and understanding.

Euclidean distance has limitations in stylometric analysis and other applications. One of its key limitations is that it is sensitive to scale. This means that the distances computed might be skewed depending on the units of the features. Another limitation is that as the dimensionality of the data increases, Euclidean distance tends to become less effective – a phenomenon known as the "curse of dimensionality." Euclidean distance also does not work well when the data are sparse. The vectors should be non-zero in approximately 75% of attributes for Euclidean distance to be usable, a limitation that applies to any other norm-induced distance measures – such as Manhattan distance, discussed below. In textual data, other distance measures are typically preferred over Euclidean distance, because Euclidean distance depends on the vector's magnitude instead of the angle between the vectors. The angle measure is more resilient to variations of occurrence counts between semantically similar terms, whereas occurrence counts and heterogeneity of word neighborhoods influence the magnitude of vectors. Even if the two text documents are far apart by Euclidean distance, in other words, there is still a chance that they are close to each other in terms of their context.

2.2 Cosine Similarity

Cosine similarity is a pivotal concept in mathematics and computational linguistics, measuring the similarity between vectors by calculating the cosine of the angle between them, commonly used in stylometry for comparing textual documents represented as Bag-of-Words (BoW) or Term Frequency-Inverse Document Frequency (TF-IDF) representations.[4] These representations transform textual documents into numerical vectors, where each element corresponds to the frequency or importance of a specific term within the document. By employing such representations, complex textual information is condensed into structured numerical formats, facilitating computational analysis and comparison. Cosine similarity operates by calculating the cosine of the angle between two non-zero vectors. Mathematically, it is expressed as the dot product of the vectors divided by the product of their magnitudes. This formulation enables the comparison of documents based on their directional similarity rather than their magnitudes, making it robust to variations in document lengths and word frequencies.

This similarity metric, ranging from −1 to 1, determines perfect alignment at 1, perfect dissimilarity at −1, and orthogonality at 0, making it robust for tasks like document clustering, authorship attribution, plagiarism detection, and information retrieval in natural language processing. In collaborative filtering-based recommendation systems, cosine similarity helps identify users or items having similar preferences. This technique is commonly used in movie, music, and product recommendations. Cosine similarity can also be used in text classification tasks such as spam detection, sentiment analysis, or topic modeling. Search engines use cosine similarity to match user queries with relevant documents by comparing the query vector to document vectors. This approach is commonly employed in information retrieval. Finally, cosine similarity is useful for clustering similar data points together, for example to group similar news articles or social media posts.

Cosine similarity is defined by the following:

$$\text{Similarity}(A, B) = \cos(\theta) = \frac{A \cdot B}{||A|| ||B||} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2 \sum_{i=1}^{n} B_i^2}}.$$
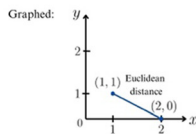
---

4 Al-Hagery and Hassan, "Google Search Filter using Cosine Similarity Measure to Find All Relevant Documents of a Specific Research."

Below is a simple example of cosine similarity:

Sentence 1: Hello world   Sentence 2: Hello hello

Mapping them to vector:

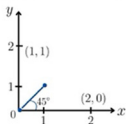|	Hello	World
Sentence 1	1	1
Sentence 2	2	0

Graphed:

$(1, 1)$ Euclidean distance
$(2, 0)$

Applying Euclidean distance formula:

Euclidean Distance $= \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$
$= \sqrt{(2 - 1)^2 + (1 - 0)^2} = \sqrt{2} \approx 1.4142$

Applying cosine similarity:

$(1,1)$
$45°$ $(2, 0)$

Applying formula:

$\cos(45) = 0.7071$

This is the same even if sentence 2 is "hello hello hello"

Multi-Dimensional Array Example:

Cosine similarity $= \dfrac{A \cdot B}{\|A\| \cdot \|B\|}$ } $\dfrac{\text{Dot product of vectors A and B}}{\text{Magnitude (Euclidean norm) of vectors A and B}}$

Sentence 1: I am Iron Man   Sentence 2: Iron can rust

Vector matrix:

	I	Am	Iron	Man	Can	Rust
Sentence 1	1	1	1	1	0	0
Sentence 2	0	0	1	0	1	1

Cosine similarity $= \dfrac{\sum_{i=1}^{n} A_i B_i}{\sum_{i=1}^{n} A_i^2 \times \sum_{i=1}^{n} B_i^2} = \dfrac{(1 \times 0) + (1 \times 0) + (1 \times 1) + (1 \times 0) + (0 \times 1) + (0 \times 1)}{\sqrt{(1^2 + 1^2 + 1^2 + 1^2 + 0^2 + 0^2) \times (0^2 + 0^2 + 1^2 + 0^2 + 1^2 + 1^2)}}$

$= \dfrac{1}{\sqrt{4} \times \sqrt{3}} = \dfrac{1}{2 \times (1.732)} = \dfrac{1}{3.46} = 0.289$

Cosine similarity is a fundamental concept in mathematics and computational linguistics, offering a powerful tool for measuring the similarity between two vectors in an inner product space. In the context of stylometry, which concerns the quantitative analysis of literary style, cosine similarity serves as a cornerstone for comparing textual documents based on their vectorized representations.

Cosine similarity is easy to compute, especially with sparse matrices, and it can capture the overall similarity of the documents regardless of their length. Furthermore, cosine similarity is remarkably versatile. It can be applied not only to pairs of points but also to a diverse array of data types, including text documents, images, and numerical data. This versatility renders cosine similarity a powerful tool in various domains. However, it lacks consideration for the magnitude of vectors, focusing solely on their direction – it yields the same value regardless of the size of the vectors being compared as long as the angle between them is the same. This can be problematic when comparing vectors of different sizes, such as when comparing the risk profiles of different companies. It also does not capture the semantic meaning of words, making it less effective in understanding context. Additionally, in high-dimensional spaces where data are sparse, cosine similarity can be less reliable – it might not accurately reflect the true similarity between data points. Another limitation is that cosine similarity depends on vector normalization. Different normalization methods can yield different results, making comparisons sensitive to preprocessing choices.[5]

2.3 Jaccard Similarity

Jaccard similarity is a key tool in data analysis and information retrieval, quantifying the similarity between sets by comparing their overlap.[6] It calculates the ratio of intersecting elements to the total elements in the sets, providing a score from 0 to 1 where 1 signifies perfect similarity, and 0 indicates no common elements. It is defined by the following equation:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}.$$

---

5 Ale, *Drawbacks with Cosine Similarity*. Cross Validated.
6 Shtovba and Petrychko, "Jaccard index-Based Assessing the Similarity of Research Fields in Dimensions."

Here is a simple example of Jaccard similarity:

Calculated as: $\dfrac{\text{Number of obvservations in both sets}}{\text{Number in either set}}$ Notation form: $J(A, B) = \dfrac{|A \cap B|}{|A \cup B|}$

Example:

$E = [\text{cat, dog, hippo, monkey}]$ # in both: $[\text{monkey}] = 1$ $\left.\begin{array}{c}\\\\\end{array}\right\}$ Jaccard similarity $= \dfrac{1}{7} \approx 0.142857$

$F = [\text{monkey, rhino, ostrich, salmon}]$ # in either: $= 7$ Since this number is low, it indicates that the two sets are quite dissimilar.

To calculate Jaccard similarity, the number of common elements (intersection) between two sets is divided by the total number of distinct elements (union) present in both sets. This ratio ranges from 0 to 1, where 1 signifies perfect overlap and 0 indicates no common elements.

Jaccard similarity is a fundamental concept in data analysis and information retrieval, offering a straightforward method to quantify the similarity between sets. Its applications span various domains, including text analysis, genomic analysis, social network analysis, and data deduplication. Jaccard similarity offers a different perspective in stylometric analysis, enabling the comparison of text documents based on their word content. By representing texts as sets of words and computing the ratio of intersection to union, analysts can discern similarities and differences in writing styles.

Jaccard similarity is a useful metric that has applications in various fields. One of its popular applications is in Natural Language Processing (NLP), where it is used to compare the similarity of two text documents. This technique is used in document clustering, plagiarism detection, and recommendation systems. Jaccard similarity is also used in recommendation systems based on collaborative filtering. It helps identify users with similar preferences by comparing the sets of items each user has interacted with. This way, recommendations can be made based on the preferences of similar users. In the field of genomic analysis, biologists use Jaccard similarity to measure the similarity between genetic sequences. By identifying common elements and mutations, Jaccard similarity aids in understanding evolutionary relationships. Social network analysis similarly uses Jaccard similarity to help analyze social networks by comparing the sets of friends or connections between individuals, thereby more easily identifying users with similar social circles. Finally, Jaccard similarity can also be used in data cleaning and preprocessing tasks. It can detect and remove duplicate records by comparing the sets of attributes within records to help in data deduplication, which is essential in maintaining data quality.

This metric is highly recommended for comparing sets of data for more conceptual reasons. One of the most significant advantages of using Jaccard similarity is that it is scale-invariant, which means that the size of the sets being compared does not affect the metric's performance. This feature makes it a robust metric for comparing sets of varying sizes. Another advantage of Jaccard similarity is that it is easy to understand. Its simple concept makes it accessible to both data scientists and non-technical stakeholders. Finally, Jaccard similarity is versatile and can be applied to a wide range of data types and domains. This makes it a highly flexible metric for similarity measurement, from text data to biological data.

However, Jaccard similarity has certain limitations. First, it treats all elements equally without considering their frequency of occurrence. This means that words appearing multiple times in a text receive the same weight as words appearing only once, which can be a limiting factor as word frequency often carries important stylistic and authorial information. Second, Jaccard similarity is sensitive to the size of the sets being compared. Even slight variations in the size of the sets can lead to significant changes in the similarity score, which may not always accurately reflect the similarity between texts. Third, Jaccard similarity only considers the presence or absence of elements, completely ignoring their order or arrangement within the sets. This can be problematic in stylometric analysis, where the sequential arrangement of words, phrases, or stylistic features may carry important information about an author's style. Finally, Jaccard similarity is based on set operations and does not capture semantic relationships between words, which is a significant limitation. Due to this, it may not effectively capture similarities between texts that use synonyms, paraphrasing, or related concepts.

2.4 Manhattan Distance

Manhattan distance is a practical measure in mathematics and computer science that is useful in grid-based systems where movement is constrained to horizontal and vertical paths. Stylometric analysis quantifies textual similarities by summing the absolute differences between corresponding linguistic features, aiding tasks like authorship attribution and route planning across various fields. It is defined by the following:
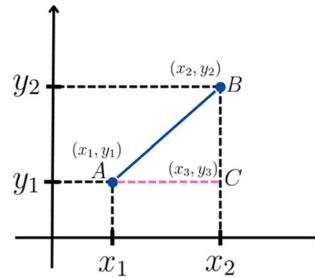
$$\text{Manhattan distance} = \sum_{i=1}^{n} |x_i - y_i|.$$

A simple example of Manhattan distance:

Assuming:

$AC = 2\text{m}$ and $BC = 1\text{m}$

$$\text{Manhattan distance} = |x_2 - x_1| + |y_2 - y_1|$$
$$(AB)$$
$$= |2| + |1|, \; AB = 3$$

In stylometric analysis, Manhattan distance can be a valuable tool. By representing texts as vectors, with each element corresponding to the frequency of certain linguistic features (such as word occurrences, character n-grams, or syntactic structures), analysts can compute the Manhattan distance between these vectors to quantify the dissimilarity or similarity between texts.

Manhattan distance provides a straightforward and versatile method for measuring distance within grid-based systems, and its application extends beyond mathematical contexts to fields such as stylometry, where it facilitates the quantitative analysis of textual similarities and differences. Once the Pauline letters are represented as vectors in the feature space, the Manhattan distance can be calculated between pairs of letters. For each pair of letters attributed to different authors (e.g., Paul and another potential author), the Manhattan distance is computed by summing the absolute differences between the corresponding components of their feature vectors. Moreover, Manhattan distance is advantageous in scenarios where the data are naturally discrete or when the cost of movement along different dimensions is not uniform. This makes it a preferred choice in many practical applications, including route planning, image recognition, and clustering algorithms.

Manhattan distance is less affected by outliers compared to Euclidean distance. It assigns equal weight to each dimension and is more robust to variations in individual coordinates. If your data have outliers or you want to reduce the impact of extreme values, Manhattan distance might be a better choice. Also, you might want to consider the Manhattan distance if the input variables are not similar in type (such as age, gender, height, etc.). Due to the curse of dimensionality, we know that Euclidean distance becomes a poor choice as the number of dimensions increases. As it does not take any squares, it does not amplify the differences between any of the features. It also does not ignore any features.

Manhattan distance is a distance metric that treats each dimension or coordinate equally, regardless of the scale or importance of the feature. However, in cases where certain dimensions are more significant than others, this can lead to inaccurate distance measurements. Additionally, Manhattan distance restricts movement to horizontal and vertical paths, which can be inadequate for situations where diagonal movement is allowed or relevant. In such cases, other distance metrics like Euclidean distance might provide a better representation of true distances. It's worth noting that while Manhattan distance excels in grid-based systems like city maps or game boards, it might not be the most appropriate metric for continuous or non-grid systems where movement is not constrained to specific paths.

2.5 Burrows' Delta Method

Burrows' Delta Method is a measure of the distance between a text whose authorship we want to ascertain and some other corpus. This method is designed to compare an anonymous text (or set of texts) to many different authors' signatures at the same time.[7] More precisely, this method measures how the anonymous text and sets of texts written by an arbitrary number of known authors all diverge from the average of all of them put together.[8] Furthermore, the Delta Method gives equal weight to every feature that it measures, thus avoiding the problem of common words overwhelming the results.

There are several steps to execute Burrows' Delta method. Initially, a reference corpus comprising texts authored by various individuals is essential. The corpus should represent a diverse array of writing styles and subjects to ensure its efficacy in capturing distinctive patterns of function word usage. Next, the text under examination, whose authorship needs identification, is compared to the reference corpus. This comparison involves calculating the frequency of function words in both the text being analyzed and the reference corpus. Common function words include "the," "of," "and," "to," "in," "that," "it," "for," "you," and many others. Subsequently, statistical techniques are applied to gauge the discrepancy between the frequency distributions of function words in the analyzed text and the reference corpus. Burrows' Delta, a statistical measure, is often utilized for this purpose. It quantifies the differences between the text's function word usage and the average usage across the reference corpus.

A high delta (= difference) value suggests substantial dissimilarity between the text and the reference corpus in terms of function word frequency, indicating potential divergence in authorship. Conversely, a low delta value may imply similarity in writing styles and, thus, shared authorship characteristics. Finally, interpretation of the delta value demands careful consideration of contextual factors, including the size of the reference corpus, the genre and subject matter of the texts, and the potential presence of other linguistic features that could influence function word usage.

3 Data Analysis: Analyzing Paul's Letters Using Multiple Methods

In this section, we outline the steps taken to compute and analyze the similarities and differences among various texts using the distance and similarity measures above: Euclidean distance, cosine similarity, Manhattan distance, Jaccard similarity, and Burrows' Delta Method.

First, we defined the feature vectors for each text. These vectors represented the frequency of specific textual features. For instance, the feature vectors for texts like "1 Corinthians," "2 Corinthians," and others were compiled into a dictionary, allowing for structured and systematic analysis. Here is an example of the feature vectors defined for each text:

```
corinthians1 = [35, 36, 5, 4, 42, 6, 13, 10, 29, 5, 2, 0, 1, 9, 6, 36, 25, 22, 5, 16]
corinthians2 = [17, 27, 1, 1, 14, 6, 4, 3, 13, 1, 10, 2, 2, 8, 3, 5, 9, 3, 5, 2]
# Other vectors...
```

We then created DataFrames from this dictionary to facilitate computational analysis. Using the DataFrames, we calculated the Euclidean distance, cosine similarity, and Manhattan distance. The calculation of these metrics involved the following code:

```
euclidean_matrix = euclidean_distances(df.T)
cosine_matrix = cosine_similarity(df.T)

euclidean_df = pd.DataFrame(euclidean_matrix, columns=df.columns, index=df.columns)
cosine_df = pd.DataFrame(cosine_matrix, columns=df.columns, index=df.columns)
# Other DataFrames…
```

---

7 Argamon, "Interpreting Burrows's Delta: Geometric and Probabilistic Foundations," 131–47.
8 Evert et al., "Understanding and Explaining Delta Measures for Authorship Attribution."

The Jaccard similarity was calculated using a custom function designed to handle numerical arrays:

```python
def jaccard_similarity(x, y):
    intersection = np.minimum(x, y).sum()
    union = np.maximum(x, y).sum()
    return intersection / union

jaccard_matrix = np.array(
    [[jaccard_similarity(df[text1], df[text2]) for text2 in texts.keys()] for text1 in
texts.keys()])

jaccard_df = pd.DataFrame(jaccard_matrix, columns=df.columns, index=df.columns)
```

The code for Burrows' Delta Method involved several steps, including calculating *z*-scores and delta scores:

```python
# Calculate percentages for each feature in each subcorpus
percentages = df.div(df.sum(axis=0), axis=1) * 100

# Calculate mean and standard deviation for each feature
feature_means = percentages.mean(axis=1)
feature_stds = percentages.std(axis=1)

# Calculate z-scores for each feature in each subcorpus
z_scores = (percentages.sub(feature_means, axis=0)).div(feature_stds, axis=0)

# Calculate Delta scores
delta_scores = pd.DataFrame(index=z_scores.columns, columns=z_scores.columns)

for text1 in z_scores.columns:
    for text2 in z_scores.columns:
        delta_scores.at[text1, text2] = np.mean(np.abs(z_scores[text1] -
z_scores[text2]))
```

In this code, several Python libraries were used to allow for easy and accurate calculations.[9, 10] The Pandas library was important for data manipulation and analysis, providing helpful data structures like DataFrame, which are essential for handling large datasets. NumPy was another library used in our code, supporting large, multi-dimensional arrays and matrices. It offers a collection of mathematical functions to operate on these arrays, which was needed for performing numerical operations and calculations necessary for the analysis. SciKit-Learn, a machine learning library, was used extensively to calculate the Euclidean distance, cosine similarity, and Manhattan distance. It offers simple and efficient tools, enabling ease of application of the various distance and similarity measures. Finally, the OS module, a standard utility in Python, was used to interact with the operating system. It aided with reading from and writing to files, ensuring smooth handling of input and output operations.

4 Discussion of Euclidean Distance Results

In analyzing the Euclidean distances between the various texts, we can draw several important conclusions about their relationships and differences. In our context, each text is represented as a point in a 20-dimensional space where each dimension corresponds to the frequency of a specific rhetorical feature – 20 features in all. By examining these distances, we can infer the degree of similarity or difference between the texts based on their rhetorical profiles.

The numerical range of Euclidean distances can vary significantly. We identify the maximum and minimum distances observed to set a context for what counts as high or low (Figure 1).

Low distance pairs such as 1 Corinthians and Galatians, with a distance of 64.76, indicate a close relationship, suggesting that these texts share similar patterns and frequencies for the comparanda. Similarly, 2

---

9 https://www.python.org/downloads/.

10 Pedregosa et al., *SciKit-Learn: Machine Learning in Python.*

	1 Cor	2 Cor	1 Thess	2 Thess	4 Macc	Arist Pan	Epict Disc	Gal	Romans	Philem	Phil	Philo Death	Philo Piety	Seneca	Damascus	Hebrews
1 Cor	0	58.89822	79.27168	79.14544	69.33974	70.53368	522.2078	64.7611	42.91853	89.31965	78.89867	85.7613	79.25276	63.44289	72.97945	45.18849
2 Cor	58.89822	0	30.54505	40.28647	31.85906	61.26989	564.2818	22.51666	31.73326	40.58325	29.15476	40.84116	53.74012	34.35113	46.76537	60.37384
1 Thess	79.27168	30.54505	0	21.81742	24.57641	63.58459	581.9167	17.49286	51.59457	15.55635	11	20.51828	53.03772	30.54505	56.08921	74.44461
2 Thess	79.14544	40.28647	21.81742	0	34.75629	70.04998	584.3295	24.45404	53.88877	22.31591	23.08679	28.68798	61.87891	29.41088	67.34983	76.10519
4 Macc	69.33974	31.85906	24.57641	34.75629	0	45.26588	566.7619	18.38478	41.06093	32.55764	27.87472	23.30236	41.55719	24.43358	51.49757	63.16645
Arist Pan	70.53368	61.26989	63.58459	70.04998	45.26588	0	557.5249	56.20498	56.92978	71.56116	66.25708	64.23395	64.49806	57.63679	64.11708	64.64519
Epict Disc	522.2078	564.2818	581.9167	584.3295	566.7619	557.5249	0	569.678	551.5696	590.6276	582.4208	582.2697	563.0524	560.8351	558.5633	534.1339
Gal	64.7611	22.51666	17.49286	24.45404	18.38478	56.20498	569.678	0	37.30952	27.0555	19.87461	26.13427	48.4252	21.23676	50.69517	62.73755
Romans	42.91853	31.73326	51.59457	53.88877	41.06093	56.92978	551.5696	37.30952	0	60.63003	51.68172	55.6507	52.39275	40.36087	51.80734	42.30839
Philem	89.31965	40.58325	15.55635	22.31591	32.55764	71.56116	590.6276	27.0555	60.63003	0	16.15549	18.46619	60.17475	36.56501	67.33498	86.98276
Phil	78.89867	29.15476	11	23.08679	27.87472	66.25708	582.4208	19.87461	51.68172	16.15549	0	22.13594	54.40588	32.68027	57.92236	76.43952
Philo Death	85.7613	40.84116	20.51828	28.68798	23.30236	64.23395	582.2697	26.13427	55.6507	18.46619	22.13594	0	46.19524	33.34666	62.44197	80.75271
Philo Piety	79.25276	53.74012	53.03772	61.87891	41.55719	64.49806	563.0524	48.4252	52.39275	60.17475	54.40588	46.19524	0	56.10704	47.1487	68.66586
Seneca	63.44289	34.35113	30.54505	29.41088	24.43358	57.63679	560.8351	21.23676	40.36087	36.56501	32.68027	33.34666	56.10704	0	63.34824	62.74552
Damascus	72.97945	46.76537	56.08921	67.34983	51.49757	64.11708	558.5633	50.69517	51.80734	67.33498	57.92236	62.44197	47.1487	63.34824	0	59.5147
Hebrews	45.18849	60.37384	74.44461	76.10519	63.16645	64.64519	534.1339	62.73755	42.30839	86.98276	76.43952	80.75271	68.66586	62.74552	59.5147	0

Figure 1: Euclidean distances.

Corinthians and 1 Thessalonians, with a distance of 30.54, show a high degree of similarity, indicating that their usage of the comparanda is quite alike.

Moderate distance pairs like Romans and 1 Corinthians, with a distance of 42.92, and Hebrews and 1 Corinthians, with a distance of 45.19, suggest some degree of similarity but also notable differences. These moderate distances imply that while there are commonalities between these texts, they are not as closely related as the low-distance pairs.

High-distance pairs, such as Epictetus' *Discourses*, consistently show high distances from all other sets, with the smallest distance being 522.21 to 1 Corinthians. This indicates that Epictetus' *Discourses* are quite distinct from the other texts in terms of the frequency and pattern of the comparanda. Additionally, Philemon and 1 Corinthians, with a distance of 89.32, suggest significant differences, making them less comparable.

Interpreting these distances, low numbers (0–30) indicate strong similarities between the sets, suggesting that the compared documents share a high frequency of similar comparanda. Moderate numbers (30–60) suggest a moderate level of similarity, implying that while there are commonalities, there are also notable differences. High numbers (60 and above) indicate significant dissimilarities. For instance, values above 500 suggest that the documents are quite different in terms of the chosen comparanda, which might reflect different authors, genres, or thematic focuses.

5 Discussion of Manhattan Distance Results

In the context of rhetorical feature analysis, Manhattan distance provides insights into the total difference in feature counts between texts. Unlike Euclidean distance, which considers the shortest path in a multi-dimensional space, Manhattan distance strictly follows the axes, offering a different perspective on textual similarity.[11]

Manhattan distances can range from 0 to a large number, depending on the feature counts. Lower values indicate greater similarity, while higher values denote greater dissimilarity. For practical analysis, distances closer to 0 can be considered indicative of high similarity, while larger values reflect more substantial differences (Figure 2).

The low Manhattan distance of 98 between 1 Corinthians and Galatians suggests that these two texts have very similar frequency distributions of the selected comparanda, indicating a close thematic and rhetorical resemblance. This implies that these letters may share similar contexts or purposes, potentially reflecting common authorship or similar influences in their composition. Similarly, the distance of 110 between Romans

---

11 Suwanda et al. "Analysis of Euclidean distance and Manhattan Distance in the K-Means Algorithm for Variations Number of Centroid K."

	1 Cor	2 Cor	1 Thess	2 Thess	4 Macc	Arist Pan	Epict Disc	Gal	Romans	Philem	Phil	Philo Death	Philo Piety	Seneca	Damascus	Hebrews
1 Cor	0	193	256	262	198	231	1769	210	132	300	259	269	257	209	232	146
2 Cor	193	0	87	125	93	190	1956	69	105	129	88	132	160	106	143	189
1 Thess	256	87	0	58	78	171	2023	62	146	50	37	61	149	95	150	244
2 Thess	262	125	58	0	114	185	2029	84	156	38	59	63	177	97	192	246
4 Macc	198	93	78	114	0	123	1957	60	112	108	99	75	119	81	150	198
Arist Pan	231	190	171	185	123	0	1878	155	185	191	182	168	194	174	203	215
Epict Disc	1769	1956	2023	2029	1957	1878	0	1969	1885	2065	2028	2032	1908	1948	1883	1783
Gal	210	69	62	84	60	155	1969	0	106	98	73	89	139	69	156	208
Romans	132	105	146	156	112	185	1885	106	0	182	155	159	171	101	154	142
Philem	300	129	50	38	108	191	2065	98	182	0	51	43	163	119	182	282
Phil	259	88	37	59	99	182	2028	73	155	51	0	68	158	110	159	247
Philo Death	269	132	61	63	75	168	2032	89	159	43	68	0	128	110	177	251
Philo Piety	257	160	149	177	119	194	1908	139	171	163	158	128	0	160	149	233
Seneca	209	106	95	97	81	174	1948	69	101	119	110	110	160	0	201	213
Damascus	232	143	150	192	150	203	1883	156	154	182	159	177	149	201	0	174
Hebrews	146	189	244	246	198	215	1783	208	142	282	247	251	233	213	174	0

Figure 2: Manhattan distances.

and Galatians also reflects a high degree of similarity, supporting the notion that these letters are closely related in their use of rhetorical and thematic elements.

In the moderate distance category, the distance of 165 between 2 Corinthians and 1 Thessalonians suggests some differences in the frequency of comparanda, indicating that while there are shared themes, there are also divergent elements that differentiate these texts. This could reflect variations in the contexts or audiences addressed by these letters. The distance of 180 between Philippians and 1 Thessalonians further supports this, hinting at some shared rhetorical strategies but also distinct variations in content and style.

High Manhattan distances, such as the 1,236 between Epictetus' *Discourses* and 2 Thessalonians, indicate significant differences in the frequency of comparanda. This suggests that these texts have distinct themes and rhetorical strategies, likely reflecting different genres, purposes, or authorship. Similarly, the high distance of 934 between Philodemus' *On Death* and Epictetus' *Discourses* highlights substantial differences in content and style, reinforcing the idea that these texts are markedly distinct in their rhetorical and thematic approaches.

The interpretation of these distance scores helps in understanding the degree of similarity and difference between the texts. Low distances (below 200) indicate high similarity in the frequency distributions of comparanda, suggesting thematic and rhetorical consistency. Moderate distances (200–600) suggest some similarities but also notable differences, reflecting a mix of shared and distinct elements. High distances (above 600) highlight significant differences in the frequency distributions, pointing to distinct themes, topics, and rhetorical strategies.

6 Discussion of Cosine Similarity Results

In analyzing the cosine similarity between the various texts, we can determine the degree of alignment or similarity in their rhetorical features. Cosine similarity measures the cosine of the angle between two vectors, which in this context are the rhetorical feature frequency vectors of the texts. Cosine similarity values range from −1 to 1, though in the context of text analysis, values typically range from 0 to 1. A similarity close to 1 indicates high similarity, while a value near 0 indicates low similarity. For example, values above 0.8 could be considered high, indicating strong similarity, whereas values below 0.3 could be considered low, indicating significant differences (Figure 3).

High similarity pairs such as 1 Corinthians and Galatians, with a cosine similarity score of 0.945160, suggest a strong thematic and rhetorical congruence between these texts. This high score indicates that the frequency of the chosen comparanda in these letters is quite similar, pointing toward a consistently shared conceptual focus and rhetorical style. Similarly, the high similarity score of 0.900998 between Romans and Hebrews highlights a strong resemblance in the usage of comparanda, suggesting shared theological themes and rhetorical strategies, possibly indicative of similar authorship or common theological discourse. Another pair exhibiting high similarity is 1 Corinthians and Romans, with a score of 0.912127. This reinforces the idea that Paul's letters share significant thematic and stylistic consistency. Such high similarity scores suggest

	1 Cor	2 Cor	1 Thess	2 Thess	4 Macc	Arist Pan	Epict Disc	Gal	Romans	Philem	Phil	Philo Death	Philo Piety	Seneca	Damascus	Hebrews
1 Cor	1	0.861268	0.737964	0.603348	0.758147	0.649114	0.795826	0.94516	0.912127	0.614495	0.721603	0.38125	0.517405	0.830793	0.615122	0.873853
2 Cor	0.861268	1	0.782958	0.355685	0.665782	0.530966	0.674286	0.864956	0.882908	0.530544	0.814049	0.288244	0.511406	0.633643	0.746072	0.800038
1 Thess	0.737964	0.782958	1	0.466304	0.700937	0.587765	0.58378	0.821173	0.691721	0.618943	0.811141	0.350482	0.586226	0.60027	0.796545	0.84893
2 Thess	0.603348	0.355685	0.466304	1	0.294127	0.252015	0.334328	0.57639	0.511023	0.493447	0.418704	0.097214	0.167056	0.623131	0.234086	0.621659
4 Macc	0.758147	0.665782	0.700937	0.294127	1	0.887193	0.763078	0.834774	0.79056	0.309737	0.549445	0.737015	0.768077	0.769228	0.696673	0.839478
Arist Pan	0.649114	0.530966	0.587765	0.252015	0.887193	1	0.522281	0.694339	0.649345	0.22171	0.435646	0.530327	0.543012	0.608028	0.587363	0.693817
Epict Disc	0.795826	0.674286	0.58378	0.334328	0.763078	0.522281	1	0.783043	0.679416	0.395961	0.522137	0.521599	0.509707	0.833921	0.527536	0.694617
Gal	0.94516	0.864956	0.821173	0.57639	0.834774	0.694339	0.783043	1	0.921596	0.559085	0.724144	0.453869	0.635325	0.827963	0.758342	0.93125
Romans	0.912127	0.882908	0.691721	0.511023	0.79056	0.649345	0.679416	0.921596	1	0.47318	0.66245	0.47251	0.638717	0.778175	0.691029	0.900998
Philem	0.614495	0.530544	0.618943	0.493447	0.309737	0.22171	0.395961	0.559085	0.47318	1	0.779631	0.160357	0.386269	0.381173	0.518107	0.523449
Phil	0.721603	0.814049	0.811141	0.418704	0.549445	0.435646	0.522137	0.724144	0.66245	0.779631	1	0.287834	0.502471	0.494701	0.680302	0.715893
Philo Death	0.38125	0.288244	0.350482	0.097214	0.737015	0.530327	0.521599	0.453869	0.47251	0.160357	0.287834	1	0.85932	0.463403	0.461997	0.502423
Philo Piety	0.517405	0.511406	0.586226	0.167056	0.768077	0.543012	0.509707	0.635325	0.638717	0.386269	0.502471	0.85932	1	0.438379	0.743271	0.633743
Seneca	0.830793	0.633643	0.60027	0.623131	0.769228	0.608028	0.833921	0.827963	0.778175	0.381173	0.494701	0.463403	0.438379	1	0.413348	0.797965
Damascus	0.615122	0.746072	0.796545	0.234086	0.696673	0.587363	0.527536	0.758342	0.691029	0.518107	0.680302	0.461997	0.743271	0.413348	1	0.741124
Hebrews	0.873853	0.800038	0.84893	0.621659	0.839478	0.693817	0.694617	0.93125	0.900998	0.523449	0.715893	0.502423	0.633743	0.797965	0.741124	1

Figure 3: Cosine similarity scores.

that these letters might have been intended for similar audiences or contexts, emphasizing consistent theological teachings and rhetorical methods.

On the other hand, moderate similarity pairs such as 2 Corinthians and 1 Thessalonians, with a score of 0.782958, indicate some divergence in themes or rhetorical approaches while still maintaining notable similarities. This score suggests that while these letters share some common elements, there are also distinct differences possibly arising from different contexts or shifts in the author's focus. Similarly, Philippians and Galatians, with a score of 0.724144, exhibit moderate similarity, suggesting shared rhetorical elements and theological points but also highlighting variations that could stem from different situational contexts.

Low similarity pairs, such as Philodemus' *On Death* and 2 Thessalonians, with a score of 0.097214, indicate significant differences in the frequency of comparanda between these texts. This low score suggests that the rhetorical styles and themes in these works are very different, reflecting distinct genres or thematic focuses. Similarly, the low similarity score of 0.334328 between Epictetus' *Discourses* and 2 Thessalonians perhaps underscores the contrast between the philosophical nature of Epictetus' work versus the more religious content of 2 Thessalonians.

In interpreting these similarity scores, high similarity (0.8–1) indicates a high degree of thematic and rhetorical consistency, suggesting common authorship, similar audience, or consistent theological focus. Moderate similarity (0.5–0.8) suggests shared elements with notable differences, possibly arising from different writing contexts or shifts in rhetorical strategy. Low similarity (0–0.5) highlights significant differences in content and style, indicating distinct themes, purposes, or rhetorical techniques, and potentially different authors. It is important to note that there is no universally defined threshold for determining what constitutes a low or high cosine similarity score. The classification of a score as high or low is contingent upon the specific data being analyzed and is derived from internal comparison to other scores instead of an objective bar for significance.

7 Discussion of Jaccard Similarity Results

The Jaccard similarity index provides a different perspective on the relationship between texts by focusing on the presence or absence of specific features rather than their frequency. In this context, the sets are the rhetorical features present in each text. Jaccard similarity ranges from 0 to 1, where 0 indicates no common features and 1 indicates identical feature sets (Figure 4).

High similarity pairs include 1 Corinthians and Galatians with a Jaccard similarity score of 0.579618. This relatively high score indicates a significant overlap in the presence of comparanda between these two letters,

	1 Cor	2 Cor	1 Thess	2 Thess	4 Macc	Arist Pan	Epict Disc	Gal	Romans	Philem	Phil	Philo Death	Philo Piety	Seneca	Damascus	Hebrews
1 Cor	1	0.393082	0.174194	0.14658	0.36129	0.368852	0.147059	0.324759	0.579618	0.022801	0.169872	0.126623	0.293956	0.346875	0.362637	0.606469
2 Cor	0.393082	1	0.378571	0.183007	0.459302	0.269231	0.060519	0.554839	0.511628	0.051471	0.375887	0.142857	0.304348	0.420765	0.388889	0.384365
1 Thess	0.174194	0.378571	1	0.275	0.376	0.189573	0.02553	0.446429	0.255102	0.122807	0.513158	0.227848	0.194595	0.311594	0.242424	0.172881
2 Thess	0.14658	0.183007	0.275	1	0.167883	0.127358	0.021225	0.282051	0.2	0.155556	0.271605	0.148649	0.082902	0.270677	0.098592	0.151724
4 Macc	0.36129	0.459302	0.376	0.167883	1	0.430556	0.055502	0.571429	0.461538	0.06087	0.272059	0.347826	0.40201	0.49375	0.339207	0.342193
Arist Pan	0.368852	0.269231	0.189573	0.127358	0.430556	1	0.093629	0.317181	0.348592	0.02551	0.16129	0.164179	0.297101	0.292683	0.307167	0.383954
Epict Disc	0.147059	0.060519	0.02553	0.021225	0.055502	0.093629	1	0.050169	0.09069	0.003378	0.024531	0.019305	0.079151	0.059846	0.091216	0.139479
Gal	0.324759	0.554839	0.446429	0.282051	0.571429	0.317181	0.050169	1	0.47	0.066667	0.381356	0.239316	0.318627	0.536913	0.306667	0.30897
Romans	0.579618	0.511628	0.255102	0.2	0.461538	0.348592	0.09069	0.47	1	0.037037	0.228856	0.180412	0.347328	0.512077	0.421053	0.541935
Philem	0.022801	0.051471	0.122807	0.155556	0.06087	0.02551	0.003378	0.066667	0.037037	1	0.12069	0.044444	0.023952	0.048	0.037037	0.024221
Phil	0.169872	0.375887	0.513158	0.271605	0.272059	0.16129	0.024531	0.381356	0.228856	0.12069	1	0.180723	0.168421	0.246575	0.216749	0.16835
Philo Death	0.126623	0.142857	0.227848	0.148649	0.347826	0.164179	0.019305	0.239316	0.180412	0.044444	0.180723	1	0.228916	0.19708	0.128079	0.134483
Philo Piety	0.293956	0.304348	0.194595	0.082902	0.40201	0.297101	0.079151	0.318627	0.347328	0.023952	0.168421	0.228916	1	0.285714	0.406375	0.3207
Seneca	0.346875	0.420765	0.311594	0.270677	0.49375	0.292683	0.059846	0.536913	0.512077	0.048	0.246575	0.19708	0.285714	1	0.217899	0.319489
Damascus	0.362637	0.388889	0.242424	0.098592	0.339207	0.307167	0.091216	0.306667	0.421053	0.037037	0.216749	0.128079	0.406375	0.217899	1	0.466258
Hebrews	0.606469	0.384365	0.172881	0.151724	0.342193	0.383954	0.139479	0.30897	0.541935	0.024221	0.16835	0.134483	0.3207	0.319489	0.466258	1

Figure 4: Jaccard similarity scores.

suggesting that similar themes and rhetorical elements are prominent in both. Another high similarity pair is 4 Maccabees and Galatians, which have a Jaccard similarity score of 0.571429. This high similarity points to common rhetorical or thematic elements in these texts, despite their different contexts.

Moderate similarity pairs include 2 Corinthians and 1 Thessalonians, which have a Jaccard similarity score of 0.378571. This score suggests a moderate level of overlap in the comparanda used, indicating some shared themes or rhetorical strategies but also notable differences. Another moderate similarity pair is Philippians and 1 Thessalonians, with a Jaccard similarity score of 0.513158. This score reflects a moderate degree of overlap, hinting at shared but not identical themes or rhetorical approaches.

Low similarity pairs include Philodemus' *On Death* and Epictetus' *Discourses*, with a Jaccard similarity score of 0.019305. This very low similarity indicates minimal overlap in the comparanda, suggesting significant differences in content and style. Another low similarity pair is Philodemus' *On Death* and 2 Thessalonians, which have a Jaccard similarity score of 0.148649. Similarly, a low similarity score highlights substantial differences between these texts.

High similarity scores (0.4–0.6) indicate a significant overlap in the presence of comparanda, suggesting shared themes, topics, or rhetorical strategies. Moderate similarity scores (0.2–0.4) suggest a moderate degree of overlap, indicating some common elements but also substantial differences. Low similarity scores (0–0.2) highlight minimal overlap, pointing to distinct differences in themes, topics, and rhetorical strategies.

8 Discussion of Burrows' Delta Method Results

In Burrows' Delta method, lower scores indicate greater stylistic similarity, while higher scores suggest more substantial stylistic differences. Each cell in the matrix represents the delta score between two texts (Figure 5).

|	1 Cor	2 Cor	1 Thess	2 Thess	4 Macc	Arist Pan	Epict Disc	Gal	Romans	Philem	Phil	Philo Death	Philo Piety	Seneca	Damascus	Hebrews
1 Cor	0	0.730895	0.840493	0.979054	0.818196	1.084194	0.926968	0.512446	0.555584	1.183252	0.926276	1.117094	0.99305	0.659728	1.045161	0.676825
2 Cor	0.730895	0	0.837534	1.138492	0.831172	1.304233	1.067349	0.725717	0.730896	1.513546	0.948462	1.37154	1.200959	1.011082	0.995342	0.856611
1 Thess	0.840493	0.837534	0	1.049012	1.009013	1.155673	1.303729	0.975669	0.989949	1.312899	0.837303	1.331085	1.20071	1.113394	0.844385	0.74249
2 Thess	0.979054	1.138492	1.049012	0	1.136715	1.263649	1.407538	1.081472	0.855401	1.049578	0.845887	1.06801	1.166358	1.100695	1.162679	0.660779
4 Macc	0.818196	0.831172	1.009013	1.136715	0	0.771301	0.943867	0.704018	0.742436	1.554118	1.149566	0.833699	0.785896	0.856543	0.875856	0.694064
Arist Pan	1.084194	1.304233	1.155673	1.263649	0.771301	0	1.542245	1.231395	0.994524	1.388597	1.254353	1.150326	1.272173	1.282446	1.066728	0.946892
Epict Disc	0.926968	1.067349	1.303729	1.407538	0.943867	1.542245	0	1.069143	1.231528	1.849916	1.524115	1.221356	1.295843	0.880264	1.369233	1.202581
Gal	0.512446	0.725717	0.975669	1.081472	0.704018	1.231395	1.069143	0	0.656122	1.434159	1.083956	1.096991	1.025827	0.875039	1.013408	0.709969
Romans	0.555584	0.730896	0.989949	0.855401	0.742436	0.994524	1.231528	0.656122	0	1.180697	0.925685	0.941671	0.966041	0.778559	0.876786	0.464114
Philem	1.183252	1.513546	1.312899	1.049578	1.554118	1.388597	1.849916	1.434159	1.180697	0	0.801129	1.370074	1.264088	1.625133	1.190328	1.18283
Phil	0.926276	0.948462	0.837303	0.845887	1.149566	1.254353	1.524115	1.083956	0.925685	0.801129	0	1.280067	1.071193	1.339896	1.001652	0.791626
Philo Death	1.117094	1.37154	1.331085	1.06801	0.833699	1.150326	1.221356	1.096991	0.941671	1.370074	1.280067	0	0.723533	1.089422	0.97128	0.927753
Philo Piety	0.99305	1.200959	1.20071	1.166358	0.785896	1.272173	1.295843	1.025827	0.966041	1.264088	1.071193	0.723533	0	1.180123	0.919919	0.996559
Seneca	0.659728	1.011082	1.113394	1.100695	0.856543	1.282446	0.880264	0.875039	0.778559	1.625133	1.339896	1.089422	1.180123	0	1.395967	0.909764
Damascus	1.045161	0.995342	0.844385	1.162679	0.875856	1.066728	1.369233	1.013408	0.876786	1.190328	1.001652	0.97128	0.919919	1.395967	0	0.666065
Hebrews	0.676825	0.856611	0.74249	0.660779	0.694064	0.946892	1.202581	0.709969	0.464114	1.18283	0.791626	0.927753	0.996559	0.909764	0.666065	0

Figure 5: Burrows' delta method scores.

The delta score between 1 Corinthians and Galatians is 0.512, indicating a high degree of stylistic similarity. This suggests that the two texts share many stylistic features, possibly reflecting similar authorship or a close temporal or contextual relationship. The score between Hebrews and 1 Corinthians is 0.676, also suggesting a significant stylistic similarity. These low delta scores, ranging from 0.512 to 0.800, suggest that the texts compared have a high degree of stylistic similarity. For instance, texts like 1 Corinthians and Galatians (0.512) and 1 Corinthians and Hebrews (0.676) might have been written by the same author or in a similar style.

In the realm of moderate similarity pairs, the delta score between 1 Thessalonians and 2 Corinthians is 0.837, indicating a moderate stylistic difference. This suggests that while there are some differences in stylistic features, the two texts are still relatively similar compared to others. The score between Romans and Galatians is 0.656, showing moderate similarity, which might reflect shared authorship or thematic connections. These medium delta scores, ranging from 0.800 to 1.100, indicate moderate stylistic differences. Texts in this range, such as 2 Thessalonians and Romans (0.855), while not identical in style, share some common stylistic features. This might indicate that different authors are writing in a similar style or that the same author is writing in different contexts.

Low similarity pairs include Epictetus' *Discourses* and 1 Thessalonians, with a delta score of 1.303, indicating substantial stylistic differences. This is expected as these texts likely belong to different genres and authors. Similarly, the score of 1.084 between Aelius Aristides' *Panathenaicus* and 1 Corinthians reflects significant stylistic differences, which could be attributed to different authorships as well as the highly different contexts and social purposes. High delta scores, ranging from 1.100 and above, reflect significant stylistic differences. Examples such as Epictetus' *Discourses* and 1 Thessalonians (1.303) or Aristides' *Panathenaicus* and 1 Corinthians (1.084) suggest different authors, genres, or contexts, highlighting distinct stylistic features.

9 Notable Similarities and Divergences

The visualizations and detailed analysis of textual similarities and divergences across different metrics provide a comprehensive understanding of the relationships between the texts. By examining the top five and bottom five similarity and distance scores, we can draw several notable insights and conclusions.
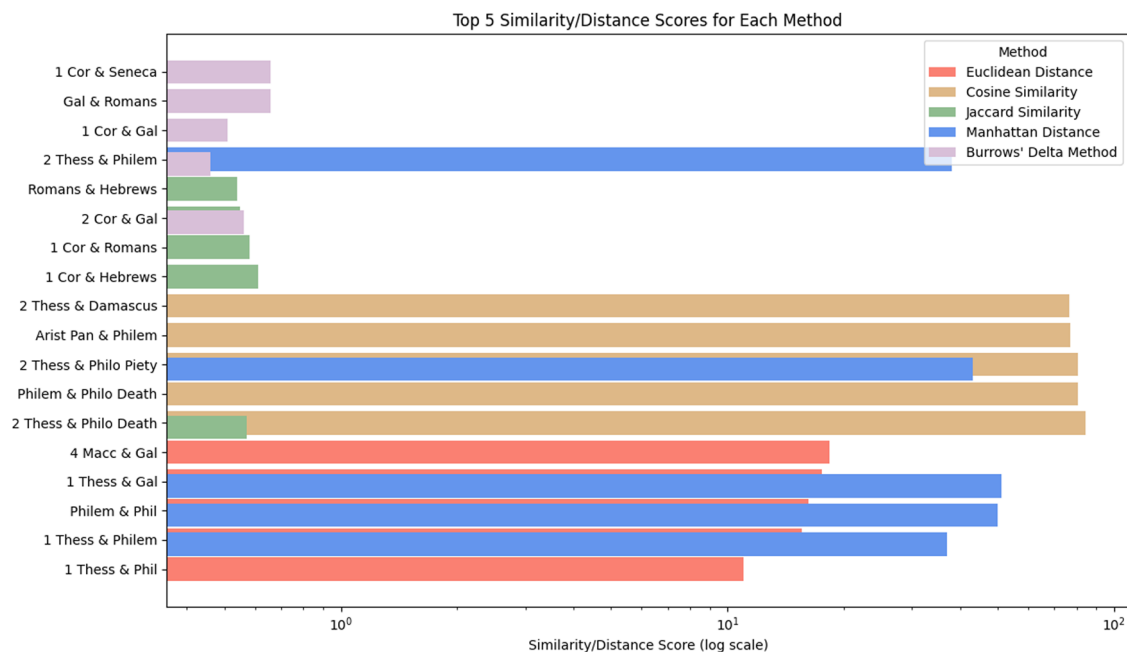
Graph of the top five pairs of texts that are most similar according to each method.
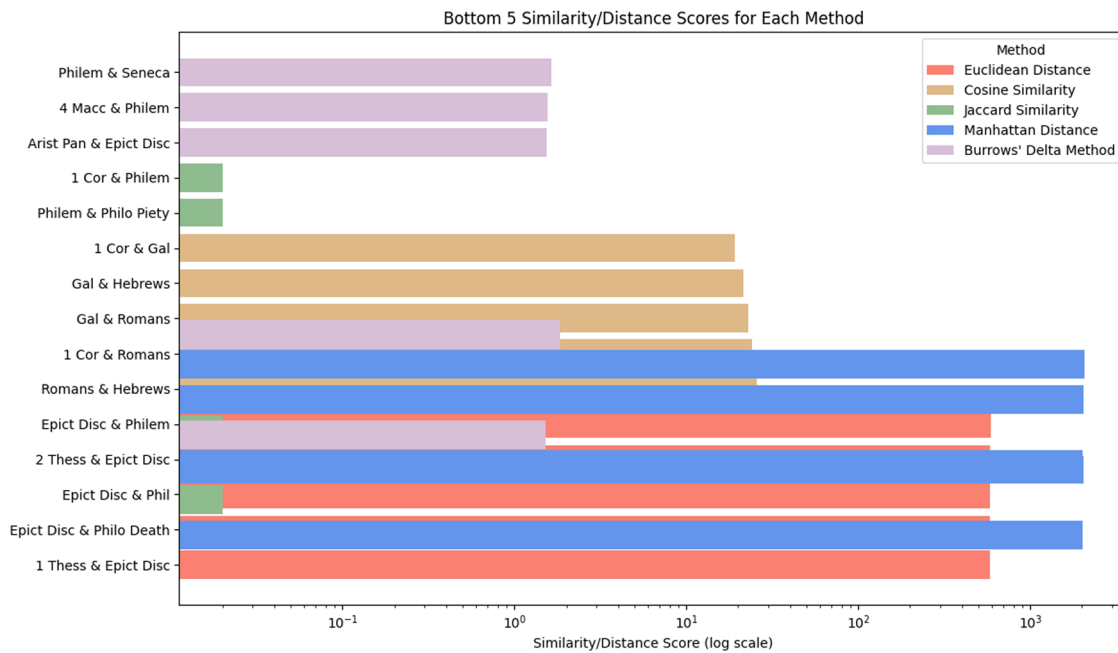
Graph of the bottom five pairs of texts that are least similar according to each method.

The high similarity between 1 Corinthians and Galatians is evident across multiple metrics, including cosine similarity, Jaccard similarity, and Burrows' Delta Method. This indicates that these texts share a significant number of common features and have a similar distribution of these features. For instance, the cosine similarity score between these texts is among the highest, reflecting their close alignment in terms of feature vectors. Similarly, the Jaccard similarity score shows a substantial overlap in features, while the Burrows' Delta score indicates minimal stylistic differences. The consistent high similarity scores across different metrics suggest that 1 Corinthians and Galatians are thematically and stylistically close. This is consistent with the known historical and contextual connections between these letters, both written by Paul addressing similar issues in early Christian communities.

Conversely, Epictetus' *Discourses* consistently shows high distances and low similarity scores with other texts across all metrics. For example, it has a high Euclidean distance of 522.21 with 1 Corinthians and a low Jaccard similarity of 0.061 with 2 Corinthians. This indicates that Epictetus' *Discourses* is markedly different from the other texts in terms of both feature counts and distributions, as well as stylistic elements. The consistent identification of Epictetus' *Discourses* as an outlier across multiple metrics highlights its unique characteristics and suggests that it may belong to a different genre or context compared to the other texts analyzed.[12]

The analysis also reveals that 1 Corinthians and 2 Corinthians have a close similarity across various metrics. For instance, their Euclidean distance is relatively small, indicating that the overall magnitude of differences in feature counts between these texts is minimal. Their high cosine similarity score further supports this, showing that the direction of their feature vectors is closely aligned. Additionally, the Jaccard similarity score indicates a substantial overlap in features, reflecting the shared themes and stylistic elements. This high degree of similarity is expected given that both letters were written by Paul to the same community, addressing related issues.

---

12 Cf. some of the conclusions in Robertson, *Paul's Letters and Contemporary Greco-Roman Literature*, which highlighted Epictetus' *Discourses* as a better comparison than Greco-Roman oratory and so-called Jewish apocalyptic.

Another notable finding is the similarity between 1 Thessalonians and Philippians. These texts show high similarity across metrics, with particularly high Manhattan and Euclidean distance scores, suggesting minimal differences in feature magnitudes. The Jaccard similarity score highlights a significant overlap in features, consistent with the expectation that these letters, written by the same author to the same community, would be thematically and stylistically similar. Furthermore, Romans and Hebrews show a high similarity score across cosine similarity, Jaccard similarity, and Burrows' Delta Method, indicating a significant overlap in themes and stylistic features. The Euclidean and Manhattan distances between these texts are moderate, suggesting that while there are differences in feature magnitudes, the overall distribution patterns are similar. This relationship may reflect shared theological themes or stylistic influences between these texts.

On the other hand, Philemon shows lower similarity scores and higher distance measures with most other texts. For instance, its Jaccard similarity with 1 Corinthians is very low, reflecting fewer shared features. The Euclidean and Manhattan distances are also relatively high, indicating more substantial differences in feature counts. This suggests that Philemon may have a unique set of features, or a different style compared to the other texts analyzed.

The bottom five similarity scores highlight texts with minimal overlap or significant differences. For example, the low similarity between Philemon and Seneca's *Natural Questions* across multiple metrics underscores the distinct thematic and stylistic elements of these texts. Similarly, the high Euclidean and Manhattan distances between Epictetus' *Discourses* and other texts further emphasize its unique characteristics and outlier status.

Triangulation, where three or more methods indicate high similarity between the same pairs of texts, further validates these findings. For instance, the high similarity between 1 Corinthians and Galatians is supported by cosine similarity, Jaccard similarity, and Burrows' Delta Method, providing robust evidence of their close relationship. Similarly, the outlier status of Epictetus' *Discourses* is consistently highlighted across all metrics, reinforcing its distinctiveness.

The visualizations illustrate these patterns clearly. The top five similarity scores graph shows texts like 1 Thessalonians with high similarity across multiple metrics, indicating its close relationship with multiple other texts. In contrast, the graph of the bottom five similarity scores highlights texts like Epictetus' *Discourses* and Philemon, which show significant differences from others. The graphs are plotted on a log scale to effectively represent the wide range of similarity and distance scores, allowing for a clearer visualization of both small and large values. Using a log scale helps to compress the range, making it easier to compare values that vary by order of magnitude. This approach is particularly useful in text analysis, where the similarity and distance measures can differ significantly between text pairs. For those interested, Appendix B includes all pairs, providing a complete overview of the relationships among the analyzed texts.

10 Conclusion

The comprehensive analysis of Paul's letters using multiple similarity and distance measures, including Euclidean distance, cosine similarity, Manhattan distance, Jaccard similarity, and Burrows' Delta Method, reveals significant insights into the textual and stylistic characteristics of these epistles. This study underscores the importance of employing diverse analytical methods to capture the nuanced relationships between texts. Each metric highlighted different aspects of similarity and dissimilarity, which can provide a multi-faceted understanding of the letters attributed to Paul.

The application of Euclidean distance provided a measure of overall dissimilarity where magnitude matters, revealing significant differences between texts such as Epictetus' *Discourses* and others, emphasizing the unique feature distribution of this text. Cosine similarity, which is ideal for documents with varying lengths, showed high similarity between 1 Corinthians and Galatians, suggesting strong thematic and rhetorical consistency. This consistency was further evidenced by high Jaccard similarity scores, indicating significant overlap in the presence of comparanda, such as between 1 Corinthians and Hebrews. Manhattan distance, known for its robustness to outliers, highlighted consistent feature count differences, particularly

between 1 Corinthians and Philemon, suggesting divergent themes or rhetorical approaches. Burrows' Delta Method, focusing on stylistic features, showed high similarity scores between Romans and Hebrews, indicating shared authorship or thematic connections.

These findings can usefully deeper insights into the Pauline corpus beyond what traditional qualitative analysis can offer. By quantitatively measuring the similarities and differences among Paul's letters, we can make more informed assertions about their authorship, thematic coherence, and rhetorical strategies. For instance, the consistently high similarity scores between certain letters reinforce the argument for their inclusion in Paul's authentic corpus. In contrast, others, like 2 Thessalonians, show significant stylistic differences, supporting the hypothesis of different authorship.

The implications of these findings extend to biblical scholarship, textual criticism, and historical theology. They offer a robust framework for analyzing ancient texts, combining computational methods with traditional literary analysis. This dual approach not only enhances our understanding of Paul's letters but also provides a model for studying other corpora with similar complexity. Furthermore, putting these methods side by side allows us to triangulate findings, providing a more reliable and comprehensive view of the texts. The integration of various metrics reveals patterns and relationships that might be overlooked when using a single method. For example, while Euclidean distance highlights overall dissimilarities, cosine similarity, and Jaccard similarity focus on thematic and pattern overlaps, offering a balanced perspective. Ultimately, this study illustrates the value of combining quantitative and qualitative approaches in textual analysis. It shows that by leveraging multiple methods, researchers can uncover richer, more nuanced insights into ancient texts, enhancing our understanding of their historical and literary contexts.

Author contributions: Alessandra Luce researched and compared the different statistical methods, ran the primary statistics, and drafted the majority of the paper. Paul Robertson provided the initial conceptual underpinnings, aided in the drafting and editing of Luce's sections, and drafted and edited other substantial sections of the paper.

Conflict of interest: Authors state no conflict of interest.

References

Al-Anazi, Sumayia, Hind AlMahmoud, and Isra Al-Turaiki. "Finding Similar Documents using Different Clustering Techniques." *Procedia Computer Science* 82 (2016), 28–34. doi: 10.1016/j.procs.2016.04.005.

Ale (https://stats.stackexchange.com/users/152762/ale). *Drawbacks with Cosine Similarity*. Cross Validated, 2017, March 12. https://stats.stackexchange.com/q/266979.

Argamon, Shlomo. "Interpreting Burrows's Delta: Geometric and Probabilistic Foundations." *Literary and Linguistic Computing* 23:2 (2007), 131–47. doi: 10.1093/llc/fqn003.

Al-Hagery, Mohammed Abdullah Hassan. "Google Search Filter using Cosine Similarity Measure to Find All Relevant Documents of a Specific Research." *ResearchGate*, 2016. https://www.researchgate.net/publication/309312557_Google_Search_Filter_Using_Cosine_Similarity_Measure_to_Find_All_Relevant_Documents_of_a_Specific_Research_Topic.

Evert, Stefan, Thomas Proisl, Fotis Jannidis, Isabella Reger, Steffen Pielström, Christof Schöch, and Thorsten Vitt. "Understanding and Explaining Delta Measures for Authorship Attribution." *Digital Scholarship in the Humanities* 32:suppl_2 (2017), ii4–16. doi: 10.1093/llc/fqx023.

Liberti, Leo, Carlile Lavor, Nelson Maculan, and Antonio Mucherino. "Euclidean Distance Geometry and Applications." *SIAM Review* 56:1 (2014), 3–69. doi: 10.1137/120875909.

Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, Édouard Duchesnay. *SciKit-Learn: Machine Learning in Python*, 2011. https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html.

Robertson, Paul Mark. *Paul's Letters and Contemporary Greco-Roman Literature: Theorizing a New Taxonomy*. Leiden: Brill, 2016.

Shtovba, Serhiy and Mykola Petrychko. "Jaccard index-Based Assessing the Similarity of Research Fields in Dimensions." *ResearchGate*, 2019. https://www.researchgate.net/publication/338458200_Jaccard_index-Based_Assessing_the_Similarity_of_Research_Fields_in_Dimensions.

Suwanda, Rizki, Zulfahmi Syahputra, and Elvi M. Zamzami. "Analysis of Euclidean distance and Manhattan Distance in the K-Means Algorithm for Variations Number of Centroid K." *Journal of Physics. Conference Series* 1566:1 (2020), 012058. doi: 10.1088/1742-6596/1566/1/012058.

Appendix A

Euclidean Distance

What it provides: Euclidean distance measures the "straight-line" distance between two points in a multi-dimensional space. It is a straightforward, intuitive way to understand how similar or dissimilar two texts are based on their feature frequencies.

Why use it: It is widely used in various fields such as pattern recognition, machine learning, and bioinformatics due to its simplicity and interpretability. Ideal for continuous data where the magnitude of differences is important.

Where it is used: Clustering algorithms like k-means. Nearest neighbor searches.

How it calculates the score: By taking the square root of the sum of the squared differences between corresponding feature values in two vectors (texts).

Why it finds different things: Euclidean distance emphasizes the absolute differences in feature magnitudes, making it sensitive to large variations in any feature. It doesn't consider the direction of differences, only their magnitude, so texts with similar feature distributions but different scales might appear very different.

Cosine Similarity

What it provides: cosine similarity measures the cosine of the angle between two vectors in a multi-dimensional space, reflecting the orientation rather than the magnitude.

Why use it: Effective for text analysis where the focus is on the similarity in the direction of feature vectors, regardless of their magnitude. Useful when dealing with high-dimensional sparse data, common in text mining.

Where it is used: Information retrieval and text mining. Document clustering and classification.

How it calculates the score: By dividing the dot product of two vectors by the product of their magnitudes.

Why it finds different things: cosine similarity captures the similarity in the shape of the feature distribution, making it robust to differences in text length. It is less affected by the absolute number of features, focusing instead on the relative distribution.

Jaccard Similarity

What it provides: Jaccard similarity measures the similarity between two sets, calculated as the size of the intersection divided by the size of the union of the sets.

Why use it: It is particularly useful for binary or categorical data where you want to know how many features are shared between two texts. Helps in scenarios where the presence or absence of features is more significant than their frequency.

Where it is used: Set-based comparisons, such as in market basket analysis and clustering.

How it calculates the score: By dividing the intersection of the sets (common features) by the union of the sets (total features present in either text).

Why it finds different things: Jaccard similarity focuses on shared features, ignoring how often they occur. Suitable for tasks where the overlap of features, rather than their frequency, is crucial.

Manhattan Distance

What it provides: Manhattan distance, also known as L1 distance or city block distance, measures the sum of the absolute differences between the features of two texts.

Why use it: Simple and intuitive, often used when data can have outliers or when the absolute differences are meaningful. It is robust to small changes in feature values and useful for categorical and ordinal data.

Where it is used: Various machine learning algorithms and optimization problems. Clustering and nearest neighbor searches in grid-based systems.

How it calculates the score: By summing the absolute differences between corresponding features in two vectors (texts).

Why it finds different things: Manhattan distance is more sensitive to individual feature differences than Euclidean distance. It is less affected by large differences in a single feature, providing a more balanced measure of overall dissimilarity.

Summary of Why Different Metrics Find Different Things

- **Data Emphasis:** Different metrics emphasize different aspects of the data. Euclidean distance emphasizes overall magnitude, while cosine similarity focuses on distribution patterns, Jaccard similarity on shared features, and Manhattan Distance on absolute differences.
- **Sensitivity to Scale:** Euclidean distance is sensitive to the scale of features, whereas cosine similarity is not. This can lead to different interpretations of similarity.
- **Feature Presence vs Frequency:** Jaccard similarity considers only the presence of features, ignoring their frequency, which contrasts with the other metrics that consider both presence and frequency.
- **Handling of Variability:** Manhattan distance handles variability differently than Euclidean distance, being less sensitive to large differences in any single feature, which can lead to different results in texts with outlier features.

Appendix B

Euclidean Distance

---

1 Thess & Phil: 11.0
1 Thess & Philem: 15.556
Philem & Phil: 16.155
1 Thess & Gal: 17.493
4 Macc & Gal: 18.385
Philem & Philo Death: 18.466
Gal & Phil: 19.875
1 Thess & Philo Death: 20.518
Gal & Seneca: 21.237
1 Thess & 2 Thess: 21.817
Phil & Philo Death: 22.136
2 Thess & Philem: 22.316
2 Cor & Gal: 22.517
2 Thess & Phil: 23.087
4 Macc & Philo Death: 23.302
4 Macc & Seneca: 24.434
2 Thess & Gal: 24.454
1 Thess & 4 Macc: 24.576
Gal & Philo Death: 26.134
Gal & Philem: 27.055
4 Macc & Phil: 27.875
2 Thess & Philo Death: 28.688
2 Cor & Phil: 29.155
2 Thess & Seneca: 29.411
1 Thess & Seneca: 30.545
2 Cor & 1 Thess: 30.545
2 Cor & Romans: 31.733
2 Cor & 4 Macc: 31.859
4 Macc & Philem: 32.558
Phil & Seneca: 32.68
Philo Death & Seneca: 33.347
2 Cor & Seneca: 34.351

1 Thess & Damascus: 56.089
Philo Piety & Seneca: 56.107
Arist Pan & Gal: 56.205
Arist Pan & Romans: 56.93
Arist Pan & Seneca: 57.637
Phil & Damascus: 57.922
1 Cor & 2 Cor: 58.898
Damascus & Hebrews: 59.515
Philem & Philo Piety: 60.175
2 Cor & Hebrews: 60.374
Romans & Philem: 60.63
2 Cor & Arist Pan: 61.27
2 Thess & Philo Piety: 61.879
Philo Death & Damascus: 62.442
Gal & Hebrews: 62.738
Seneca & Hebrews: 62.746
4 Macc & Hebrews: 63.166
Seneca & Damascus: 63.348
1 Cor & Seneca: 63.443
1 Thess & Arist Pan: 63.585
Arist Pan & Damascus: 64.117
Arist Pan & Philo Death: 64.234
Arist Pan & Philo Piety: 64.498
Arist Pan & Hebrews: 64.645
1 Cor & Gal: 64.761
Arist Pan & Phil: 66.257
Philem & Damascus: 67.335
2 Thess & Damascus: 67.35
Philo Piety & Hebrews: 68.666
1 Cor & 4 Macc: 69.34
2 Thess & Arist Pan: 70.05
1 Cor & Arist Pan: 70.534

---

2 Thess & 4 Macc: 34.756

Philem & Seneca: 36.565

Gal & Romans: 37.31

2 Cor & 2 Thess: 40.286

Romans & Seneca: 40.361

2 Cor & Philem: 40.583

2 Cor & Philo Death: 40.841

4 Macc & Romans: 41.061

4 Macc & Philo Piety: 41.557

Romans & Hebrews: 42.308

1 Cor & Romans: 42.919

1 Cor & Hebrews: 45.188

4 Macc & Arist Pan: 45.266

Philo Death & Philo Piety: 46.195

2 Cor & Damascus: 46.765

Philo Piety & Damascus: 47.149

Gal & Philo Piety: 48.425

Gal & Damascus: 50.695

4 Macc & Damascus: 51.498

1 Thess & Romans: 51.595

Romans & Phil: 51.682

Romans & Damascus: 51.807

Romans & Philo Piety: 52.393

1 Thess & Philo Piety: 53.038

2 Cor & Philo Piety: 53.74

2 Thess & Romans: 53.889

Phil & Philo Piety: 54.406

Romans & Philo Death: 55.651

Arist Pan & Philem: 71.561

1 Cor & Damascus: 72.979

1 Thess & Hebrews: 74.445

2 Thess & Hebrews: 76.105

Phil & Hebrews: 76.44

1 Cor & Phil: 78.899

1 Cor & 2 Thess: 79.145

1 Cor & Philo Piety: 79.253

1 Cor & 1 Thess: 79.272

Philo Death & Hebrews: 80.753

1 Cor & Philo Death: 85.761

Philem & Hebrews: 86.983

1 Cor & Philem: 89.32

1 Cor & Epict Disc: 522.208

Epict Disc & Hebrews: 534.134

Epict Disc & Romans: 551.57

Arist Pan & Epict Disc: 557.525

Epict Disc & Damascus: 558.563

Epict Disc & Seneca: 560.835

Epict Disc & Philo Piety: 563.052

2 Cor & Epict Disc: 564.282

4 Macc & Epict Disc: 566.762

Epict Disc & Gal: 569.678

1 Thess & Epict Disc: 581.917

Epict Disc & Philo Death: 582.27

Epict Disc & Phil: 582.421

2 Thess & Epict Disc: 584.33

Epict Disc & Philem: 590.628

Cosine Similarity

1 Cor & Gal: 19.06 dg

Gal & Hebrews: 21.37 dg

Gal & Romans: 22.84 dg

1 Cor & Romans: 24.2 dg

Romans & Hebrews: 25.71 dg

4 Macc & Arist Pan: 27.48 dg

2 Cor & Romans: 28.0 dg

1 Cor & Hebrews: 29.09 dg

2 Cor & Gal: 30.12 dg

1 Cor & 2 Cor: 30.54 dg

Philo Death & Philo Piety: 30.76 dg

1 Thess & Hebrews: 31.9 dg

4 Macc & Hebrews: 32.91 dg

4 Macc & Gal: 33.41 dg

Epict Disc & Seneca: 33.5 dg

1 Cor & Seneca: 33.82 dg

Gal & Seneca: 34.11 dg

1 Thess & Gal: 34.8 dg

2 Cor & Phil: 35.51 dg

1 Thess & Phil: 35.79 dg

2 Thess & Seneca: 51.45 dg

2 Thess & Hebrews: 51.56 dg

1 Thess & Philem: 51.76 dg

1 Cor & Damascus: 52.04 dg

1 Cor & Philem: 52.08 dg

Arist Pan & Seneca: 52.55 dg

1 Cor & 2 Thess: 52.89 dg

1 Thess & Seneca: 53.11 dg

1 Thess & Arist Pan: 54.0 dg

Arist Pan & Damascus: 54.03 dg

1 Thess & Philo Piety: 54.11 dg

1 Thess & Epict Disc: 54.28 dg

2 Thess & Gal: 54.8 dg

Gal & Philem: 56.01 dg

4 Macc & Phil: 56.67 dg

Arist Pan & Philo Piety: 57.11 dg

2 Cor & Arist Pan: 57.93 dg

2 Cor & Philem: 57.96 dg

Arist Pan & Philo Death: 57.97 dg

Epict Disc & Damascus: 58.16 dg

2 Cor & Hebrews: 36.87 dg

Seneca & Hebrews: 37.06 dg

1 Thess & Damascus: 37.2 dg

1 Cor & Epict Disc: 37.27 dg

4 Macc & Romans: 37.76 dg

Epict Disc & Gal: 38.46 dg

2 Cor & 1 Thess: 38.47 dg

Philem & Phil: 38.77 dg

Romans & Seneca: 38.91 dg

4 Macc & Seneca: 39.72 dg

4 Macc & Philo Piety: 39.82 dg

4 Macc & Epict Disc: 40.26 dg

Gal & Damascus: 40.68 dg

1 Cor & 4 Macc: 40.7 dg

2 Cor & Damascus: 41.75 dg

Philo Piety & Damascus: 41.99 dg

Damascus & Hebrews: 42.17 dg

1 Cor & 1 Thess: 42.44 dg

4 Macc & Philo Death: 42.52 dg

Gal & Phil: 43.6 dg

1 Cor & Phil: 43.81 dg

Phil & Hebrews: 44.28 dg

1 Thess & 4 Macc: 45.5 dg

4 Macc & Damascus: 45.84 dg

Epict Disc & Hebrews: 46.0 dg

Arist Pan & Gal: 46.03 dg

Arist Pan & Hebrews: 46.07 dg

1 Thess & Romans: 46.23 dg

Romans & Damascus: 46.29 dg

Phil & Damascus: 47.13 dg

Epict Disc & Romans: 47.2 dg

2 Cor & Epict Disc: 47.6 dg

2 Cor & 4 Macc: 48.26 dg

Romans & Phil: 48.51 dg

Arist Pan & Romans: 49.51 dg

1 Cor & Arist Pan: 49.53 dg

Romans & Philo Piety: 50.3 dg

Gal & Philo Piety: 50.56 dg

Philo Piety & Hebrews: 50.67 dg

2 Cor & Seneca: 50.68 dg

Philem & Hebrews: 58.44 dg

Arist Pan & Epict Disc: 58.51 dg

Epict Disc & Phil: 58.52 dg

Epict Disc & Philo Death: 58.56 dg

Philem & Damascus: 58.79 dg

1 Cor & Philo Piety: 58.84 dg

2 Cor & Philo Piety: 59.24 dg

2 Thess & Romans: 59.27 dg

Epict Disc & Philo Piety: 59.36 dg

Phil & Philo Piety: 59.84 dg

Philo Death & Hebrews: 59.84 dg

Phil & Seneca: 60.35 dg

2 Thess & Philem: 60.43 dg

Romans & Philem: 61.76 dg

Romans & Philo Death: 61.8 dg

1 Thess & 2 Thess: 62.21 dg

Philo Death & Seneca: 62.39 dg

Philo Death & Damascus: 62.48 dg

Gal & Philo Death: 63.01 dg

Philo Piety & Seneca: 64.0 dg

Arist Pan & Phil: 64.17 dg

2 Thess & Phil: 65.25 dg

Seneca & Damascus: 65.58 dg

Epict Disc & Philem: 66.67 dg

Philem & Philo Piety: 67.28 dg

1 Cor & Philo Death: 67.59 dg

Philem & Seneca: 67.59 dg

2 Cor & 2 Thess: 69.16 dg

1 Thess & Philo Death: 69.48 dg

2 Thess & Epict Disc: 70.47 dg

4 Macc & Philem: 71.96 dg

2 Thess & 4 Macc: 72.89 dg

2 Cor & Philo Death: 73.25 dg

Phil & Philo Death: 73.27 dg

2 Thess & Arist Pan: 75.4 dg

2 Thess & Damascus: 76.46 dg

Arist Pan & Philem: 77.19 dg

2 Thess & Philo Piety: 80.38 dg

Philem & Philo Death: 80.77 dg

2 Thess & Philo Death: 84.42 dg

Jaccard Similarity

1 Cor & Hebrews: 0.606

1 Cor & Romans: 0.58

4 Macc & Gal: 0.571

2 Cor & Gal: 0.555

Romans & Hebrews: 0.542

Gal & Seneca: 0.537

1 Thess & Phil: 0.513

Romans & Seneca: 0.512

Gal & Philo Death: 0.239

Philo Death & Philo Piety: 0.229

Romans & Phil: 0.229

1 Thess & Philo Death: 0.228

Seneca & Damascus: 0.218

Phil & Damascus: 0.217

2 Thess & Romans: 0.2

Philo Death & Seneca: 0.197

2 Cor & Romans: 0.512

4 Macc & Seneca: 0.494

Gal & Romans: 0.47

Damascus & Hebrews: 0.466

4 Macc & Romans: 0.462

2 Cor & 4 Macc: 0.459

1 Thess & Gal: 0.446

4 Macc & Arist Pan: 0.431

Romans & Damascus: 0.421

2 Cor & Seneca: 0.421

Philo Piety & Damascus: 0.406

4 Macc & Philo Piety: 0.402

1 Cor & 2 Cor: 0.393

2 Cor & Damascus: 0.389

2 Cor & Hebrews: 0.384

Arist Pan & Hebrews: 0.384

Gal & Phil: 0.381

2 Cor & 1 Thess: 0.379

1 Thess & 4 Macc: 0.376

2 Cor & Phil: 0.376

1 Cor & Arist Pan: 0.369

1 Cor & Damascus: 0.363

1 Cor & 4 Macc: 0.361

Arist Pan & Romans: 0.349

4 Macc & Philo Death: 0.348

Romans & Philo Piety: 0.347

1 Cor & Seneca: 0.347

4 Macc & Hebrews: 0.342

4 Macc & Damascus: 0.339

1 Cor & Gal: 0.325

Philo Piety & Hebrews: 0.321

Seneca & Hebrews: 0.319

Gal & Philo Piety: 0.319

Arist Pan & Gal: 0.317

1 Thess & Seneca: 0.312

Gal & Hebrews: 0.309

Arist Pan & Damascus: 0.307

Gal & Damascus: 0.307

2 Cor & Philo Piety: 0.304

Arist Pan & Philo Piety: 0.297

1 Cor & Philo Piety: 0.294

Arist Pan & Seneca: 0.293

Philo Piety & Seneca: 0.286

2 Thess & Gal: 0.282

1 Thess & 2 Thess: 0.275

4 Macc & Phil: 0.272

2 Thess & Phil: 0.272

2 Thess & Seneca: 0.271

2 Cor & Arist Pan: 0.269

1 Thess & Romans: 0.255

Phil & Seneca: 0.247

1 Thess & Damascus: 0.242

1 Thess & Philo Piety: 0.195

1 Thess & Arist Pan: 0.19

2 Cor & 2 Thess: 0.183

Phil & Philo Death: 0.181

Romans & Philo Death: 0.18

1 Cor & 1 Thess: 0.174

1 Thess & Hebrews: 0.173

1 Cor & Phil: 0.17

Phil & Philo Piety: 0.168

Phil & Hebrews: 0.168

2 Thess & 4 Macc: 0.168

Arist Pan & Philo Death: 0.164

Arist Pan & Phil: 0.161

2 Thess & Philem: 0.156

2 Thess & Hebrews: 0.152

2 Thess & Philo Death: 0.149

1 Cor & Epict Disc: 0.147

1 Cor & 2 Thess: 0.147

2 Cor & Philo Death: 0.143

Epict Disc & Hebrews: 0.139

Philo Death & Hebrews: 0.134

Philo Death & Damascus: 0.128

2 Thess & Arist Pan: 0.127

1 Cor & Philo Death: 0.127

1 Thess & Philem: 0.123

Philem & Phil: 0.121

2 Thess & Damascus: 0.099

Arist Pan & Epict Disc: 0.094

Epict Disc & Damascus: 0.091

Epict Disc & Romans: 0.091

2 Thess & Philo Piety: 0.083

Epict Disc & Philo Piety: 0.079

Gal & Philem: 0.067

4 Macc & Philem: 0.061

2 Cor & Epict Disc: 0.061

Epict Disc & Seneca: 0.06

4 Macc & Epict Disc: 0.056

2 Cor & Philem: 0.051

Epict Disc & Gal: 0.05

Philem & Seneca: 0.048

Philem & Philo Death: 0.044

Romans & Philem: 0.037

Philem & Damascus: 0.037

1 Thess & Epict Disc: 0.026

Arist Pan & Philem: 0.026

Epict Disc & Phil: 0.025

Philem & Hebrews: 0.024

Philem & Philo Piety: 0.024

1 Cor & Philem: 0.023

2 Thess & Epict Disc: 0.021

Epict Disc & Philo Death: 0.019

Epict Disc & Philem: 0.003

Manhattan Distance

1 Thess & Phil: 37
2 Thess & Philem: 38
Philem & Philo Death: 43
1 Thess & Philem: 50
Philem & Phil: 51
1 Thess & 2 Thess: 58
2 Thess & Phil: 59
4 Macc & Gal: 60
1 Thess & Philo Death: 61
1 Thess & Gal: 62
2 Thess & Philo Death: 63
Phil & Philo Death: 68
Gal & Seneca: 69
2 Cor & Gal: 69
Gal & Phil: 73
4 Macc & Philo Death: 75
1 Thess & 4 Macc: 78
4 Macc & Seneca: 81
2 Thess & Gal: 84
2 Cor & 1 Thess: 87
2 Cor & Phil: 88
Gal & Philo Death: 89
2 Cor & 4 Macc: 93
1 Thess & Seneca: 95
2 Thess & Seneca: 97
Gal & Philem: 98
4 Macc & Phil: 99
Romans & Seneca: 101
2 Cor & Romans: 105
2 Cor & Seneca: 106
Gal & Romans: 106
4 Macc & Philem: 108
Philo Death & Seneca: 110
Phil & Seneca: 110
4 Macc & Romans: 112
2 Thess & 4 Macc: 114
4 Macc & Philo Piety: 119
Philem & Seneca: 119
4 Macc & Arist Pan: 123
2 Cor & 2 Thess: 125
Philo Death & Philo Piety: 128
2 Cor & Philem: 129
1 Cor & Romans: 132
2 Cor & Philo Death: 132
Gal & Philo Piety: 139
Romans & Hebrews: 142
2 Cor & Damascus: 143
1 Cor & Hebrews: 146

Phil & Damascus: 159
2 Cor & Philo Piety: 160
Philo Piety & Seneca: 160
Philem & Philo Piety: 163
Arist Pan & Philo Death: 168
Romans & Philo Piety: 171
1 Thess & Arist Pan: 171
Arist Pan & Seneca: 174
Damascus & Hebrews: 174
Philo Death & Damascus: 177
2 Thess & Philo Piety: 177
Arist Pan & Phil: 182
Philem & Damascus: 182
Romans & Philem: 182
Arist Pan & Romans: 185
2 Thess & Arist Pan: 185
2 Cor & Hebrews: 189
2 Cor & Arist Pan: 190
Arist Pan & Philem: 191
2 Thess & Damascus: 192
1 Cor & 2 Cor: 193
Arist Pan & Philo Piety: 194
4 Macc & Hebrews: 198
1 Cor & 4 Macc: 198
Seneca & Damascus: 201
Arist Pan & Damascus: 203
Gal & Hebrews: 208
1 Cor & Seneca: 209
1 Cor & Gal: 210
Seneca & Hebrews: 213
Arist Pan & Hebrews: 215
1 Cor & Arist Pan: 231
1 Cor & Damascus: 232
Philo Piety & Hebrews: 233
1 Thess & Hebrews: 244
2 Thess & Hebrews: 246
Phil & Hebrews: 247
Philo Death & Hebrews: 251
1 Cor & 1 Thess: 256
1 Cor & Philo Piety: 257
1 Cor & Phil: 259
1 Cor & 2 Thess: 262
1 Cor & Philo Death: 269
Philem & Hebrews: 282
1 Cor & Philem: 300
1 Cor & Epict Disc: 1769
Epict Disc & Hebrews: 1783
Arist Pan & Epict Disc: 1878

1 Thess & Romans: 146

1 Thess & Philo Piety: 149

Philo Piety & Damascus: 149

1 Thess & Damascus: 150

4 Macc & Damascus: 150

Romans & Damascus: 154

Romans & Phil: 155

Arist Pan & Gal: 155

Gal & Damascus: 156

2 Thess & Romans: 156

Phil & Philo Piety: 158

Romans & Philo Death: 159

Epict Disc & Damascus: 1883

Epict Disc & Romans: 1885

Epict Disc & Philo Piety: 1908

Epict Disc & Seneca: 1948

2 Cor & Epict Disc: 1956

4 Macc & Epict Disc: 1957

Epict Disc & Gal: 1969

1 Thess & Epict Disc: 2023

Epict Disc & Phil: 2028

2 Thess & Epict Disc: 2029

Epict Disc & Philo Death: 2032

Epict Disc & Philem: 2065

Burrows' Delta Method

Romans & Hebrews: 0.464

1 Cor & Gal: 0.512

1 Cor & Romans: 0.556

Gal & Romans: 0.656

1 Cor & Seneca: 0.66

2 Thess & Hebrews: 0.661

Damascus & Hebrews: 0.666

1 Cor & Hebrews: 0.677

4 Macc & Hebrews: 0.694

4 Macc & Gal: 0.704

Gal & Hebrews: 0.71

Philo Death & Philo Piety: 0.724

2 Cor & Gal: 0.726

1 Cor & 2 Cor: 0.731

2 Cor & Romans: 0.731

4 Macc & Romans: 0.742

1 Thess & Hebrews: 0.742

4 Macc & Arist Pan: 0.771

Romans & Seneca: 0.779

4 Macc & Philo Piety: 0.786

Phil & Hebrews: 0.792

Philem & Phil: 0.801

1 Cor & 4 Macc: 0.818

2 Cor & 4 Macc: 0.831

4 Macc & Philo Death: 0.834

1 Thess & Phil: 0.837

2 Cor & 1 Thess: 0.838

1 Cor & 1 Thess: 0.84

1 Thess & Damascus: 0.844

2 Thess & Phil: 0.846

2 Thess & Romans: 0.855

4 Macc & Seneca: 0.857

2 Cor & Hebrews: 0.857

Gal & Seneca: 0.875

4 Macc & Damascus: 0.876

Romans & Damascus: 0.877

Gal & Philo Piety: 1.026

1 Cor & Damascus: 1.045

1 Thess & 2 Thess: 1.049

2 Thess & Philem: 1.05

Arist Pan & Damascus: 1.067

2 Cor & Epict Disc: 1.067

2 Thess & Philo Death: 1.068

Epict Disc & Gal: 1.069

Phil & Philo Piety: 1.071

2 Thess & Gal: 1.081

Gal & Phil: 1.084

1 Cor & Arist Pan: 1.084

Philo Death & Seneca: 1.089

Gal & Philo Death: 1.097

2 Thess & Seneca: 1.101

1 Thess & Seneca: 1.113

1 Cor & Philo Death: 1.117

2 Thess & 4 Macc: 1.137

2 Cor & 2 Thess: 1.138

4 Macc & Phil: 1.15

Arist Pan & Philo Death: 1.15

1 Thess & Arist Pan: 1.156

2 Thess & Damascus: 1.163

2 Thess & Philo Piety: 1.166

Philo Piety & Seneca: 1.18

Romans & Philem: 1.181

Philem & Hebrews: 1.183

1 Cor & Philem: 1.183

Philem & Damascus: 1.19

1 Thess & Philo Piety: 1.201

2 Cor & Philo Piety: 1.201

Epict Disc & Hebrews: 1.203

Epict Disc & Philo Death: 1.221

Arist Pan & Gal: 1.231

Epict Disc & Romans: 1.232

Arist Pan & Phil: 1.254

Epict Disc & Seneca: 0.88

Seneca & Hebrews: 0.91

Philo Piety & Damascus: 0.92

Romans & Phil: 0.926

1 Cor & Phil: 0.926

1 Cor & Epict Disc: 0.927

Philo Death & Hebrews: 0.928

Romans & Philo Death: 0.942

4 Macc & Epict Disc: 0.944

Arist Pan & Hebrews: 0.947

2 Cor & Phil: 0.948

Romans & Philo Piety: 0.966

Philo Death & Damascus: 0.971

1 Thess & Gal: 0.976

1 Cor & 2 Thess: 0.979

1 Thess & Romans: 0.99

1 Cor & Philo Piety: 0.993

Arist Pan & Romans: 0.995

2 Cor & Damascus: 0.995

Philo Piety & Hebrews: 0.997

Phil & Damascus: 1.002

1 Thess & 4 Macc: 1.009

2 Cor & Seneca: 1.011

Gal & Damascus: 1.013

2 Thess & Arist Pan: 1.264

Philem & Philo Piety: 1.264

Arist Pan & Philo Piety: 1.272

Phil & Philo Death: 1.28

Arist Pan & Seneca: 1.282

Epict Disc & Philo Piety: 1.296

1 Thess & Epict Disc: 1.304

2 Cor & Arist Pan: 1.304

1 Thess & Philem: 1.313

1 Thess & Philo Death: 1.331

Phil & Seneca: 1.34

Epict Disc & Damascus: 1.369

Philem & Philo Death: 1.37

2 Cor & Philo Death: 1.372

Arist Pan & Philem: 1.389

Seneca & Damascus: 1.396

2 Thess & Epict Disc: 1.408

Gal & Philem: 1.434

2 Cor & Philem: 1.514

Epict Disc & Phil: 1.524

Arist Pan & Epict Disc: 1.542

4 Macc & Philem: 1.554

Philem & Seneca: 1.625

Epict Disc & Philem: 1.85