A Lexicographic Analysis of Similarities in the Pauline Epistles

Jacob H. Nie

April 27, 2020

Contents

Disclaimer 1: This was mostly a personal project to practice my programming skills. I was surprised by the results, especially since I did not need to optimize anything to attain them. You may be equally surprised! However, I did not do any research into how other people have analyzed the New Testament with computer methods, so it is possible that my methods are either too simple or too wrong to be considered any kind of relevant contribution to the complex, divided field of New Testament studies.

Disclaimer 2: Any information about current scholarly consensus should be taken with a grain of salt. First of all, most outside information in this paper came from Wikipedia. At the very best, it's a bad research practice and at the very worst, the information is wrong. Also, please note that both Wikipedia and modern critical analysis are skewed against conservative religious thought due to the general overrepresentation of progressive scholarship in universities. Modern scholars are extremely quick to reject claims to authorship on the basis of occasionally minute details.

Disclaimer 3: For the Christian, rejecting the teachings of certain books because they are pseudepigraphic is dangerous business. Many times, they are considered pseudepigraphic largely in part due to the presence of "inconvenient" doctrines (misogyny is the biggest culprit) or simply because it seems different. "If we lose inerrancy, then it becomes a very slippery slope then to making the Bible be whatever we want it to be. We just say that everything we don't like is not authentic. In that case the Bible is no longer the Word of God but just a source to defend what you already believe." –Larry Lin

With that said, let's dive right in.

1   Introduction

New Testament authorship is a tricky subject. Most books have claims to authorship (in other words, the author directly writes, "It is me, Jacob, who wrote this"), but other books have either no claims, or only indirect claims (like in John), to authorship. And even among the books with claims to authorship, many modern scholars now dispute these claims, suggesting that these books were actually written by a later person pretending to be the claimed author. (The scholarly term is "pseudepigraphical.")

Here's the general consensus among modern day scholars.

1.1   Modern Scholarship

We can split up the New Testament into five sections: the four Gospels (each giving an account of the ministry, crucifixion, and resurrection of Christ), the Book of Acts (a history of the early church following the resurrection), the Pauline Epistles (letters written by Paul to churches and specific people), the Catholic Epistles (letters written to the "catholic", or general, church not to be confused with the modern day Roman Catholic Church), and the Book of Revelation (a book containing end-time prophecies.)

The Gospels are ascribed to the authors named in the title. Among the four Gospels, we call the first three (Matthew, Mark, and Luke) the Synoptic Gospels. The fourth, John,

reads somewhat differently from the first three.

The Book of Acts does not claim an author, but most scholars agree that it is written by the same person who wrote Luke (who we think is Luke but very well might not be).

The Epistles are a complex territory. Seven letters are considered genuine works by Paul: 1 Thessalonians, Galatians, 1 Corinthians, Philippians, Philemon, 2 Corinthians, and Romans. Scholars are split on the authenticity of two letters by Paul: Colossians and 2 Thessalonians. Most scholars believe four letters are "fakes" (pseudepigraphic): Ephesians, 1 Timothy, 2 Timothy, and Titus. Hebrews is anonymous but is usually categorized with the Pauline Epistles, but nobody thinks it's written by Paul, not even the church.

And the Catholic Epistles. James is written by somebody named James, but there are a couple of James's. The opinion is split. 1 Peter is very contested as well. (Peter was a fisherman, how could he write so well?? But there is evidence for it still.) 2 Peter is generally considered to be written by somebody other than Peter. 1, 2, and 3 John are generally considered to be written by the same person who wrote the Gospel of John (but not necessarily John!). Finally, Jude is pretty disputed.

The Book of Revelation is attributed to John, and traditionally this John has been thought to be the same John that wrote the Gospel of John, but now this is disputed due to significant stylistic differences.

1.2   My analysis

For our analysis, we will only consider the Pauline epistles. Trying to form any claim about authorship in other books is impossible because there aren't any books written by the same author that can be compared to. For example, it would be pointless to try to determine whether James wrote James from a purely stylistic analysis: we have no other writings by James to compare it to.

Why can't we examine the Johannine literature? Because comparing the books of John and 1, 2, 3 John and Revelation is like comparing apples to oranges. The books have different purposes and different audiences—any difference in language would not speak much. Furthermore, the sizes of the Johannine books vary too widely. (Yes, size does matter! Refer to the discussion about Philemon below.) Nevertheless, you can try and form your own claims about the authorship of Johannine books by taking a look at Appendix B.

So to summarize, the big question is: *Which Pauline Epistles were actually written by Paul?*

2   Method

Our method is pretty simple (almost too simple) and only involves looking at the word frequencies in the original Greek. (No examination of grammatical structures or word combinations.) Here is the general structure of our method.

2.1 Summary

First, we determine the frequencies of each Greek word as it appears in each book. (We use root words. For example, in English we would consider *run*, *ran*, *running*, and *runs* to be identical.)

Then, to compare two books, we will have something that I call the *degree of dissimilarity*. If one word has a high frequency in the first text but a low frequency in the second text, then this will add significantly to the degree of dissimilarity. If another word appears equally frequently or infrequently in both texts, then this will add zero to the degree of dissimilarity.

Note: we will limit ourselves to the most common words (the 10 most common in each book). Here is why. We are looking for stylistic differences and not topical differences. For example, Paul talks about circumcision often in Romans but never in 2 Corinthians. But that should not discredit the stylistic similarities between the two books! So we stick to the 10 most common words in each book, which happen to be mostly prepositions, conjunctions, and pronouns. (Refer to Appendix A for a table of the 20 most common words in each book, along with the frequencies.)

One might think, how could you possibly expect to glean any kind of results simply by looking at the frequencies of the most insignificant words like "is," "but," and "the"??? I know it sounds ridiculous, but you might be surprised.

2.2 A bit more detail

Now I will explain the method in more detail.

Let $\{m_i\}$ be the set of book M's top 10 words. Let $\{n_i\}$ be the set of book N's top 10 words. Then let $\{w_i\}$ be the union of $\{m_i\}$ and $\{n_i\}$. (The size of $\{w_i\}$ is anywhere between 10 and 20. This skews our final number, but in the right direction. You'll see.)

Let $f_m(w_i)$ be the frequency of word w_i in book M (occurrences divided by total words). Let $f_n(w_i)$ be the frequency of word w_i in book N. Note that it is possible, but unlikely, that word w_i doesn't appear in one of the books. In that case, we take note of $\{w_i\}$ but omit it from the set. (We can't do $\ln 0$ as you will see.)

My formula for the degree of dissimilarity, a, is:

$$a = \sum s_i = \sum |\ln f_m(w_i) - \ln f_n(w_i)|.$$

A small a means the two books are similar, a large a means the two books are not similar. Unfortunately, we can't look at a specific value of a and determine whether this means the books are similar or dissimilar. We can only look at two pairs of books and determine which pair is more similar.

Note: I use a logarithm to give proper weight to less common words. For example, if one word has $f_m(w_i) = 0.15$ and $f_n(w_i) = 0.13$ and another word has $f_m(w_j) = 0.02$ and $f_n(w_j) = 0.001$, then the absence of a logarithm would make these two words contribute equally to a: $|f_m(w_i) - f_n(w_i)| = |f_m(w_j) - f_n(w_j)|$. However, it's clear that w_j shows the largest, most significant stylistic difference. The logarithm solves that issue.

2.3 A clarifying example

I'll make this in English and not Greek so it's a bit more clear. And also, instead of the top ten words in each book, we'll take the top three words instead. And the two books we compare will be: *The Catcher in the Rye* and *The Old Man and the Sea.* (Data is fake, if that's not clear.)

The Catcher in the Rye	$f_1(w)$	*The Old Man and the Sea*	$f_2(w)$
the	0.15	the	0.14
and	0.07	for	0.06
is	0.02	and	0.03
$\vdots$	$\vdots$	$\vdots$	$\vdots$
for	0.01	is	0.005

Using the same notation as we did above, $\{m_i\} = \{\text{the, and, is}\}$, and $\{n_i\} = \{\text{the, for, and}\}$. Then $\{w_i\} = \{\text{the, and, is, for}\} = \{w_1, w_2, w_3, w_4\}$.

We calculate:

$$s_1 = |\ln 0.15 - \ln 0.14| = 0.069$$
$$s_2 = |\ln 0.07 - \ln 0.03| = 0.847$$
$$s_3 = |\ln 0.02 - \ln 0.005| = 1.386$$
$$s_4 = |\ln 0.01 - \ln 0.06| = 1.792$$

Hence, our degree of dissimilarity is $a = 0.069 + 0.847 + 1.386 + 1.792 = 4.094$. The word that contributes most, the word that has the largest discrepancy, is "for".

2.4 Finding similar works in a group of works

(This section explains what the function **mostSimilarIn** (Section 3) does.)

Suppose we have five books, (A, B, C, D, and E). We've calculated the a-values for each pair of these five books, and we want to find the three out of the five books that are most similar to one another.

How do we do this? We test each possible size 3 subset of the five books (ABC, ABD, ABE, etc) and see which one has the lowest "total a-value." In other words, which subset has the lowest sum of all the pairwise a-values. Then, this subset is the subset of books that is most similar to one another.

In the program I wrote, you would invoke:

```
mostSimilarIn([A, B, C, D, E], 3)
>>> ([A, C, D], 5.313151)
```

where the second number is the average pairwise a value. (As before, lower means more similar!)

3 Computer Work

I used Python 3. I downloaded a .csv file of the entire Greek New Testament (each line containing one word and its Strong's Number along with morphological tags that I ignored) from <u>here</u>.

Next, I split up the .csv file into books. Each book had a corresponding 1-D array of the words that it contained.

Following that, I wrote a function that would return, in array form, each unique word along with the frequency with which it occurred in the book, ordered from most common to least common. I also created a dictionary for each book that would allow me to quickly access the frequency of each word.

Then, I wrote a function (**compare1**) to return the a value for any two input books, along with the five words that contributed most significantly to a (i.e. the words for which $s_i = |\ln f_m(w_i) - \ln f_n(w_i)|$ was the largest).

Next, I wrote a function (**mostSimilarTo**) that would take in an input book and return the a values for it vs. the 26 other books, then ranking which books were most similar and most dissimilar to the input book. (Results in Appendix B)

Finally, I wrote a function (**mostSimilarIn**) that would take a sublist of n books and return the $k \leq n$ most similar books within that sublist. The returned subset would be the subset of size k that minimized the sum of all $\binom{k}{2}$ degrees of dissimilarity between all pairs of books within that subset. Furthermore, the function also returns the average value of a (the degree of dissimilarity) within all book-pairs within the returned subset.

The complete Python code is found in Appendix C.

4 Results

Please see Appendix A and B for the raw results. Appendix A lists the top 20 most common words in each book, along with each word's frequency. Appendix B lists, for each book, the a values for that book and all 26 other books.

Let's try to see how our data answers our main question. Of course, we can't expect a clear, definite answer using only the limited analysis provided, but we can get a sense of how consistent our method is with current scholarship.

Which Pauline Epistles were actually written by Paul?

To answer this question, we will use the function **mostSimilarIn**, which is the same method as explained in Section 2.4.

We will examine the 13 Pauline Epistles, not including Hebrews. First, we will look for the two most similar works. Then, we will look for the three most similar works. (In other words, out of these 13 books we try every possible combination of three books and see which combination is the most similar to one another.) Hopefully, by the time we get to the five and six most similar works, a clear "in-group" of similar books will begin to take shape.

Towards the end, when we start looking for the 11 and 12 most similar books, the books that only just begin to appear are going to be the clear odd ones out.

If our developing "in-group" of similar books is consistent with what scholarship believes to be authentic Pauline literature, then we can estimate an approximate ranking of which books are "most Pauline" and "least Pauline."

The results are listed in the table below. (For your convenience, I have bolded the books that "have just appeared", i.e. the books that are not in the row directly above it.)

k	Most Similar k books	Average a (Average Degree of Dissimilarity)
2	Eph, Col	2.361
3	**Rom**, **1Cor**, **Gal**	3.358
4	Rom, 1Cor, **2Cor**, Gal	3.639
5	Rom, 1Cor, 2Cor, Gal, **2Tim**	4.429
6	Rom, 2Cor, **Phil**, **1Thess**, **2Thess**, 2Tim	4.826
7	Rom, **1Cor**, 2Cor, Phil, 1Thess, 2Thess, 2Tim	5.204
8	Rom, 1Cor, 2Cor, **Gal**, Phil, 1Thess, 2Thess, 2Tim	5.464
9	Rom, 1Cor, 2Cor, Gal, **Eph**, Phil, 1Thess, 2Thess, 2Tim	5.804
10	Rom, 1Cor, 2Cor, Gal, Eph, Phil, **Col**, 1Thess, 2Thess, 2Tim	6.253
11	Rom, 1Cor, 2Cor, Gal, Eph, Phil, Col, 1Thess, 2Thess, 2Tim, **Phlm**	6.67
12	Rom, 1Cor, 2Cor, Gal, Eph, Phil, Col, 1Thess, 2Thess, 2Tim, **Titus**, Phlm	7.101
13	Rom, 1Cor, 2Cor, Gal, Eph, Phil, Col, 1Thess, 2Thess, **1Tim**, 2Tim, Titus, Phlm	7.410

From this, we can order (approximately) the 13 books from most Pauline to least Pauline: Romans, 1 Corinthians, Galatians, 2 Corinthians, 2 Timothy, Philippians, 1 Thessalonians, 2 Thessalonians, Ephesians, Colossians, Philemon, Titus, and 1 Timothy.

(Despite Ephesians and Colossians appearing first, I have chosen not to consider them most Pauline because they appear to only be similarly dissimilar. We can pretty safely do this because a very clear in-group began to develop in $k = 3, 4, 5$.)

4.1   Comparison to results of modern scholarship

As stated above, modern scholarship believes that Romans, 1 Corinthians, 2 Corinthians, Galatians, Philippians, 1 Thessalonians, and Philemon are authentic. Colossians and 2 Thessalonians are disputed, and Ephesians, 1 Timothy, 2 Timothy, and Titus are believed to be pseudepigraphic.

The agreement with modern scholarship is very remarkable, especially considering that these determinations come from only the frequencies of the top 10-15 most common words in the books (words that, in English, would be words like the, he, is, and, but, for, etc).

The notable outliers in our analysis are: 2 Timothy and Philemon.

4.1.1   2 Timothy

Our analysis shows that **2 Timothy is much much more Pauline than modern scholarship thinks it is**. Indeed, modern scholarship is much more divided on 2 Timothy than 1 Timothy and Titus, the two other pastoral epistles. **Our analysis may be very revealing.**

Often, modern scholarship discounts the Pauline authorship of 2 Timothy by grouping 1 Timothy, 2 Timothy, and Titus into one collection and discounting all three of them at the same time. However, our analysis shows significant differences between 2 Timothy and the other two books. (As you can see in Appendix B, 1 Timothy and Titus are similar to each other but very different from the Pauline epistles, whereas 2 Timothy is far closer to the "authentic" Pauline epistles than it is to 1 Timothy.)

These differences can be seen in the content of 2 Timothy, which aligns much better with more "traditional" Pauline concepts, compared to the overtly misognystic non-sola fide in 1 Timothy (2:15).

Our analysis maintains the claim that 1 Timothy and Titus are not Pauline works but provides significant evidence against grouping 2 Timothy with the two other Pastoral Epistles.

4.1.2   Philemon

Our analysis shows that **Philemon is much less Pauline than modern scholarship thinks it is**. Modern scholarship rarely questions the authorship of Philemon, so it appears that **our analysis may have erred in this case.** One possible reason is the book's shortness. Almost all Pauline epistles have a greeting, introduction, thanksgiving, and conclusion, which contains a different style than the main body of the epistle. Since Philemon is so short, the greeting, introduction, thanksgiving, and conclusion take up almost half of the entire epistle. This can significantly skew the frequency counts, thereby causing a possible erroneous result.

4.2   Disclaimers (again)

These are important disclaimers, so I will say them again.

Disclaimer 1: This was mostly a personal project to practice my programming skills. I was surprised by the results, especially since I did not need to optimize anything to attain them. You may be equally surprised! However, I did not do any research into how other people have analyzed the New Testament with computer methods, so it is possible that my methods are either too simple or too wrong to be considered any kind of relevant contribution to the complex, divided field of New Testament studies.

Disclaimer 2: Any information about current scholarly consensus should be taken with a grain of salt. First of all, most outside information in this paper came from Wikipedia. At the very best, it's a bad research practice and at the very worst, the information is wrong. Also, please note that both Wikipedia and modern critical analysis are skewed against conservative religious thought due to the general overrepresentation of progressive scholarship

in universities. Modern scholars are extremely quick to reject claims to authorship on the basis of occasionally minute details.

Disclaimer 3: For the Christian, rejecting the teachings of certain books because they are pseudepigraphic is dangerous business. Many times, they are considered pseudepigraphic largely in part due to the presence of "inconvenient" doctrines (misogyny is the biggest culprit) or simply because it seems different. "If we lose inerrancy, then it becomes a very slippery slope then to making the Bible be whatever we want it to be. We just say that everything we don't like is not authentic. In that case the Bible is no longer the Word of God but just a source to defend what you already believe." –Larry Lin

5 Sources

The Greek New Testament in csv format was downloaded from here.

I made use of this website to translate Strong's numbers to English (done in Appendix A).

Most of the information about New Testament scholarship was taken from these Wikipedia pages: New Testament, Authorship of the Pauline Epistles, Authorship of the Petrine Epistles, Second Epistle to Timothy, Epistle to Philemon, Authorship of the Johannine works.

In writing the section on 2 Timothy, I used these two articles from the New World Encyclopedia: Second Epistle to Timothy and First Epistle to Timothy.

A   Top Twenty Words in Each Book

Matt

Strong's No., English	Frequency
3588, this	0.15164763101808842
2532, and	0.0638832723099623
846, himself	0.04967484561997924
1161, but	0.026996010710967813
3004, say	0.025903054811738345
4771, you	0.02481009891250888
1772, in	0.016011803923711678
1510, to be	0.015738564948904313
1473, I	0.01420842668998306
1519, into	0.011858571506639708
3756, not	0.011093502377179081
2424, Jesus	0.008197169244220997
3778, this	0.007978578064375103
3754, that	0.007705339089567736
3708, see	0.007486747909721842
3361, not lest	0.007322804524837423
3956, all	0.006994917755068583
3739, who	0.006830974370184163
1063, for	0.00677632657522269
1909, upon	0.006667030985299743

Mark

Strong's No., English	Frequency
3588, this	0.1339895362241731
2532, and	0.09621353196772191
846, himself	0.0671277822115811
3004, say	0.025450031036623216
1510, to be	0.01711448080163164
1519, into	0.0148975791433892
4771, you	0.014542874878070409
1161, but	0.01445419881174071
1772, in	0.01223729715349827
1473, I	0.011439212556530992
3756, not	0.010641127959563713
3754, that	0.009133634831958854
3739, who	0.007714817770683693

2064, come	0.007626141704353995
2424, Jesus	0.007271437439035204
3778, this	0.0070054092400461111
3361, not lest	0.006916733173716414
5101, what	0.006384676775738229
1909, upon	0.006384676775738229
2192, to have	0.006207324643078833

Luke

Strong's No., English	Frequency
3588, this	0.13563939144736842
2532, and	0.07529810855263158
846, himself	0.05556126644736842
1161, but	0.027857730263157895
3004, say	0.0263671875
4771, you	0.022872121710526317
1510, to be	0.0185546875
1772, in	0.01845189144736842
1473, I	0.014494243421052632
3778, this	0.01177014802631579
1519, into	0.01166735197368421
3739, who	0.009765625
3756, not	0.008891858552631578
3754, that	0.008891858552631578
4314, forward	0.008532072368421052
1909, upon	0.008275082236842105
3956, all	0.008120888157894737
3361, not lest	0.007401315789473684
3708, see	0.007092927631578948
1096, made	0.006630345394736842

John

Strong's No., English	Frequency
3588, this	0.13910375247714632
2532, and	0.05312280253148373
846, himself	0.04845617848238829
1473, I	0.032666368343668095
3004, say	0.03042894585437576
1510, to be	0.028383302435594195

4771, you	0.02614587994630186
3756, not	0.017963306271175605
3754, that	0.017260116345969443
2424, Jesus	0.015534104711372499
3778, this	0.015342325640861727
1772, in	0.014447356645144793
1161, but	0.013360608578917087
3767, then	0.012657418653710926
1519, into	0.012018155085341686
1537, out from	0.010483922521255514
3739, who	0.010292143450744742
2064, come	0.010036438023397046
2443, in order that	0.00926932174135396
3962, father	0.008630058172984722

Acts

Strong's No., English	Frequency
3588, this	0.14603381721307018
2532, and	0.06013157179361714
846, himself	0.03767737726308922
1161, but	0.030337628445604305
1473, I	0.016636763986299135
1519, into	0.016310552927744252
1510, to be	0.015005708693524711
1772, in	0.015005708693524711
4771, you	0.014135812537378351
3778, this	0.01283096830315881
3004, say	0.012287283205567335
3739, who	0.012232914695808188
3956, all	0.009405752188332518
1909, upon	0.00913390963953678
2316, God	0.008970804110259337
5037, and	0.007937802424835536
4314, forward	0.007285380307725765
3972, Paul	0.006959169249170886
1096, made	0.006796063719893438
3754, that	0.006687326700375143

Rom

Strong's No., English	Frequency
3588, this	0.15535211267605634
2532, and	0.038450704225352114
1772, in	0.02436619718309859
846, himself	0.02211267605633803
2316, God	0.02140845070422535
1473, I	0.020985915492957745
1161, but	0.020704225352112676
1063, for	0.02028169014084507
4771, you	0.018028169014084508
3756, not	0.017183098591549296
1519, into	0.016760563380028169
1510, to be	0.01591549295774648
1223, through	0.012816901408450704
3739, who	0.012816901408450704
3361, not lest	0.011126760563380281
3551, law	0.010422535211267606
3956, all	0.009859154929577466
235, but	0.009718309859154929
5547, Christ	0.009154929577464789
1537, out from	0.008450704225352112

1Cor

Strong's No., English	Frequency
3588, this	0.12683284457478006
2532, and	0.04090909090909091
1161, but	0.030791788856304986
1772, in	0.024780058651026393
1510, to be	0.023607038123167154
3756, not	0.02287390029325513
4771, you	0.02258064516129032
1473, I	0.0187683284457478
3956, all	0.016422287390029325
1063, for	0.015542521994134898
2316, God	0.015395894428152493
3361, not lest	0.01407624633431085
846, himself	0.012609970674486803
235, but	0.010557184750733138
3778, this	0.00997067448680352
2962, supreme	0.009824046920821115

Strong's No., English	Frequency
5547, Christ	0.009237536656891495
1487, if	0.009237536656891495
3739, who	0.008944281524926686
3754, that	0.008797653958944282

2Cor

Strong's No., English	Frequency
3588, this	0.12329380174535691
2532, and	0.04408144998881181
1772, in	0.036025956589841125
1473, I	0.036025956589841125
4771, you	0.03490713806220631
3756, not	0.021705079436115462
2316, God	0.01767733273663012
1519, into	0.017453569031103153
1063, for	0.016782277914522265
1161, but	0.0165585142089953
235, but	0.01521593197583352
846, himself	0.013873349742671738
1510, to be	0.013873349742671738
3956, all	0.011635712687402103
3754, that	0.01141194898187514
3361, not lest	0.010964421570821213
5547, Christ	0.010516894159767285
1223, through	0.010069366748713358
2443, in order that	0.009845603043186395
3778, this	0.009621839337659432

Gal

Strong's No., English	Frequency
3588, this	0.12387791741472172
2532, and	0.03231597845601436
1473, I	0.026481149012567325
1161, but	0.026032315978456014
1510, to be	0.02378815080789946
4771, you	0.02333931777378815
1772, in	0.018402154398563735
5547, Christ	0.017055655296229804
3756, not	0.017055655296229804

1063, for	0.01615798922800718
1537, out from	0.01570915619389587
3551, law	0.01436265709156194
2316, God	0.013464991023339317
1519, into	0.013464991023339317
3754, that	0.013016157989228007
846, himself	0.011669658886894075
3361, not lest	0.010771992818671455
3739, who	0.010323159784560144
235, but	0.010323159784560144
4102, faith	0.009874326750448833

Eph

Strong's No., English	Frequency
3588, this	0.17858619264158743
2532, and	0.05663497312939231
1772, in	0.05002066969822241
846, himself	0.02728400165357586
3956, all	0.02108309218685407
4771, you	0.020256304257957835
1510, to be	0.019842910293509715
5547, Christ	0.019016122364613478
1473, I	0.017775940471269118
1519, into	0.015295576684580404
3739, who	0.01364200082678793
2316, God	0.012815212897891691
2962, supreme	0.010748243075651096
2596, according	0.009921455146754857
2443, in order that	0.009508061182306739
1223, through	0.0086812732534105
1161, but	0.00826787928896238
2424, Jesus	0.00826787928896238
3778, this	0.007441091360066143
3361, not lest	0.006614303431169905

Phil

Strong's No., English	Frequency
3588, this	0.11779141104294479
2532, and	0.0656441717791411

1772, in	0.03987730061349693
1473, I	0.03558282208588957
4771, you	0.03190184049079755
5547, Christ	0.022699386503067485
3956, all	0.020245398773006136
846, himself	0.01901840490797546
1161, but	0.016564417177914112
2316, God	0.014723926380368098
1519, into	0.01411042944785276
2424, Jesus	0.013496932515337423
3754, that	0.012883435582822086
1510, to be	0.010429447852760737
3739, who	0.009202453987730062
2962, supreme	0.009202453987730062
3778, this	0.009202453987730062
235, but	0.009202453987730062
1223, through	0.008588957055214725
1487, if	0.007975460122699387

Col

Strong's No., English	Frequency
3588, this	0.16444444444444445
2532, and	0.06412698412698413
1772, in	0.05523809523809524
4771, you	0.035555555555555556
846, himself	0.02857142857142857
3956, all	0.024761904761904763
3739, who	0.02412698412698413
1510, to be	0.01650793650793651
5547, Christ	0.015873015873015872
1473, I	0.015873015873015872
2316, God	0.013333333333333334
1519, into	0.012063492063492064
2962, supreme	0.010158730158730159
1223, through	0.008888888888888889
2596, according	0.008888888888888889
2443, in order that	0.008253968253968255
1537, out from	0.006984126984126984
3361, not lest	0.006349206349206349
575, from	0.005714285714285714

4561, flesh	0.005714285714285714

1Thess

Strong's No., English	Frequency
3588, this	0.1296673455532926
2532, and	0.06788866259334692
4771, you	0.05702647657841141
1772, in	0.03598099117447386
1473, I	0.03394433129667346
2316, God	0.024439918533604887
1519, into	0.017651052274270197
2962, supreme	0.016293279022403257
846, himself	0.016293279022403257
1063, for	0.015614392396469789
80, brother	0.012898845892735914
3756, not	0.012219959266802444
3956, all	0.012219959266802444
2424, Jesus	0.010862186014935505
1161, but	0.010183299389002037
3361, not lest	0.010183299389002037
2531, as	0.008825526137135099
3754, that	0.008825526137135099
235, but	0.008825526137135099
1492, know	0.008825526137135099

2Thess

Strong's No., English	Frequency
3588, this	0.1362530413625304
2532, and	0.06082725060827251
4771, you	0.04866180048661801
1772, in	0.031630170316301706
1473, I	0.029197080291970802
2962, supreme	0.0267639902676399
846, himself	0.021897810218978103
2316, God	0.021897810218978103
3956, all	0.019464720194647202
1519, into	0.0170316301703163
2424, Jesus	0.015815085158150853
3739, who	0.014598540145985401

3361, not lest	0.01338199513381995
3754, that	0.01338199513381995
1161, but	0.01338199513381995
5547, Christ	0.012165450121654502
1223, through	0.012165450121654502
575, from	0.010948905109489052
80, brother	0.010948905109489052
3756, not	0.009732360097323601

1Tim

Strong's No., English	Frequency
3588, this	0.09949622166246852
2532, and	0.05793450881612091
1772, in	0.02707808564231738
1161, but	0.018891687657430732
1510, to be	0.018261964735516372
3361, not lest	0.015113350125944584
3956, all	0.014483627204030227
3739, who	0.013853904282115869
2316, God	0.013853904282115869
4102, faith	0.011964735516372796
1519, into	0.011964735516372796
3778, this	0.011335012594458438
5100, some	0.011335012594458438
2570, good	0.010075566750629723
2443, in order that	0.009445843828715366
5547, Christ	0.009445843828715366
1473, I	0.009445843828715366
4771, you	0.008816120906801008
2192, to have	0.008816120906801008
2424, Jesus	0.008816120906801008

2Tim

Strong's No., English	Frequency
3588, this	0.12045270816491511
2532, and	0.0549717057396928
1473, I	0.03395311236863379
1772, in	0.029911075181891674
1161, but	0.019401778496362168

4771, you	0.016976556184316895
3739, who	0.016168148746968473
1519, into	0.014551333872271624
3956, all	0.014551333872271624
2962, supreme	0.012934518997574777
846, himself	0.012934518997574777
1510, to be	0.012934518997574777
1063, for	0.01131770412287793
5547, Christ	0.010509296685529508
2424, Jesus	0.010509296685529508
2316, God	0.010509296685529508
1223, through	0.009700889248181084
235, but	0.009700889248181084
3756, not	0.009700889248181084
3778, this	0.00889248181083266

Titus

Strong's No., English	Frequency
3588, this	0.09270516717325228
2532, and	0.0547112462006079
1473, I	0.028875379939209727
1510, to be	0.0243161094224924
3956, all	0.02127659574468085
3361, not lest	0.019756838905775075
2443, in order that	0.019756838905775075
1772, in	0.019756838905775075
2316, God	0.019756838905775075
3739, who	0.013677811550151976
4771, you	0.0121580547112462
2041, work	0.0121580547112462
1161, but	0.0121580547112462
846, himself	0.0121580547112462
2596, according	0.0121580547112462
4102, faith	0.00911854103343465
4990, savior	0.00911854103343465
1063, for	0.00911854103343465
3778, this	0.007598784194528876
3056, word	0.007598784194528876

Phlm

Strong's No., English	Frequency
3588, this	0.10746268656716418
4771, you	0.07164179104477612
1473, I	0.0626865671641791
2532, and	0.05373134328358209
1772, in	0.029850746268656716
5547, Christ	0.023880597014925373
1161, but	0.01791044776119403
2424, Jesus	0.01791044776119403
3739, who	0.014925373134328358
2962, supreme	0.014925373134328358
1223, through	0.011940298507462687
3754, that	0.011940298507462687
2443, in order that	0.011940298507462687
80, brother	0.011940298507462687
2192, to have	0.011940298507462687
5613, as	0.011940298507462687
3778, this	0.008955223880597015
4160, make	0.008955223880597015
4968, bowels	0.008955223880597015
3972, Paul	0.008955223880597015

Heb

Strong's No., English	Frequency
3588, this	0.1408678102926337
2532, and	0.05126135216952573
846, himself	0.028456104944500505
1063, for	0.018365287588294654
1519, into	0.014934409687184663
3739, who	0.014934409687184663
1161, but	0.013925327951564077
2316, God	0.013723511604439959
3756, not	0.013319878910191726
1473, I	0.013319878910191726
1772, in	0.013118062563067608
4771, you	0.011907164480322906
1223, through	0.011503531786074672
1510, to be	0.011099899091826439
3956, all	0.010696266397578204
3778, this	0.00887991927346115

3361, not lest	0.00887991927346115
2596, according	0.008274470232088799
2192, to have	0.007870837537840566
3004, say	0.007669021190716448

Jas

Strong's No., English	Frequency
3588, this	0.13225991949396204
2532, and	0.06325474410580793
4771, you	0.02702702702702703
846, himself	0.023001725129384705
1772, in	0.02185163887291547
1161, but	0.02127659574468085
1510, to be	0.018401380103507763
3756, not	0.015526164462334674
3361, not lest	0.013801035077630822
1473, I	0.012650948821161587
80, brother	0.010925819436457734
2316, God	0.009200690051753882
4102, faith	0.009200690051753882
3754, that	0.009200690051753882
1519, into	0.008625646923519264
2041, work	0.008625646923519264
5100, some	0.008625646923519264
1063, for	0.008625646923519264
2962, supreme	0.008050603795284647
1537, out from	0.007475560667050029

1Pet

Strong's No., English	Frequency
3588, this	0.11813842482100238
2532, and	0.042362768496420046
4771, you	0.031026252983293555
1772, in	0.029832935560859187
1519, into	0.025059665871121718
2316, God	0.023269689737470168
846, himself	0.02028639618138425
3739, who	0.018496420047732696
1161, but	0.016706443914081145

5613, as	0.016109785202863963
5547, Christ	0.013126491646778043
3956, all	0.010739856801909307
1223, through	0.010739856801909307
235, but	0.00954653937947494
1487, if	0.008949880668257757
3754, that	0.008949880668257757
3361, not lest	0.008353221957040573
2443, in order that	0.007756563245823389
3756, not	0.007756563245823389
3778, this	0.0065632458233890216

2Pet

Strong's No., English	Frequency
3588, this	0.11293260473588343
2532, and	0.05737704918032787
1772, in	0.039162112932604735
846, himself	0.026411657559198543
3778, this	0.020036429872495445
4771, you	0.01912568306010929
1161, but	0.01912568306010929
3739, who	0.017304189435336976
1473, I	0.017304189435336976
1063, for	0.01366120218579235
2962, supreme	0.012750455373406194
1510, to be	0.011839708561020037
3756, not	0.01092896174863388
2250, day	0.01092896174863388
1519, into	0.010018214936247723
5613, as	0.009107468123861567
2424, Jesus	0.00819672131147541
5547, Christ	0.007285974499089253
3956, all	0.006375227686703097
1223, through	0.006375227686703097

1John

Strong's No., English	Frequency
3588, this	0.16807116104868913
2532, and	0.060393258426966294

846, himself	0.047752808988764044
1510, to be	0.046348314606741575
1772, in	0.03698501872659176
3754, that	0.035580524344569285
2316, God	0.02902621722846442
1473, I	0.026217228464419477
3756, not	0.02247191011235955
3778, this	0.018258426966292134
4771, you	0.015917602996254682
1537, out from	0.015917602996254682
3739, who	0.014044943820224719
2192, to have	0.013108614232209739
25, to love	0.013108614232209739
3956, all	0.012640449438202247
1097, to know	0.011704119850187267
3306, to stay	0.011235955056179775
2889, world	0.010767790262172285
1437, if	0.010299625468164793

2John

Strong's No., English	Frequency
3588, this	0.1346938775510204
2532, and	0.0653061224489796
4771, you	0.0326530612244898
1772, in	0.0326530612244898
846, himself	0.02857142857142857
1473, I	0.024489795918367346
1510, to be	0.024489795918367346
3361, not lest	0.02040816326530612
3756, not	0.02040816326530612
225, truth	0.02040816326530612
3778, this	0.02040816326530612
2443, in order that	0.02040816326530612
1785, commandment	0.0163265306122449
3962, father	0.0163265306122449
2192, to have	0.0163265306122449
235, but	0.0163265306122449
5547, Christ	0.012244897959183673
3306, to stay	0.012244897959183673
4314, forward	0.012244897959183673

1519, into	0.012244897959183673

3John

Strong's No., English	Frequency
3588, this	0.1324200913242009
2532, and	0.0502283105022831
4771, you	0.045662100456621
1473, I	0.0273972602739726
225, truth	0.0273972602739726
3739, who	0.0228310502283105
3140, to witness	0.0182648401826484
3778, this	0.0182648401826484
3756, not	0.0182648401826484
846, himself	0.0182648401826484
27, beloved	0.0182648401826484
1125, write	0.0136986301369863
80, brother	0.0136986301369863
1577, church	0.0136986301369863
235, but	0.0136986301369863
2316, God	0.0136986301369863
1772, in	0.0136986301369863
4160, make	0.0136986301369863
3686, name	0.0091324200913242
3708, see	0.0091324200913242

Jude

Strong's No., English	Frequency
3588, this	0.12035010940919037
2532, and	0.045951859956236324
1161, but	0.0262582056892779
4771, you	0.024070021881838075
3778, this	0.019693654266958426
1772, in	0.0175054704595186
3956, all	0.0175054704595186
846, himself	0.0175054704595186
1473, I	0.015317286652078774
2962, supreme	0.015317286652078774
2424, Jesus	0.01312910284463895
1438, himself	0.01312910284463895

5547, Christ	0.01312910284463895
1519, into	0.01312910284463895
4012, about	0.010940919037199124
5083, guard	0.010940919037199124
3739, who	0.010940919037199124
40, sacred	0.0087527352297593
2316, God	0.0087527352297593
1510, to be	0.0087527352297593

Rev

Strong's No., English	Frequency
3588, this	0.19172091131000812
2532, and	0.1136086248982913
846, himself	0.04465012205044752
1772, in	0.016069975589910496
1909, upon	0.014646053702196907
1537, out from	0.013730675345809601
1510, to be	0.011187957689178194
2192, to have	0.01017087062652563
3004, say	0.01017087062652563
1473, I	0.00986574450772986
2316, God	0.009662327095199348
3708, see	0.009052074857607811
1093, land	0.008340113913751017
3173, great	0.008136696501220505
1519, into	0.008136696501220505
4771, you	0.008136696501220505
3739, who	0.007119609438567941
5613, as	0.007017900732302685
32, angel	0.006814483319772173
3756, not	0.006814483319772173

B Book Similarities between all 27 books

Matt

Book	a-value
Luke	1.2498858510252246
Mark	2.808491884758681
Acts	2.863939222914629
Jas	4.4346324976013465
John	4.678567860690274
Jude	6.961803874406476
Heb	7.88402478953559
1Cor	8.27718596926664
Rom	8.344439915059338
2Pet	9.082797677205104
Gal	9.693678966679508
2Tim	9.785473187567574
Eph	10.432615562269314
Rev	10.503923386527115
2Cor	10.569479552521868
1Pet	11.01860472708464
2John	11.411612200679972
2Thess	12.009534851244338
Col	12.526479987093426
Phil	12.982965547698754
1Thess	13.166040445500222
1John	13.319468398025837
Phlm	14.564184398991955
1Tim	15.001133513719347
Titus	15.37829088943846
3John	19.944348788153366

Mark

Book	a-value
Matt	2.808491884758681
Luke	3.0822798286148467
Acts	4.037392272772234
John	5.65640388695672
Jas	6.182932982948504
Heb	6.841319253495689
Rev	9.052949729869212

Rom	9.224729133939874
Jude	10.453955571142572
2Pet	10.584373854975077
1Cor	10.81424890587908
1Pet	11.240397808725703
2Cor	11.416286750616639
Eph	11.478587930258778
2Tim	11.516673884347284
2John	11.563196454726096
Gal	12.224830001939202
1John	13.325773953034545
2Thess	13.55507990735967
3John	13.581649263068126
Titus	14.109114845386244
Phil	14.217385143550265
Col	14.342928234537267
1Thess	14.585334331558329
1Tim	14.731344772719343
Phlm	17.67813790680798

Luke

Book	a-value
Matt	1.2498858510252246
Acts	2.9954302131011126
Mark	3.0822798286148467
Jas	5.291714777343677
John	5.870433429065313
Jude	6.473497387400187
1Cor	7.820456175164511
Heb	8.067114764522275
Rom	8.516943816226547
2Pet	8.62071208429699
2Tim	9.519565236668662
1Pet	10.589280497236954
2Cor	10.793233528392742
2John	10.996672989658158
Gal	11.148630974497753
Rev	11.15379430845007
1John	11.17264101763682
Eph	11.197209346847472

2Thess	12.04492799547027
Phil	12.728676716661727
Titus	12.83754770994
1Thess	13.204306172337871
1Tim	13.302681404703307
Col	13.670925587289256
Phlm	14.904853202539176
3John	18.053114680370125

John

Book	a-value
Matt	4.678567860690274
Mark	5.65640388695672
Luke	5.870433429065313
Acts	8.078242143693412
1John	8.411323797616504
2John	8.609675216567437
Jas	8.721868817722687
3John	10.265949844393411
Rom	10.309927511788267
Gal	10.404714409545871
1Cor	10.907198570530603
2Cor	11.061894508320831
Heb	11.210480742875674
Jude	11.469453846497478
2Pet	12.024894659343758
2Tim	12.148982543463408
2Thess	12.46014091223983
Phlm	12.762588361337974
Rev	12.775960411976842
1Pet	13.477308495380738
Phil	13.49163960435913
Eph	13.492052186245019
1Thess	13.71369308283147
1Tim	14.946319353391026
Col	15.799653009352925
Titus	16.259662280026276

Acts

Book	*a*-value
Matt	2.863939222914629
Luke	2.9954302131011126
Mark	4.037392272772234
Jude	4.916097429993837
2Pet	5.082873396765076
Heb	5.416026847796947
2Tim	5.528398577945508
Jas	6.472835406060057
Rom	6.505012963443187
1Cor	7.687001356338429
Eph	7.847942989942897
John	8.078242143693412
2Cor	8.397518606367868
2Thess	8.711506654093402
Phil	9.061621886988657
1Pet	9.629174665717688
Gal	9.700489665529872
1Thess	9.789277095129279
Col	10.418140064290256
1Tim	10.567553742690878
1John	10.738945261034436
Rev	11.117753366041498
2John	11.486116236413995
Titus	12.743731064008266
Phlm	13.471742486015694
3John	16.380265256484805

Rom

Book	*a*-value
2Cor	3.304624866250097
1Cor	3.3634385915073337
Heb	3.589732464510037
Gal	3.762412087232254
Jas	3.9067570184953815
2Pet	5.068264652963615
2Tim	5.347651691181487
1Thess	5.562672736939286
Acts	6.505012963443187
Phil	6.559658832524622

2Thess	6.597659163278408
1Pet	7.751629182606049
Eph	7.797311634922189
Matt	8.344439915059338
1Tim	8.444144360376796
Luke	8.516943816226547
2John	8.643309820482035
Jude	8.81927986491159
Mark	9.224729133939874
Titus	9.601923689084177
Phlm	9.771174850275099
Col	9.98144735009302
1John	10.30088385814545
John	10.309927511788267
3John	13.368059306084888
Rev	15.409420118108146

1Cor

Book	a-value
Gal	2.947045962736011
Rom	3.3634385915073337
Jas	4.260601660084346
2Cor	4.470531192075281
2Tim	5.599865151542004
2Pet	6.236601819943239
Phil	6.8095621465672505
Heb	7.061425841167152
Jude	7.0657256993816775
Acts	7.687001356338429
Luke	7.820456175164511
1Thess	7.895460709124209
Titus	8.199243241982249
Matt	8.27718596926664
1Tim	8.283817030355692
Eph	8.619380545971659
2Thess	8.761443809163133
2John	9.057562497589577
Phlm	9.486700703180663
Col	10.43528571770649
1Pet	10.801428149132054

Mark	10.81424890587908
John	10.907198570530603
1John	10.964978350169366
3John	16.707154203096355
Rev	17.659429830249145

2Cor

Book	a-value
Rom	3.304624866250097
1Thess	3.5831709969695367
Gal	3.9831036963231763
1Cor	4.470531192075281
Phil	4.507446899462842
2Tim	4.72576351900792
Heb	5.707156188036583
2Thess	5.750838561889809
Jas	5.860917943917871
2Pet	6.368628606851557
1Pet	7.306147537348039
2John	7.9806914148624415
Phlm	8.268911865836756
Acts	8.397518606367868
Eph	8.437749735109865
Jude	9.317805067829227
1Tim	9.53777224016234
Col	10.172539684962665
Matt	10.569479552521868
Titus	10.640134011635926
Luke	10.793233528392742
John	11.061894508320831
Mark	11.416286750616639
1John	11.669816665478802
3John	13.617121499640714
Rev	18.032964178582887

Gal

Book	a-value
1Cor	2.947045962736011
Rom	3.762412087232254

2Cor	3.9831036963231763
Jas	6.597426691669961
2Tim	6.785928983326744
Phil	6.927831934108866
2Pet	7.163029618166046
Heb	7.386919153066945
1Tim	7.770543696461224
2John	7.943318349523352
Phlm	8.329605437730203
Eph	8.612294442945819
1Thess	8.899858487169308
Matt	9.693678966679508
Acts	9.700489665529872
Titus	9.804481544422595
Jude	10.03478412469002
2Thess	10.379014777612305
John	10.404714409545871
1Pet	10.552026086581963
Luke	11.148630974497753
Col	11.244820523085805
Mark	12.224830001939202
1John	13.041947903321775
3John	13.333865818090239
Rev	19.824662106823926

Eph

Book	a-value
Col	2.3609806708099352
Phil	4.074692422709013
2Tim	5.1576756379507955
2Thess	5.32091262046331
Jude	6.429519356344312
2Pet	7.094787445303593
Rom	7.797311634922189
Acts	7.847942989942897
1Thess	7.932384582943782
2John	8.25250664675242
2Cor	8.437749735109865
Gal	8.612294442945819
1Cor	8.619380545971659

Titus	8.652917841104854
1Pet	8.685381698472085
Heb	9.032483682085626
1Tim	9.098224706866691
Jas	9.517516729155506
Phlm	9.904483746355698
Matt	10.432615562269314
1John	11.093272998064585
Luke	11.197209346847472
3John	11.387942654416358
Mark	11.478587930258778
John	13.492052186245019
Rev	14.632467106599286

Phil

Book	a-value
2Thess	3.743576074890499
Eph	4.074692422709013
2Tim	4.080985743135747
2Cor	4.507446899462842
1Thess	5.187291306780544
Jude	5.3529973370136865
Col	5.739007455473995
Phlm	6.212910605262698
Rom	6.559658832524622
1Cor	6.8095621465672505
Gal	6.927831934108866
2Pet	6.976244901588798
1Pet	7.362437600257963
Titus	8.625706983866053
Heb	8.996867241213454
Acts	9.061621886988657
Jas	9.559517594566442
3John	9.879233855901957
1Tim	10.056139192073712
2John	10.34884582024838
1John	10.959145595487586
Luke	12.728676716661727
Matt	12.982965547698754
John	13.49163960435913

Mark	14.217385143550265
Rev	18.02813071464653

Col

Book	a-value
Eph	2.3609806708099352
2Thess	5.468475270981354
Phil	5.739007455473995
2Tim	7.198377447288609
Jude	8.550672535950751
Titus	8.753522659821979
2Pet	8.861531020872318
2John	9.03950380212461
1Pet	9.779128098094167
1Thess	9.83325767733729
Rom	9.98144735009302
2Cor	10.172539684962665
Acts	10.418140064290256
1Cor	10.43528571770649
Phlm	10.883532588172633
Heb	10.943680002361875
1Tim	10.980561680292086
Gal	11.244820523085805
Jas	11.474849252405225
1John	11.770717308863905
Matt	12.526479987093426
Luke	13.670925587289256
3John	13.73372401856001
Mark	14.342928234537267
Rev	15.213140460105995
John	15.799653009352925

1Thess

Book	a-value
2Thess	2.9432926622927016
2Cor	3.5831709969695367
Phil	5.187291306780544
Rom	5.562672736939286
2Tim	6.1053473627754

2John	6.732221577764145
Jas	7.754277794253226
3John	7.826868985385091
1Cor	7.895460709124209
Eph	7.932384582943782
2Pet	8.493948313235821
Jude	8.636760589021716
Gal	8.899858487169308
Heb	9.06111904251555
Phlm	9.212284950091842
Acts	9.789277095129279
1Pet	9.801746744189156
Col	9.83325767733729
Titus	11.013048909349884
1John	11.757880225841074
1Tim	12.788256579171469
Matt	13.166040445500222
Luke	13.204306172337871
John	13.71369308283147
Mark	14.585334331558329
Rev	17.2606162680904

2Thess

Book	a-value
1Thess	2.9432926622927016
Phil	3.743576074890499
2Tim	4.3939890439105405
Eph	5.32091262046331
Col	5.468475270981354
2Cor	5.750838561889809
Jude	6.482519630870105
Rom	6.597659163278408
2John	7.080979993815129
Titus	7.208837011402443
Jas	7.430148101067357
1Pet	7.609416839352782
Phlm	7.690568180297394
3John	8.322927419278681
Heb	8.33268287101203
2Pet	8.477922793882962

Acts	8.711506654093402
1Cor	8.761443809163133
1John	9.31977990060063
1Tim	10.134941214578092
Gal	10.379014777612305
Matt	12.009534851244338
Luke	12.04492799547027
John	12.46014091223983
Mark	13.55507990735967
Rev	17.769507232903926

1Tim

Book	a-value
Titus	4.317441163201198
2Tim	6.445178946323636
Jas	7.680119535701941
Gal	7.770543696461224
Heb	7.78076490539881
1Cor	8.283817030355692
Rom	8.444144360376796
Eph	9.098224706866691
2Cor	9.53777224016234
Jude	9.620437615505121
2John	9.68241725096013
Phil	10.056139192073712
2Pet	10.075827813779586
2Thess	10.134941214578092
1Pet	10.350272295366853
Acts	10.567553742690878
Col	10.980561680292086
Phlm	11.440024142005205
1Thess	12.788256579171469
Luke	13.302681404703307
3John	14.593689402632535
Mark	14.731344772719343
1John	14.774390912321072
John	14.946319353391026
Matt	15.001133513719347
Rev	19.502434620726405

2Tim

Book	a-value
Phil	4.080985743135747
Jude	4.108539867238225
2Pet	4.179558098638855
2Thess	4.3939890439105405
2Cor	4.72576351900792
Eph	5.1576756379507955
Rom	5.347651691181487
Acts	5.528398577945508
1Cor	5.599865151542004
Phlm	5.63084202844056
1Thess	6.1053473627754
Heb	6.330060367286448
1Tim	6.445178946323636
Gal	6.785928983326744
Col	7.198377447288609
2John	7.611731874976947
Jas	7.852435772385855
1Pet	7.891086685830416
Titus	8.1936759996299
3John	8.322314089730803
Luke	9.519565236668662
Matt	9.785473187567574
1John	10.901522953046426
Mark	11.516673884347284
John	12.148982543463408
Rev	17.05793508686522

Titus

Book	a-value
1Tim	4.317441163201198
2Thess	7.208837011402443
Jude	7.31632222961266
2Tim	8.1936759996299
1Cor	8.199243241982249
Phil	8.625706983866053
Eph	8.652917841104854
Col	8.753522659821979
2John	9.586225223159214

Rom	9.601923689084177
Gal	9.804481544422595
2Cor	10.640134011635926
1Thess	11.013048909349884
Phlm	11.217562020686868
Heb	11.42775968242117
2Pet	11.457095749268587
Jas	11.647886495817415
3John	12.400194926351592
1John	12.63064240324077
Acts	12.743731064008266
1Pet	12.796446553015791
Luke	12.83754770994
Mark	14.109114845386244
Rev	15.137645500505732
Matt	15.37829088943846
John	16.259662280026276

Phlm

Book	a-value
3John	4.857375518366468
2Tim	5.63084202844056
Phil	6.212910605262698
2John	7.283312552750726
Jude	7.465168215982802
2Pet	7.643641327173146
2Thess	7.690568180297394
2Cor	8.268911865836756
Gal	8.329605437730203
1Thess	9.212284950091842
1Cor	9.486700703180663
Rom	9.771174850275099
Eph	9.904483746355698
Col	10.883532588172633
1Pet	11.171906885552115
Titus	11.217562020686868
1Tim	11.440024142005205
John	12.762588361337974
Acts	13.471742486015694
Jas	13.638259975930449

Heb	14.025403014218568
1John	14.272362925822446
Matt	14.564184398991955
Luke	14.904853202539176
Mark	17.67813790680798
Rev	21.254361996129752

Heb

Book	a-value
Rom	3.589732464510037
2Pet	5.175658371080429
Acts	5.416026847796947
2Cor	5.707156188036583
2Tim	6.330060367286448
Jas	6.406746235804077
Mark	6.841319253495689
1Cor	7.061425841167152
Gal	7.386919153066945
1Tim	7.78076490539881
Matt	7.88402478953559
Luke	8.067114764522275
1Pet	8.193683070161079
2Thess	8.33268287101203
Phil	8.996867241213454
Eph	9.032483682085626
1Thess	9.06111904251555
Jude	9.495033907696877
Col	10.943680002361875
John	11.210480742875674
Titus	11.42775968242117
2John	11.907777582390002
Rev	12.46321011225798
1John	13.263495422475312
3John	13.696341786443817
Phlm	14.025403014218568

Jas

Book	a-value
Rom	3.9067570184953815

1Cor	4.260601660084346
Matt	4.4346324976013465
2John	4.939340998530797
Luke	5.291714777343677
2Cor	5.860917943917871
Mark	6.182932982948504
Heb	6.406746235804077
Acts	6.472835406060057
Gal	6.597426691669961
2Pet	6.850911851453528
Jude	7.206992489404924
2Thess	7.430148101067357
1Tim	7.680119535701941
1Thess	7.754277794253226
2Tim	7.852435772385855
John	8.721868817722687
Eph	9.517516729155506
Phil	9.559517594566442
1John	9.876051832732749
1Pet	10.062653568373358
3John	10.227760232407563
Col	11.474849252405225
Titus	11.647886495817415
Rev	11.797770304687862
Phlm	13.638259975930449

1Pet

Book	a-value
2Cor	7.306147537348039
Phil	7.362437600257963
2Thess	7.609416839352782
Rom	7.751629182606049
2Pet	7.872029236949034
2Tim	7.891086685830416
Heb	8.193683070161079
Eph	8.685381698472085
Jude	9.145894534385286
Acts	9.629174665717688
Col	9.779128098094167
1Thess	9.801746744189156

Jas	10.062653568373358
1Tim	10.350272295366853
Gal	10.552026086581963
Luke	10.589280497236954
1Cor	10.801428149132054
Matt	11.01860472708464
Phlm	11.171906885552115
Mark	11.240397808725703
3John	12.342973677911544
Titus	12.796446553015791
2John	13.229612606404983
John	13.477308495380738
Rev	15.995224823702431
1John	16.255018289712257

2Pet

Book	a-value
2Tim	4.179558098638855
Rom	5.068264652963615
Acts	5.082873396765076
Heb	5.175658371080429
Jude	5.671691541437243
1Cor	6.236601819943239
2Cor	6.368628606851557
Jas	6.850911851453528
Phil	6.976244901588798
Eph	7.094787445303593
Gal	7.163029618166046
Phlm	7.643641327173146
3John	7.7772514624107
1Pet	7.872029236949034
2John	8.10578725965267
2Thess	8.477922793882962
1Thess	8.493948313235821
Luke	8.62071208429699
Col	8.861531020872318
Matt	9.082797677205104
1Tim	10.075827813779586
Mark	10.584373854975077
1John	11.338415395563855

Titus	11.457095749268587
John	12.024894659343758
Rev	17.79850209361776

1John

Book	a-value
2John	7.667529473755046
John	8.411323797616504
2Thess	9.31977990060063
Jas	9.876051832732749
Rom	10.30088385814545
Acts	10.738945261034436
2Tim	10.901522953046426
Phil	10.959145595487586
1Cor	10.964978350169366
Eph	11.093272998064585
Luke	11.17264101763682
2Pet	11.338415395563855
Jude	11.627188554800423
2Cor	11.669816665478802
1Thess	11.757880225841074
Col	11.770717308863905
3John	12.341030897943835
Titus	12.63064240324077
Gal	13.041947903321775
Heb	13.263495422475312
Matt	13.319468398025837
Mark	13.325773953034545
Rev	14.196820733475423
Phlm	14.272362925822446
1Tim	14.774390912321072
1Pet	16.255018289712257

2John

Book	a-value
Jas	4.939340998530797
3John	4.973085154000499
1Thess	6.732221577764145
Jude	6.853020640291435

2Thess	7.080979993815129
Phlm	7.283312552750726
2Tim	7.611731874976947
1John	7.667529473755046
Gal	7.943318349523352
2Cor	7.9806914148624415
2Pet	8.10578725965267
Eph	8.25250664675242
John	8.609675216567437
Rom	8.643309820482035
Col	9.03950380212461
1Cor	9.057562497589577
Rev	9.517225302237614
Titus	9.586225223159214
1Tim	9.68241725096013
Phil	10.34884582024838
Luke	10.996672989658158
Matt	11.411612200679972
Acts	11.486116236413995
Mark	11.563196454726096
Heb	11.907777582390002
1Pet	13.229612606404983

3John

Book	a-value
Phlm	4.857375518366468
2John	4.973085154000499
Jude	5.64052290300648
2Pet	7.7772514624107
1Thess	7.826868985385091
2Tim	8.322314089730803
2Thess	8.322927419278681
Phil	9.879233855901957
Jas	10.227760232407563
John	10.265949844393411
Eph	11.387942654416358
1John	12.341030897943835
1Pet	12.342973677911544
Titus	12.400194926351592
Gal	13.333865818090239

Rom	13.368059306084888
Mark	13.581649263068126
2Cor	13.617121499640714
Heb	13.696341786443817
Col	13.73372401856001
Rev	14.124352261904734
1Tim	14.593689402632535
Acts	16.380265256484805
1Cor	16.707154203096355
Luke	18.053114680370125
Matt	19.944348788153366

Jude

Book	a-value
2Tim	4.108539867238225
Acts	4.916097429993837
Phil	5.3529973370136865
3John	5.64052290300648
2Pet	5.671691541437243
Eph	6.429519356344312
Luke	6.473497387400187
2Thess	6.482519630870105
2John	6.853020640291435
Matt	6.961803874406476
1Cor	7.0657256993816775
Jas	7.206992489404924
Titus	7.31632222961266
Phlm	7.465168215982802
Col	8.550672535950751
1Thess	8.636760589021716
Rom	8.81927986491159
1Pet	9.145894534385286
2Cor	9.317805067829227
Heb	9.495033907696877
1Tim	9.620437615505121
Gal	10.03478412469002
Mark	10.453955571142572
John	11.469453846497478
1John	11.627188554800423
Rev	16.437010253461374

Rev

Book	a-value
Mark	9.052949729869212
2John	9.517225302237614
Matt	10.503923386527115
Acts	11.117753366041498
Luke	11.15379430845007
Jas	11.797770304687862
Heb	12.46321011225798
John	12.775960411976842
3John	14.124352261904734
1John	14.196820733475423
Eph	14.632467106599286
Titus	15.137645500505732
Col	15.213140460105995
Rom	15.409420118108146
1Pet	15.995224823702431
Jude	16.437010253461374
2Tim	17.05793508686522
1Thess	17.2606162680904
1Cor	17.659429830249145
2Thess	17.769507232903926
2Pet	17.79850209361776
Phil	18.02813071464653
2Cor	18.032964178582887
1Tim	19.502434620726405
Gal	19.824662106823926
Phlm	21.254361996129752

C   Code

The reader.py file reads the csv file and creates an array for each book (contained in a dictionary) that contains all the words (by Strong's number) in sequential order.

```python
#!/usr/bin/env python3
# -*- coding: utf-8 -*-
"""
Created on Sat Apr 25 15:06:02 2020

@author: jacob
"""

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

fileName = 'nt.csv'

# Read .csv file into pandas dataframe
df = pd.read_csv(fileName, sep='\t', header=0, usecols=[0,4], \
                 index_col=False, dtype='str')

# clean up strong's numbers: 1080&5656 -> 1080
def clean(string):
    if '&' in string:
        index = string.find('&')
        new = string[0:index]
        return new
    else:
        return string
df = df.applymap(clean)

# create one array per book and create a dictionary called
# bookDict that takes the key 'title' and returns an array
# containing all the words in that book.
def bcvToBook(string):
    """ Convert 'Matt 1:1' to 'Matt' """
    if ' ' in string:
        spaceIndex = string.find(' ')
        book = string[0:spaceIndex]
        return book
    else:
        return string
df = df.applymap(bcvToBook)
array = df.to_numpy()
# find the index of the row where each book ends:
lastWordIndices = list()
for rowNum in range(len(array)):
    if rowNum < len(array)-1:
        currentBook = array[rowNum, 0]
```

```
47          nextBook = array[rowNum + 1, 0]
48          if currentBook != nextBook:
49              lastWordIndices.append(rowNum)
50      else:
51          lastWordIndices.append(rowNum)
52  # split up the books
53  bookDict = dict()
54  for i in range(len(lastWordIndices)):
55      if i == 0:
56          lastWord = lastWordIndices[i]
57          book = array[0:lastWord + 1, 1]
58          bookTitle = array[lastWord, 0]
59          bookDict[bookTitle] = book
60      else:
61          lastWord = lastWordIndices[i]
62          prevlastWord = lastWordIndices[i-1]
63          book = array[prevlastWord + 1:lastWord + 1, 1]
64          bookTitle = array[lastWord, 0]
65          bookDict[bookTitle] = book
```

The nt.py file uses the result of the reader.py file. There are a couple useful functions (descriptions in the code): freqCount, freqCountTwenty (used to create Appendix A), freqDict, compare1, mostSimilarTo, printToTextFile (used to create Appendix B), printToDict, and mostSimilarIn. The final output is the result of the table in the section 4.

```python
#!/usr/bin/env python3
# -*- coding: utf-8 -*-
"""
Created on Sat Apr 25 16:42:36 2020

@author: jacob
"""

import reader
import numpy as np
import itertools

bookDict = reader.bookDict

bookList = ['Matt', 'Mark', 'Luke', 'John', 'Acts', 'Rom', '1Cor',
            '2Cor', 'Gal', 'Eph', 'Phil', 'Col', '1Thess', '2Thess',
            '1Tim', '2Tim', 'Titus', 'Phlm', 'Heb', 'Jas', '1Pet',
            '2Pet', '1John', '2John', '3John', 'Jude', 'Rev']

gospels = ['Matt', 'Mark', 'Luke', 'John']
pauline = ['Rom', '1Cor', '2Cor', 'Gal', 'Eph', 'Phil', 'Col',
           '1Thess', '2Thess', '1Tim', '2Tim', 'Titus', 'Phlm']
paulineHeb = pauline.copy()
paulineHeb.append('Heb')
johannine = ['John', '1John', '2John', '3John', 'Rev']

def freqCount(title):
    """ Returns the words (by Strong's number)
    in the book with title "title", along with the frequency
    as a decimal number (occurrences/total_words).
    ndarray: [[word, freq], ... ]
    word is a string containing a number, freq is a float """
    book = bookDict[title]
    word, count = np.unique(book, return_counts = True)
    wordCount = np.column_stack((word, count))
    ind = np.argsort(wordCount[:,1])[::-1]
    wordCountSorted = wordCount[ind]
    total_words = len(book)
    word, count = wordCountSorted[:, 0], wordCountSorted[:, 1]
    freq = count / total_words
    freqCount = np.column_stack((word, freq))
    return freqCount

def freqCountTwenty(title):
    """ Returns the top ten most used words (by Strong's)
    in the book with title "title", along with the frequency
```

```
47    as a decimal number ( occurrences / total_words ).
48    ndarray: [[ word , freq ], ...]
49    word is a string containing a number , freq is a float """
50    return freqCount ( title ) [0:20 ,:]
51
52  def freqDict ( title ):
53    """ Returns a dictionary of the frequencies of each word (by
54    Strong 's) in the book ( given by title ).  """
55    freqDict = dict ()
56    array = freqCount ( title )
57    for i in array :
58      word = i [0]
59      frequency = i [1]
60      freqDict [ word ] = frequency
61    return freqDict
62
63  def compare1 ( book1 , book2 , n ):
64    """ Returns the tuple (a , array ([[ word1 , freq1 , freq2 , s_1 ], ... [
      word5 ,
65    freq1 , freq2 , s_5 ]])
66    a = the similarity coefficient ( float ) between the two books
67    ( Closer to zero means more similar )
68    word1 through word5 are the words that give the largest
69    ( freq1 - freq2 )^2
70
71    How similarity coefficient is calculated :
72      Let { w_i } be the union of book1 's top n words and book2 's
73      top n words .
74      a = sum of s_i = abs ( log ( freq1_i ) - log ( freq2_i ))
75    """
76    dict1 , dict2 = freqDict ( book1 ), freqDict ( book2 )
77    top_n_1 , top_n_2 = freqCount ( book1 ) [0:n ,0] , freqCount ( book2 ) [0:n ,0]
78    word_array = np . union1d ( top_n_1 , top_n_2 )
79    word_freq1_freq2_s = list ()
80    for i in word_array :
81      try :
82        freq1 = float ( dict1 [ i ])
83        freq2 = float ( dict2 [ i ])
84        sumContribution = abs ( np . log ( freq1 ) - np . log ( freq2 ))
85      except KeyError :
86        sumContribution , freq1 , freq2 = 0 , 0 , 0
87        # if i in dict1 :
88        #     print (i , 'is only in ', book1 )
89        # else :
90        #     print (i , 'is only in ', book2 )
91      w_f1_f2_s = [i , freq1 , freq2 , sumContribution ]
92      word_freq1_freq2_s . append ( w_f1_f2_s )
93    wffs = np . array ( word_freq1_freq2_s )
94    a = np . sum ( wffs [: ,3]. astype ( np . float ))
95    ind = np . argsort ( wffs [: ,3]. astype ( np . float )) [:: -1]
96    wffs = wffs [ ind ]
```

```python
97      return (a, wffs[0:5,:])

98
99  def mostSimilarTo(book):
100     """ Returns array of books and similarity coefficient ranked
101     from most similar to least similar
102     [[book1, a1],
103      ...
104      [book26, a26]]
105     """
106     unrankedList = list()
107     for book2 in bookList:
108         if book2 == book:
109             continue
110         a = compare1(book, book2, 10)[0]
111         unrankedList.append([book2, a])
112     unrankedArray = np.array(unrankedList)
113     ind = np.argsort(unrankedArray[:,1].astype(np.float))
114     rankedArray = unrankedArray[ind]
115     return rankedArray

116
117 def printToTextFile(fileName):
118     """Prints the full results of mostSimilarTo for each book
119     in fileName"""
120     file = open(fileName,"a")
121     for book in bookList:
122         file.write(book + '\n')
123         similarityArray = mostSimilarTo(book)
124         file.write(np.array2string(similarityArray))
125         file.write('\n')
126     file.close()

127
128 def printToDict():
129     """Returns a dictionary with keys (book1, book2) -> a, for all
130     combinations of books"""
131     compareDict = dict()
132     for i in range(len(bookList)):
133         for j in range(i+1,len(bookList)):
134             book1, book2 = bookList[i], bookList[j]
135             a = compare1(book1, book2, 10)[0]
136             compareDict[(book1, book2)] = a
137             compareDict[(book2, book1)] = a
138     return compareDict

139
140 compareDict = printToDict()

141
142 def mostSimilarIn(subList, k):
143     """ Given a list of n books, subList, this returns the set of
144     k < n books that are most similar to each other, along with
145     the average value of a between all pairs within the subset.

146
147     Takes (list, int) -> tuple(list, float)
```

```
148
149     Sublist has nCk subsets of size k, and the returned subset mini-
150     mizes the sum of each degree of dissimilarity between all pairs
151     of books within the subset."""
152     mina = 1e30
153     (minSubset, avga) = (list(), 0)
154     for subset in itertools.combinations(subList, k):
155         a = 0
156         for pair in itertools.combinations(subset, 2):
157             pairTup = tuple(pair)
158             s = compareDict[pairTup]
159             a = a + s
160         if a < mina:
161             mina = a
162             (minSubset, avga) = (subset, a*2/(k*(k-1)))
163     return (minSubset, avga)
164
165 for i in range(2,14):
166     print(mostSimilarIn(pauline, i))
```