

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
CURSO DE CIÊNCIA DA COMPUTAÇÃO

CASSIANO DELGADO ARAUJO BRUCKHOFF

**Estudo crítico do Novo Testamento, dos
Pais Apostólicos e suas relações**

Monografia apresentada como requisito parcial
para a obtenção do grau de Bacharel em Ciência
da Computação

Orientador: Prof^a. Dr^a. Viviane Pereira Moreira

Porto Alegre
2022

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. Carlos André Bulhões Mendes

Vice-Reitora: Prof^a. Patricia Helena Lucas Pranke

Pró-Reitora de Graduação: Prof^a. Cíntia Inês Boll

Diretora do Instituto de Informática: Prof^a. Carla Maria Dal Sasso Freitas

Coordenador do Curso de Ciência de Computação: Prof. Rodrigo Machado

Bibliotecária-chefe do Instituto de Informática: Beatriz Regina Bastos Haro

*“Onde está o sábio? Onde está o erudito?
Onde está o questionador desta era?
Acaso não tornou Deus louca a sabedoria deste mundo?”*
— PAULO, 1 CORÍNTIOS 1:20

AGRADECIMENTOS

Este trabalho é dedicado aos meus pais, que dedicaram suas vidas para me possibilitar o melhor. Parafraseando uma célebre frase de Isaac Newton: se enxerguei, foi devido a dois gigantes: a minha mãe, Silvana, que me ensinou o significado de amor incondicional e a nunca desistir dos meus sonhos, e o meu pai, Edison, por me ensinar a seguir o que é correto e a realizar sempre o melhor trabalho possível.

À minha namorada, Richelly, pela compreensão e motivação ao longo de toda a graduação. Hoje consigo ver que crescemos muito como pessoas. Espero que consigamos, juntos, construir algo cada vez maior de hoje em diante.

À professora Viviane Moreira, que orientou o desenvolvimento deste trabalho com dedicação, me dando liberdade total para realizar as minhas escolhas, sempre com máximo de compreensão e paciência durante todo o processo.

Ao professor Leandro Wives, pelas importantes críticas construtivas e, em especial, ao professor José Palazzo, pelo tempo empregado na revisão do meu texto, bem como pela motivação de ter se interessado pelo meu trabalho.

Por fim, agradeço a Deus, pois, caso exista, é a causa primeira de tudo que existe e o objetivo final da vida. Este trabalho teve como objetivo, acima de tudo, de buscar a verdade, onde espero que este seja digno de ser considerado um passo nesta direção.

RESUMO

A Bíblia é um dos livros mais influentes e estudados em toda a história. Entretanto, mesmo com uma vasta gama de trabalhos, ainda há diversos problemas em aberto na área da crítica bíblica. O presente trabalho visa obter novas informações que contribuam para a discussão de alguns dos principais problemas da crítica bíblica na área do Novo Testamento e dos Pais Apostólicos. Nossa proposta consiste em utilizar técnicas de Processamento de Linguagem Natural pertencentes a três gerações distintas: TF-IDF, Doc2Vec e SBERT para aferir a similaridade por cosseno dos textos do Novo Testamento e das obras dos Pais Apostólicos, utilizando seus respectivos textos em sua língua original. Subsequentemente, nossas análises são comparadas com as relações documentadas por especialistas em crítica bíblica e comentaristas bíblicos. Além disto, evita-se o uso de traduções e de técnicas de redução de dimensionalidade para visualização dos dados, que poderiam levar a imprecisões significativas nos resultados. Os resultados obtidos apontam para: (i) a autoria exclusiva da Epístola aos Hebreus, diferindo-se dos escritos paulinos e das demais obras analisadas (ii) a dissimilaridade do Final Longo de Marcos com o restante da obra (iii) a plausibilidade de realocação da Perícopé da Adúltera do Evangelho de Lucas para o Evangelho de João (iv) a significativa similaridade do Capítulo 21 do Evangelho de João com o restante do texto. Ademais, foram estudadas outras numerosas relações do Novo Testamento, dos Pais Apostólicos e suas interações.

Palavras-chave: Processamento de linguagem natural. atribuição de autoria. crítica bíblica.

Critical study of the New Testament, the Apostolic Fathers, and their relationships

ABSTRACT

The Bible is one of the most influential and studied books in all history. However, even with a wide range of works, there are still several open problems in the area of biblical criticism. The present work aims to obtain new information to contribute to the discussion of some main problems of biblical criticism in the area of the New Testament and the Apostolic Fathers. Our proposal consists of using Natural Language Processing techniques belonging to three distinct generations: TF-IDF, Doc2Vec, and SBERT to assess the cosine similarity of the New Testament texts and the writings of the Apostolic Fathers, using their respective texts in their original language. Subsequently, our analyses are compared with documented relationships by experts in biblical criticism and biblical commentators. Furthermore, the use of translations and dimensionality reduction techniques for data visualization, which could lead to important inaccuracies in the results, is avoided. The results obtained point to: (i) the exclusive authorship of the Epistle to the Hebrews, differing from the Pauline writings and the other analyzed works (ii) the dissimilarity of the Long Ending of Mark with the rest of the work (iii) the plausibility of relocation the Pericope of the Adulteress from the Gospel of Luke to the Gospel of John (iv) the significant similarity of Chapter 21 of the Gospel of John with the rest of the text. In addition, numerous other relationships from the New Testament, the Apostolic Fathers, and their interactions have been studied.

Keywords: Natural language processing. authorship attribution. biblical criticism.

LISTA DE FIGURAS

Figura 2.1	Relações dos evangelhos sinóticos	16
Figura 3.1	Métodos de operação do Doc2Vec	27
Figura 3.2	Arquitetura do SBERT	29
Figura 3.3	Ilustração das distorções dos resultados da análise TF-IDF utilizando técnicas de redução de dimensionalidade	31
Figura 5.1	Comparação do resultado do pré-processamento de um versículo.....	36
Figura 6.1	Ilustração do processo de <i>leave-one-out cross-validation</i>	41
Figura 6.2	Comparações de similaridade dos evangelhos pela análise Doc2Vec	45
Figura 6.3	Comparações de similaridade do Final Longo de Marcos pela análise Doc2Vec	46
Figura 6.4	Comparações de similaridade da Perícope da Adúltera pela análise Doc2Vec	46
Figura 6.5	Comparações de similaridade do Capítulo 21 do Evangelho de João pela análise Doc2Vec	47
Figura 6.6	Comparações de similaridade de Atos dos Apóstolos pela análise Doc2Vec	47
Figura 6.7	Comparações de similaridade das epístolas autênticas de Paulo pela análise Doc2Vec	48
Figura 6.8	Comparações de similaridade das epístolas deuteropaulinas pela análise Doc2Vec	49
Figura 6.9	Comparações de similaridade das epístolas pastorais pela análise Doc2Vec	50
Figura 6.10	Comparações de similaridade da Epístola aos Hebreus pela análise Doc2Vec	51
Figura 6.11	Comparações de similaridade da Epístola de Tiago pela análise Doc2Vec.	51
Figura 6.12	Comparações de similaridade das epístolas petrinas e Judas pela análise Doc2Vec	52
Figura 6.13	Comparações de similaridade das epístolas joaninas pela análise Doc2Vec	53
Figura 6.14	Comparações de similaridade do Apocalipse pela análise Doc2Vec.....	54
Figura 6.15	Comparações de similaridade das epístolas de Clemente de Roma pela análise Doc2Vec	55
Figura 6.16	Comparações de similaridade das epístolas de Inácio de Antioquia pela análise Doc2Vec	56
Figura 6.17	Comparações de similaridade da Epístola de Policarpo aos Filipenses pela análise Doc2Vec	57
Figura 6.18	Comparações de similaridade do Martírio de Policarpo pela análise Doc2Vec	57
Figura 6.19	Comparações de similaridade do Didaquê pela análise Doc2Vec	57
Figura 6.20	Comparações de similaridade da Epístola de Barnabé pela análise Doc2Vec	58
Figura 6.21	Comparações de similaridade do Pastor de Hermas pela análise Doc2Vec	58
Figura 6.22	Comparações de similaridade da Epístola a Diogneto pela análise Doc2Vec	59

LISTA DE TABELAS

Tabela 4.1	Comparação dos trabalhos relacionados	33
Tabela 6.1	Parâmetros do Doc2Vec.....	40
Tabela 6.2	Avaliação comparativa das análises utilizando amostras em blocos de palavras	42
Tabela 6.3	Lista de abreviações adotada	44

LISTA DE ABREVIATURAS E SIGLAS

BERT	Bidirectional Encoder Representations from Transformers
BOW	Bag-of-Words
CA	Correspondence Analysis
DBOW	Distributed Bag-of-Words
DM	Distributed Memory
LLE	Locally Linear Embedding
LXX	Septuaginta
MDA	Multiple Discriminant Analysis
ML	Machine Learning
NT	Novo Testamento
PCA	Principal Component Analysis
PLN	Processamento de Linguagem Natural
SBERT	Sentence Bidirectional Encoder Representations from Transformers
SBLGNT	Society of Biblical Literature Greek New Testament
SHG	Standard Hellenistic Greek
TF-IDF	Term Frequency-Inverse Document Frequency
WE	Word Embeddings

SUMÁRIO

1 INTRODUÇÃO	12
2 REFERENCIAL TEOLÓGICO	15
2.1 Novo Testamento	15
2.1.1 Evangelhos	15
2.1.2 Atos dos Apóstolos	17
2.1.3 Epístolas Paulinas	18
2.1.4 Epístola aos Hebreus	19
2.1.5 Epístola de Tiago	19
2.1.6 Epístolas Petrinas e Judas	20
2.1.7 Epístolas Joaninas	20
2.1.8 Apocalipse	21
2.2 Pais Apostólicos	21
2.2.1 Epístolas de Clemente de Roma	22
2.2.2 Epístolas de Inácio de Antioquia	22
2.2.3 Epístola de Policarpo aos Filipenses	23
2.2.4 Martírio de Policarpo	23
2.2.5 Didaquê	23
2.2.6 Epístola de Barnabé	24
2.2.7 Pastor de Hermas	24
2.2.8 Epístola a Diogneto	24
2.3 Sumário	25
3 REFERENCIAL TÉCNICO	26
3.1 TF-IDF	26
3.2 Doc2Vec	26
3.3 SBERT	28
3.4 Redução de Dimensionalidade	29
4 TRABALHOS RELACIONADOS	32
5 METODOLOGIA	35
5.1 Proposta	35
5.2 Pré-processamento	35
5.3 Pipeline TF-IDF	36
5.4 Pipeline Doc2Vec	37
5.5 Pipeline SBERT	37
5.6 Sumário	38
6 ANÁLISE EXPERIMENTAL	39
6.1 Configuração Experimental	39
6.1.1 <i>Corpora</i>	39
6.1.2 Ferramentas	39
6.1.3 Parametrização	40
6.2 Validação	41
6.3 Resultados	43
6.3.1 Evangelhos	44
6.3.2 Final Longo de Marcos	45
6.3.3 Perícope da Adúltera	46
6.3.4 Capítulo 21 do Evangelho de João	47
6.3.5 Atos dos Apóstolos	47
6.3.6 Epístolas Paulinas	48
6.3.7 Epístola aos Hebreus	50

6.3.8 Epístola de Tiago	51
6.3.9 Epístolas Petrinas e Judas	52
6.3.10 Epístolas Joaninas	53
6.3.11 Apocalipse.....	54
6.3.12 Epístolas de Clemente de Roma	54
6.3.13 Epístolas de Inácio de Antioquia	55
6.3.14 Epístola de Policarpo aos Filipenses.....	55
6.3.15 Martírio de Policarpo	56
6.3.16 Didaquê.....	57
6.3.17 Epístola de Barnabé	57
6.3.18 Pastor de Hermas	58
6.3.19 Epístola a Diogneto.....	58
6.4 Sumário.....	59
7 CONCLUSÃO	60
REFERÊNCIAS.....	62

1 INTRODUÇÃO

A Bíblia é um dos livros mais influentes e estudados de toda a história. Até mesmo Isaac Newton, famoso por suas importantes contribuições nas áreas da física e matemática, fez inúmeros estudos na área da teologia, culminando em obras como *Observações sobre as Profecias de Daniel e o Apocalipse de São João*. A confiabilidade e autoridade bíblica são importantes pontos do cristianismo, pontos estes que foram escrutinados ao longo da história. Wallace (2011) afirma que o Novo Testamento (NT) é, de longe, mais atestada que qualquer obra da literatura greco-romana antiga e que nenhuma doutrina fundamental da fé cristã foi corrompida.

Entretanto, mesmo com uma vasta gama de trabalhos, ainda há diversos problemas em aberto na área da crítica bíblica. A crítica bíblica tem como objetivo empregar critérios científicos e a razão humana para compreender o texto bíblico (HARRINGTON, 2011). O presente trabalho almeja obter novas informações para contribuir na discussão de alguns dos principais problemas da crítica bíblica na área do Novo Testamento e dos Pais Apostólicos. Um maior entendimento das relações dos textos bíblicos pode ajudar a diversos públicos, desde teólogos e cristãos, que estudam o desenvolvimento do cânon bíblico e da teologia alicerçada sobre tal texto, até historiadores, que utilizam o texto bíblico como fonte documental e estudam o seu contexto histórico.

Devido à antiguidade dos escritos, datados por volta do primeiro século (BROWN; SOARDS, 2016, p. 18), a tarefa de atribuir autores ao texto bíblico se torna uma tarefa complexa, pois há grandes perdas de informações acarretados pela destruição de escritos e informações de seu contexto pelos mais variados motivos: guerras, incêndios, disputas entre seitas cristãs e até mesmo pelo Império Romano, como no episódio da Grande Perseguição, empregada por Diocleciano no ano 303 (GADDIS, 2015, p. 29). O anonimato e a criação de pseudoepígrafos¹ também contribuem para dificultar a tarefa. Devido a isso, as maiores fontes de informações para tais inferências se tratam do próprio texto bíblico e da tradição sobrevivente que se seguiu, conhecida como Patrística.

Há diversas abordagens utilizadas atualmente para analisar os textos bíblicos. A análise manual dos escritos por especialistas em crítica textual e da história da igreja cristã levou a vastas contribuições na área, sendo que algumas delas, serão expostas e utilizadas pelo presente trabalho. Entretanto, muitas dessas análises trazem vieses de seus autores, bem como de suas crenças, denominações religiosas, escolas de pensamentos

¹Obra falsamente atribuída a uma pessoa que não escreveu tal texto.

e/ou metodologias empregadas.

Por outro lado, há uma adoção crescente de métodos computacionais na área de crítica textual. Nestes esforços, em muitos casos, há o emprego de técnicas clássicas como o modelo *Bag-of-Words* (BOW), podendo ser acompanhadas por técnicas de redução de dimensionalidade como o PCA (Principal Component Analysis) para viabilizar análises visuais. Todavia, ao utilizar técnicas de redução de dimensionalidade para esse fim sem qualquer critério, corre-se o risco de perda de informações e imprecisões importantes nos resultados, como será demonstrado na Seção 3.4. Outros trabalhos utilizam traduções dos textos, o que pode levar a perda de riqueza vocabular do texto original. Há também trabalhos que não consideram uma maior gama de textos para ter uma análise mais completa, limitando a confiabilidade dos resultados obtidos.

O presente trabalho tem dois objetivos principais: primeiramente, verificar diversas relações do Novo Testamento e dos escritos dos Pais Apostólicos, avaliando as hipóteses de especialistas em crítica textual e comentaristas bíblicos. Posteriormente, serão encaminhadas análises específicas de quatro dos principais problemas textuais do Novo Testamento:

1. Inferir a autoria da anônima Epístola aos Hebreus.
2. Avaliar a hipótese de realocação da chamada Perícope da Adúltera.
3. Aferir nível de similaridade do Final Longo de Marcos com o restante da obra.
4. Avaliar coesão textual do Capítulo 21 do Evangelho de João com as demais partes do texto.

Nossa proposta consiste em utilizar técnicas pertencentes a três gerações distintas de algoritmos de Processamento de Linguagem Natural (PLN) para aferir a similaridade por cosseno entre os textos do Novo Testamento e as obras dos Pais Apostólicos, utilizando seus respectivos textos em sua língua original. A primeira análise baseia-se na representação de vetores esparsos conhecida como *Bag-of-Words* (BOW), onde os pesos são dados pela ponderação TF-IDF (*Term Frequency-Inverse Document Frequency*). Na segunda análise, utilizamos *Paragraph Vectors*, representações vetoriais densas obtidas através do algoritmo Doc2Vec (LE; MIKOLOV, 2014). A terceira análise foi feita utilizando representações de *embeddings* contextuais obtidas a partir de grandes modelos de linguagem pré-treinados. O modelo escolhido foi o SBERT (Sentence-BERT) (REI-MERS; GUREVYCH, 2019) que atinge o estado da arte em diversas tarefas de PLN.

Os resultados gerados pelas técnicas foram comparados com as relações relatadas

por especialistas. Com isso, é esperado que o emprego de diferentes técnicas, comparações adequadas dos dados obtidos e o uso de uma maior gama de textos em sua língua original tragam resultados relevantes no estudo dos textos do Novo Testamento, dos Pais Apostólicos e suas relações.

Em relação aos resultados, no Novo Testamento podemos destacar a similaridade significativa entre o grupo de cartas autênticas de Paulo e as cartas deuteropaulinas. Nas análises dos Pais Apostólicos, a relação de dissimilaridade e assimetria das epístolas de Clemente de Roma se destaca. Dentre os quatro problemas textuais específicos mencionados nos objetivos, os resultados obtidos apontam a dissimilaridade da Epístola aos Hebreus com as cartas paulinas, a plausibilidade da realocação da Perícopé da Adúltera, a baixa afinidade do Final Longo de Marcos com o restante da obra e a coesão textual do Capítulo 21 do Evangelho de João com o restante do evangelho. Estes e outros resultados relevantes são discutidos na Seção 6.3.

O presente trabalho está organizado da seguinte forma:

- Capítulo 2: neste capítulo foram expostos os conhecimentos teológicos básicos para compreender o contexto do presente trabalho e as relações textuais estudadas.
- Capítulo 3: aqui foram introduzidos os algoritmos utilizados e detalhes técnicos necessários para embasar as escolhas realizadas ao longo do trabalho.
- Capítulo 4: neste capítulo foi realizada uma revisão bibliográfica de trabalhos relacionados, exibindo uma análise comparativa entre os trabalhos analisados e o presente trabalho.
- Capítulo 5: este capítulo detalha a metodologia adotada, procedimentos e *pipelines* necessários para obter os resultados experimentais desejados.
- Capítulo 6: neste capítulo foram detalhados os itens necessários para implementar a metodologia empregada, apresenta a validação das análises, exhibe e discute resultados obtidos, bem como apresenta hipóteses para explicar tais resultados.
- Capítulo 7: este capítulo é composto pelo resumo do trabalho e dos resultados obtidos, conclusões e elenca possíveis trabalhos futuros.

2 REFERENCIAL TEOLÓGICO

Neste Capítulo serão expostos os conhecimentos teológicos básicos para compreender o contexto do presente trabalho e as relações textuais estudadas.

2.1 Novo Testamento

Contendo 27 livros, o cânon do Novo Testamento foi escrito em sua totalidade em grego koiné ou grego helenístico, língua franca da região oriental do Mediterrâneo depois das conquistas de Alexandre Magno (BARTON; MUDDIMAN, 2001, p. 830–843). O Novo Testamento é produto de um processo orgânico ao longo da história da igreja primitiva e de seus testemunhos, culminando na cristalização da lista de livros inspirados em forma de cânons, como o Cânon Muratoriano¹. A seguir serão detalhadas as respectivas relações exploradas pelo presente trabalho.

2.1.1 Evangelhos

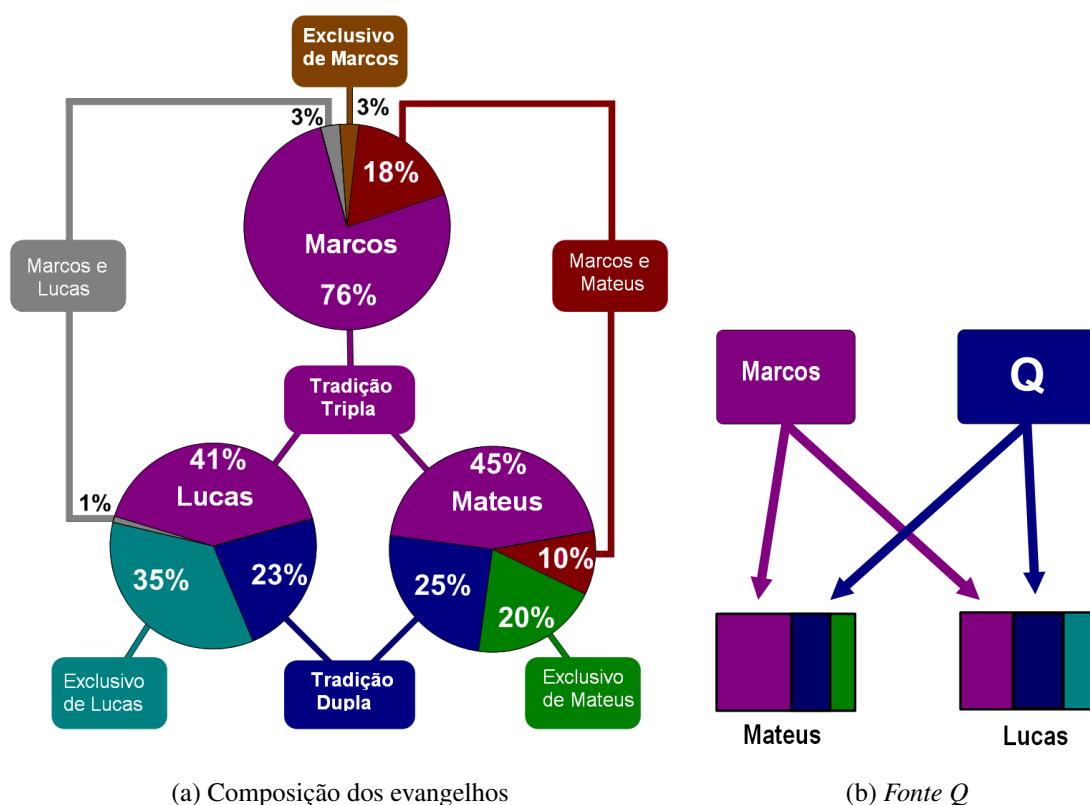
A palavra evangelho tem origem grega e significa *boa-nova*, sendo esta a mensagem da salvação contada a partir da vida, morte e ensinamentos de Jesus Cristo (BROWN; SOARDS, 2016, p. 33). Os evangelhos não tiveram a intenção de ser um relato biográfico da vida de Jesus, mas sim documentos teológicos que pudessem alicerçar a fé cristã (BROWN; SOARDS, 2016, p. 36). Para fins ilustrativos, podemos dividir o conteúdo dos evangelhos em duas partes: a primeira parte contando sobre o ministério, discursos e ensinamentos de Jesus Cristo e uma segunda parte com as acusações recebidas, prisão, crucificação e relatos de sua ressurreição.

Há quatro evangelhos considerados canônicos, ou seja, considerados inspirados e incluídos no cânon bíblico. São eles: Mateus, Marcos, Lucas e João. Os três primeiros evangelhos são chamados evangelhos sinóticos, pois têm extensas porções em comum em suas narrativas, conforme a Figura 2.1a, por esse motivo eram estudados juntos, em paralelo, de onde surgiu a palavra sinótico, que significa *ver junto*. Essa relação foi estudada, sendo denominada de problema sinótico (BROWN; SOARDS, 2016, p. 36–38), onde seus paralelos podem ser exemplificados pelos versículos idênticos compartilhados entre os

¹Uma das listas mais antigas de livros pertencentes ao Novo Testamento, datada do ano 170.

três evangelhos, conforme White (2009, p. 289). A hipótese moderna mais aceita para resolver esse problema assume a primazia do Evangelho de Marcos, que teria servido como base para os outros dois evangelhos sinóticos, que, por sua vez, teriam utilizado ainda uma possível fonte teorizada, denominada *fonte Q*, vide Figura 2.1b, entretanto, não há consenso nesse problema, pois nenhuma hipótese resolve todos os problemas levantados (BROWN; SOARDS, 2016, p. 39–43).

Figura 2.1 – Relações dos evangelhos sinóticos



Fonte: Wikipédia <<https://bit.ly/3LJnHbH>> <<https://bit.ly/3OWPrM0>>

O Evangelho de Marcos é tradicionalmente atribuído a João Marcos², cooperador dos apóstolos Pedro e Paulo. Este evangelho tem uma suposta adição em seu capítulo final que narra as aparições de Jesus após sua ressurreição. Segundo Barton e Muddiman (2001, p. 921), essa adição, conhecida como Final Longo de Marcos³, não aparece nos manuscritos mais antigos e apresenta diferenças significativas de estilo em relação ao restante do evangelho. Contudo, essa passagem apresenta um amplo respaldo da Patrística, sendo atestada por Justino Mártir, Irineu de Lyon, entre outros. Há uma teoria que Aríston de Esmirna seria o autor deste final, isso baseado em uma anotação⁴ do manuscrito

²At 12:12,25; At 15:37

³Mc 16:9-20

⁴Pequena anotação entre os versículos 16:8 e 16:9 que pode ser traduzida como: *de Aríston, o presbítero*.

armênio *Matenadaran 2374* (SNAPP JR., 2007).

Pápias de Hierápolis aponta o apóstolo Mateus como autor do chamado Evangelho de Mateus, sendo apontado como uma tradução de um texto original em hebraico ou aramaico, entretanto, esse testemunho é amplamente questionado (BARTON; MUDDIMAN, 2001, p. 844). Já o Evangelho de Lucas, teria sido composto através de uma investigação⁵ dos fatos da vida de Jesus conduzida por Lucas, cooperador de Paulo⁶. Hoje é consenso que este evangelho foi composto pela mesma pessoa que escreveu Atos dos Apóstolos (BARTON; MUDDIMAN, 2001, p. 1028).

O Evangelho de João distingue-se significativamente dos evangelhos sinóticos em diversos aspectos, como na ausência de parábolas, estilo das narrativas e detalhes. O foco principal deste evangelho é demonstrar a divindade de Jesus Cristo através de sete milagres testemunhados pelo próprio apóstolo João e outros discípulos de Jesus⁷. A tradição também elenca o apóstolo João como o autor do texto, já no fim de sua vida em Éfeso (BARTON; MUDDIMAN, 2001, p. 960–961).

Existem dois textos deste evangelho com grandes controvérsias na área da crítica textual (KOESTER, 1995, v. 2 p. 192). O primeiro texto é a chamada Perícope⁸ da Adúltera⁹, alvo de controvérsia por não estar contida nos manuscritos mais antigos hoje conhecidos e ter uma fraca atestação da Patrística, a exceção de uma possível atestação de Pápias de Hierápolis e de algumas obras posteriores (HOLMES, 2006, p. 723–727). Entretanto, há uma hipótese que essa perícope pertença ao Evangelho de Lucas e tenha sido realocada indevidamente. Essa hipótese se sustenta no fato de que o chamado grupo de manuscritos Ferrar¹⁰ trazerem a respectiva perícope no Evangelho de Lucas. O segundo texto é o Capítulo 21 de João, pois, supostamente, este último capítulo seria uma adição posterior com uma repetição exagerada do final original, que originalmente seria em João 20:30-31. Entretanto, todos os manuscritos encontrados possuem o questionado capítulo.

2.1.2 Atos dos Apóstolos

O livro de Atos dos Apóstolos segue a narrativa dos evangelhos, a partir da ascensão de Jesus, descrevendo o contexto das primeiras conversões, desenvolvimento das

⁵Lc 1:1-4

⁶2Tm 4:11; Fm 1:24

⁷Jo 20:30-31; Jo 21:24-25

⁸Conjunto de versículos que constituem uma unidade textual com conteúdo coeso.

⁹Jo 7:53-8:1-11

¹⁰Também conhecido como Família 13 de manuscritos.

primeiras comunidades cristãs, perseguições e viagens missionárias. Atos e o Evangelho de Lucas são vistos como um único volume, direcionado a alguém chamado por Teófilo, que pode ser um nome ou título honorífico de alguma autoridade. Tradicionalmente, se atribui a Lucas os dois volumes, entretanto, foram notadas divergências entre estes escritos e as cartas de Paulo, levando ao surgimento de hipóteses que sugerem que o autor deste volume não seja o cooperador de Paulo em suas viagens missionárias (BARTON; MUDDIMAN, 2001, p. 1028).

2.1.3 Epístolas Paulinas

As treze epístolas de Paulo são cartas endereçadas às igrejas e seus líderes com sistematizações e orientações doutrinárias, admoestações à fé e seus testemunhos, contribuições que são fundamentais à fé cristã. Em muitas destas cartas, são citados coautores, como Timóteo, Silvano¹¹ e Sóstenes. Paulo, provavelmente, empregava secretários, como Tércio¹², para escrever diversas de suas epístolas, algo comum na antiguidade (RICHARDS, 2004, p. 60–93). Normalmente, essas cartas são divididas em três grupos: autênticas, deuteropaulinas e pastorais (EHRMAN, 2014). A seguir serão detalhadas cada uma dessas categorias e suas peculiaridades.

São chamadas *autênticas* as sete cartas do apóstolo Paulo em que há consenso entre a tradição e os críticos modernos. São elas: Romanos, 1 e 2 Coríntios, Gálatas, Filipenses, 1 Tessalonicenses e Filemom. Estas cartas apresentam um núcleo de temáticas, teologia e estilo de escrita similares e coeso (BARTON; MUDDIMAN, 2001, p. 1077).

Há também as cartas chamadas *deuteropaulinas*, por formarem um, possível, segundo *corpus*, são epístolas disputadas e que não há uma unanimidade quanto a autoria. São elas: 2 Tessalonicenses, Colossenses e Efésios. A carta de 2 Tessalonicenses é questionada, pois, diverge da primeira carta em escolhas de palavras, estrutura e teologia, mesmo apresentando temas similares (BARTON; MUDDIMAN, 2001, p. 1213). Colossenses e Efésios são apontadas como tendo uma relação de dependência, onde é mostrado que Efésios seria uma expansão do conteúdo de Colossenses (BARTON; MUDDIMAN, 2001, p. 1165–1167).

Por fim, o grupo de epístolas *pastorais* é composto por 1 e 2 Timóteo e Tito. São

¹¹Provavelmente trata-se de uma versão latina do nome Silas. Silas foi um cooperador de Paulo em Atos dos Apóstolos, como no episódio descrito em At 15:22.

¹²Rm 16:22

chamadas assim por serem endereçadas à liderança das igrejas, e não às igrejas propriamente, como nas cartas anteriores. A crítica moderna classifica amplamente essas cartas como pseudoepígrafos que utilizaram o nome de Paulo indevidamente, consequência de discrepâncias entre o estilo de linguagem, conteúdo, entre outros fatores (MARSHALL; TOWNER, 2004, p. 79).

Apesar destes dois últimos grupos das epístolas paulinas serem questionadas em relação a sua autenticidade, elas constam no Cânon Muratoriano, além de serem citados nos escritos dos Pais Apostólicos, como Inácio de Antioquia e Policarpo de Esmirna.

2.1.4 Epístola aos Hebreus

A chamada Epístola aos Hebreus se trata de um sermão ou tratado teológico de autoria anônima que se detém no desenvolvimento de uma cristologia, trazendo paralelos com o Antigo Testamento, discutindo a relação dos cristãos com a tradição judaica e a aliança com Deus. Tradicionalmente atribuída a Paulo, hoje é consenso da crítica moderna que Paulo não seja o autor deste texto devido a diversas divergências, desde o uso de uma linguagem mais polida, sua coerência contextual e estrutura adotada difere do estilo de redação do apóstolo. Diversos outros autores foram especulados, como Clemente de Roma, Lucas, Barnabé, entre outros (BARTON; MUDDIMAN, 2001, p. 1236–1237).

2.1.5 Epístola de Tiago

A Epístola de Tiago é um texto de exortação e fonte de doutrinas em temas diversos, entretanto, sem uma grande estrutura aparente costurando o texto. Seu texto é desenvolvido utilizando um grego bem escrito, entretanto, que apresenta semitismos. Foram levantados alguns possíveis paralelos com 1 Pedro, 1 Clemente e Pastor de Hermas, entretanto, também foi notado serem paralelos em temas comuns ao texto bíblico (BARTON; MUDDIMAN, 2001, p. 1255–1257).

2.1.6 Epístolas Petrinas e Judas

As cartas de 1 e 2 Pedro são atribuídas ao apóstolo Pedro devido a sua identificação inicial nas epístolas¹³, um pescador provinciano da Galileia¹⁴, porém as duas divergem mutualmente em conceitos, estilo e terminologia, além da linguagem esperada pela cultura do autor. O autor de 1 Pedro emprega uma escrita sofisticada, o uso da tradução LXX¹⁵ e de retórica (BARTON; MUDDIMAN, 2001, p. 1263–1264), já o autor de 2 Pedro elabora seu texto com grego bem escrito, contudo, expressando um pensamento de um judeu helenizado, algo revelado por seus semitismos contidos no texto (BARTON; MUDDIMAN, 2001, p. 1270–1272).

Existem teorias que apontam o uso de secretários na escrita da epístola de 1 Pedro, onde, possivelmente, Silvano¹⁶ foi o redator, entretanto, também é relatado que essa explicação dificilmente ajuda na questão da autoria petrina.

A epístola de 2 Pedro apresenta inúmeros paralelos com a Epístola de Judas, conforme listados em Gruen (2018, p. 15–19), levando a ser especulada como uma expansão desta última. Judas é uma carta curta, porém com temáticas densas, trazendo temas do Antigo Testamento e, até mesmo, possíveis citações de pseudoepígrafos¹⁷ (BARTON; MUDDIMAN, 2001, p. 1284–1285).

2.1.7 Epístolas Joaninas

As três epístolas de João exemplificam bem a literatura hebraica, escritas com um estilo simples e frequentemente recorrendo a repetições e paralelismos. Os temas variam entre testemunhos de fé, orientações morais e denúncias de heresias. Tradicionalmente, estas cartas são agrupadas junto no aspecto de autoria, demonstrando ter linguagem e ideias em comum com o Evangelho de João (BARTON; MUDDIMAN, 2001, p. 1274–1275).

¹³ 1Pe 1:1; 2Pe 1:1

¹⁴ At 4:13

¹⁵ Conhecida também como Versão dos Setenta ou Septuaginta, trata-se da tradução mais antiga do Antigo Testamento para o grego.

¹⁶ 1Pe 5:12

¹⁷ Assunção de Moisés; 1 Enoque 1:9 cf. *The Apocrypha and Pseudepigrapha of the Old Testament*. Disponível em: <<https://bit.ly/3KTHcgq>>

2.1.8 Apocalipse

Apocalipse é um livro profético que nomeia o seu próprio gênero literário, a literatura apocalíptica, exemplificado pelo livro de Daniel, no Antigo Testamento. O livro trata de uma série de visões que João teve na ilha de Patmos¹⁸, João este que a tradição aponta ser o apóstolo João. Outros já levantaram questionamentos devido ao seu grego diferir do Evangelho de João e das epístolas, considerado ruim e gramaticalmente peculiar (LAWRENCE, 2019, p. 173), contudo, o autor apresenta vasto conhecimento das escrituras hebraicas através de alusões, elaborando o texto em uma intrincada estrutura de quiasmo¹⁹ (BARTON; MUDDIMAN, 2001, p. 1287–1288).

2.2 Pais Apostólicos

Os Pais Apostólicos são aqueles que conheceram pessoalmente pelo menos um dos apóstolos de Jesus Cristo, geralmente em uma relação de discipulado, e se tornaram autoridades e importantes testemunhas para as demais gerações de cristãos. São eles: Clemente de Roma, Inácio de Antioquia, Policarpo de Esmirna, Pápias de Hierápolis e Quadratus de Atenas. Todos tiveram obras escritas que, em muitos casos, foram altamente estimados e até mesmo citadas como canônicas. Pápias e Quadratus não tiveram suas obras preservadas, hoje, conhecidas somente por trechos de citações. As obras preservadas dos Pais Apostólicos são datadas entre o século I e II e ilustram aspectos como o início do cristianismo, seus problemas, doutrinas e testemunhos, sendo os únicos relatos extra-bíblicos de diversos acontecimentos e costumes da época (BARTON; MUDDIMAN, 2001, p. 1308). A seguir serão contextualizadas as obras atribuídas aos Pais Apostólicos ou que tem alguma relação com eles e serão utilizadas nas análises do presente trabalho.

¹⁸Ap 1:9

¹⁹Quiasmo é um variação de paralelismo realizada de forma cruzada, geralmente para enfatizar alguma parte do texto. Por exemplo, no Apocalipse, os capítulos iniciais tem paralelismos com os capítulos finais, enfatizando os capítulos centrais.

2.2.1 Epístolas de Clemente de Roma

A epístola de 1 Clemente é uma das cartas mais antigas do cristianismo, do mesmo período ou anterior a diversos escritos do Novo Testamento. A carta traz uma mensagem de admoestação moral, fazendo diversas referências ao Antigo e Novo Testamento, assim como ao martírio de Pedro e Paulo. Apesar de ser anônima, é, geralmente, aceita como escrita por Clemente de Roma, considerado o quarto papa pela Igreja Católica e que, possivelmente, foi citado no Novo Testamento como um dos cooperadores de Paulo²⁰ (HOLMES, 2006, p. 36–40). A carta de 2 Clemente, por sua vez, é considerada um pseudoepígrafo (FATHERS; GALLI, 2009, p. 11). O texto pode se tratar de uma transcrição de uma homilia, sendo então, umas das homilias mais antigas conhecidas (BARTON; MUDDIMAN, 2001, p. 1314). Os dois escritos foram incluídos no *Codex Alexandrinus*²¹.

2.2.2 Epístolas de Inácio de Antioquia

As sete epístolas de Inácio de Antioquia apresenta os últimos testemunhos de fé devido ao seu martírio eminente em Roma. Também realiza admoestações a obediência às autoridades religiosas locais e repele heresias, como o legalismo, gnosticismo e docetismo (FATHERS; GALLI, 2009, p. 11–12). Das sete cartas, seis delas são direcionadas as igrejas na rota do martírio e uma carta a Policarpo, seu amigo pessoal (BARTON; MUDDIMAN, 2001, p. 1314–1315). São elas: Efésios, Esmirniotas, Filadélfos, Magnésios, Romanos, Trálios e a Epístola a Policarpo.

Seu estilo de redação foi descrito como distintivo e que seus textos apresentam influências dos escritos paulinos, como 1 Coríntios e do Evangelho de Mateus, bem como exhibe paralelos com 1 João e com obras de outros Pais Apostólicos, como 1 e 2 Clemente e Pastor de Hermas (HOLMES, 2006, p. 87–95).

²⁰Fp 4:3

²¹Um dos quatro grandes códices unciais, ao lado do *Codex Vaticanus*, *Alexandrinus* e *Ephraemi Rescriptus*, usados como textos-base das traduções modernas da Bíblia. Os códices são os precursores dos livros modernos, que invés de serem compostos de papel, geralmente, são feitos de papiro, velino ou outros materiais. A palavra uncial refere-se ao estilo de escrita empregado, que utiliza somente letras maiúsculas e com possível embelezamento do texto, estilo este normalmente empregado entre os séculos IV e VIII.

2.2.3 Epístola de Policarpo aos Filipenses

Policarpo era o bispo de Esmirna que, segundo seu discípulo, Irineu de Lyon, foi ordenado para tal cargo pelos próprios apóstolos²². A única obra de Policarpo que sobreviveu foi a sua carta endereçada aos Filipenses. Escrita originalmente em grego, sendo hoje necessário completar lacunas utilizando uma tradução em latim, pois não foi preservada integralmente em grego. Esta obra faz uso extensivo do Antigo Testamento, em sua tradução grega LXX, e do Novo Testamento (HARTOG, 2013, p. 53–61).

2.2.4 Martírio de Policarpo

O relato anônimo do Martírio de Policarpo traz uma estrutura inicial de epístola para narrar como foi a execução de Policarpo em Esmirna, condenado a morte por fogueira por não aceitar o culto imperial imposto por Roma. O relato deste martírio teve como objetivo encorajar os cristãos a fé em Cristo e alertar contra a renegação da fé, também se destaca por ser um dos relatos mais antigos de um martírio cristão por uma fonte extra-bíblica (FATHERS; GALLI, 2009, p. 13). Também apresenta algumas intervenções miraculosas e paralelos com a vida e morte de Jesus Cristo como descritas nos evangelhos.

2.2.5 Didaquê

O Didaquê ou Doutrina dos Doze Apóstolos é um tratado de orientações teológicas de autoria anônima que pode ser dividido em duas partes: a primeira parte contém orientações morais exemplificados como escolhas entre dois caminhos, a vida e a morte. Já a segunda parte conta com orientações doutrinarias sobre o batismo, jejuns, ordenamento, entre outros. Foram notados paralelos entre o Didaquê e outros escritos como o Evangelho de Mateus, os escritos de Lucas e a Epístola de Barnabé (HOLMES, 2006, p. 157-162).

²²*Against Heresies*, III.3.4. Disponível em: <<https://bit.ly/3KW7Chl>>

2.2.6 Epístola de Barnabé

Epístola de Barnabé é um tratado ou ensaio teológico que mostra similaridades com a Epístola aos Hebreus devido ao estilo de escrita e suas temáticas como interpretação e aplicação do Antigo Testamento e sua discussão sobre a aliança com Deus, porém, utilizando uma interpretação alegórica dos textos. Também cita a analogia dos dois caminhos, agora entre a luz e as trevas, similar ao que é apresentado no Didaquê e apresenta possíveis citações do Evangelho de Mateus. Tradicionalmente atribuída a Barnabé, cooperador de Paulo, essa obra foi incluída no *Codex Sinaiticus*, contudo, críticos atribuem a autoria a algum judeu de Alexandria (HOLMES, 2006, p. 172–176).

2.2.7 Pastor de Hermas

A obra Pastor de Hermas trata-se de cinco visões, doze mandamentos e dez parábolas recebidas por um servo chamado Hermas. O Cânon Muratoriano cita o autor como o próprio Hermas, irmão do papa Pio I. Tendo similaridades com Apocalipse por seu uso de alegorias e visões, essa obra tem o intuito mais didático e instruções claras do que é esperado de um cristão (FATHERS; GALLI, 2009, p. 14–15). Este escrito foi bem recebido por seus contemporâneos e, inclusive, foi citado como canônico por Irineu de Lyon, também sendo incluído no *Codex Sinaiticus* (BARTON; MUDDIMAN, 2001, p. 1314).

2.2.8 Epístola a Diogneto

Epístola a Diogneto é um tratado apologético anônimo, onde o autor se identifica como *um discípulo dos Apóstolos*. Foram notadas semelhanças com 1 Clemente e também especulados outros autores, como Apolo e Justino Mártir²³. Holmes (2006) ainda cita Policarpo e outros Pais Apostólicos como possíveis autores, além de uma conjectura que esta epístola seria a obra Apologia de Quadratus de Atenas que foi perdida. A obra se trata do mais antigo texto apologético cristão de que se tem conhecimento, sendo uma resposta a supostas indagações de Diogneto sobre o cristianismo.

²³ *Introductory Note to the Epistle of Mathetes to Diognetus*. Disponível em: <<https://bit.ly/3whIIDO>>

2.3 Sumário

Neste Capítulo foi desenvolvida uma breve contextualização dos textos, bem como uma exposição das relações do Novo Testamento e dos Pais Apostólicos que serão estudadas. Este referencial teológico será importante para embasar as análises da Seção 6.3.

3 REFERENCIAL TÉCNICO

Neste Capítulo serão expostos os conhecimentos técnicos básicos para compreender as técnicas utilizadas e as escolhas adotadas durante o restante do presente trabalho.

3.1 TF-IDF

O TF-IDF é uma medida estatística da importância de um termo em relação a uma coleção de documentos. Apesar de sua simplicidade, essa técnica consegue obter resultados relevantes em tarefas de recuperação de informações e mineração de dados. Extensamente utilizado na indústria, o TF-IDF é o modelo de ponderação mais aplicado em sistemas de recomendação (BEEL et al., 2016).

O objetivo do cálculo é valorizar os termos que diferenciam os documentos, ao passo que penaliza termos com alta frequência que surgem na maioria dos documentos do *corpus*, como artigos, preposições e pronomes. O cálculo dessa métrica é realizada pela seguinte fórmula:

$$tf_{x,y} = \frac{\# \text{ ocorrências do termo } x \text{ no documento } y}{\text{tamanho do documento } y} \quad (3.1)$$

$$idf_x = \log \left(\frac{\# \text{ total de documentos}}{\# \text{ documentos com o termo } x} \right) \quad (3.2)$$

$$tfidf_{x,y} = tf_{x,y} * idf_x \quad (3.3)$$

Com o resultado deste cálculo para todo o *corpus* em análise, é possível construir vetores esparsos BOW com a ponderação resultante. Essa representação vetorial de cada documento viabiliza operações matemáticas comuns a Álgebra Linear e o uso de métricas como similaridade por cosseno.

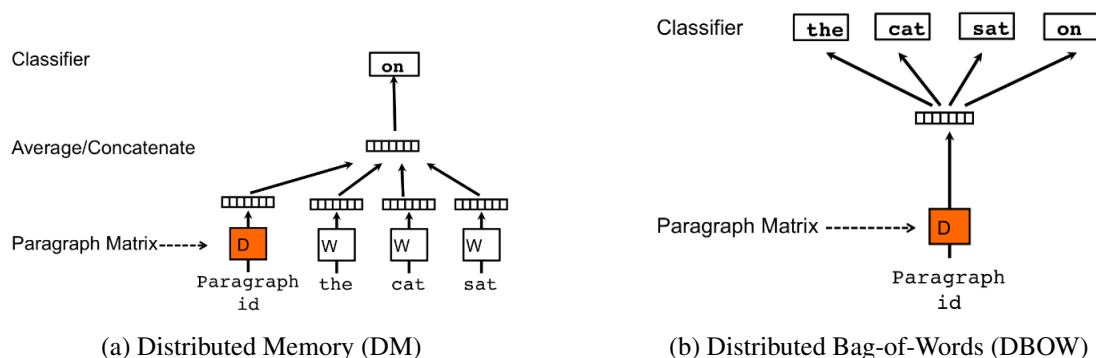
3.2 Doc2Vec

Conhecida também por *Paragraph Vectors*, a técnica Doc2Vec visa contornar dois dos principais problemas do modelo BOW: a representação vetorial esparsa e a perda de informações semânticas como ordenamento e contexto (LE; MIKOLOV, 2014). O nome

Paragraph Vectors visa exibir a capacidade do algoritmo lidar com textos de diferentes tamanhos, partindo de frases até documentos longos, onde cada texto de entrada deriva um único *Paragraph Vector*. A utilização de vetores densos de dimensionalidade fixa que capturam a semântica das palavras de um texto foi proposta em (MIKOLOV et al., 2013), trabalho que desenvolve a técnica de *word embeddings* (WE), também conhecida como Word2Vec. O Doc2Vec revela-se como uma extensão do Word2Vec, que visa levar a captura de semântica das *word embeddings* para o nível de documentos. Também foi reportado que o Doc2Vec tem melhor desempenho que uma simples ponderação de WE derivadas do modelo Word2Vec (MIKOLOV et al., 2013).

Similarmente ao modelo Word2Vec, os *Paragraph Vectors* são obtidos como resultado indireto de uma tarefa sintética de classificação, utilizada para ensinar um modelo de *machine learning* (ML) a prever a próxima palavra de uma janela de texto em análise. O modelo Doc2Vec prevê dois métodos de operação para os *Paragraph Vectors*, conforme a Figura 3.1:

Figura 3.1 – Métodos de operação do Doc2Vec



Fonte: Le e Mikolov (2014)

1. *Distributed Memory (DM)*: utiliza as palavras de contexto e a ordem das palavras, concatenando o *Paragraph Vector* e os respectivos *Word Vectors* da janela em análise para prever a próxima palavra.
2. *Distributed Bag-of-Words (DBOW)*: ignora as demais palavras da janela, mostrando uma palavra aleatória da janela em análise para realizar a predição das palavras de contexto. Por ser mais simples, esse método armazena menos dados e tem o treinamento mais rápido que o método anterior.

O método DM foi recorrentemente melhor que o método DBOW nos experimentos dos idealizadores da técnica. Entretanto, também foi relatado que o uso combinado dos dois métodos exibe, consistentemente, melhores resultados durante as análises (MI-

KOLOV et al., 2013).

3.3 SBERT

BERT (Bidirectional Encoder Representations from Transformers) (DEVLIN et al., 2018) é um modelo de linguagem com desempenho estado da arte em diversas tarefas de PLN que utiliza enormes datasets não-rotulados para pré-treinar um modelo de *deep learning* baseado em *transformers*. O BERT implementa a arquitetura e o mecanismo de atenção descrita em Vaswani et al. (2017). O pré-treinamento do modelo é realizado utilizando duas tarefas:

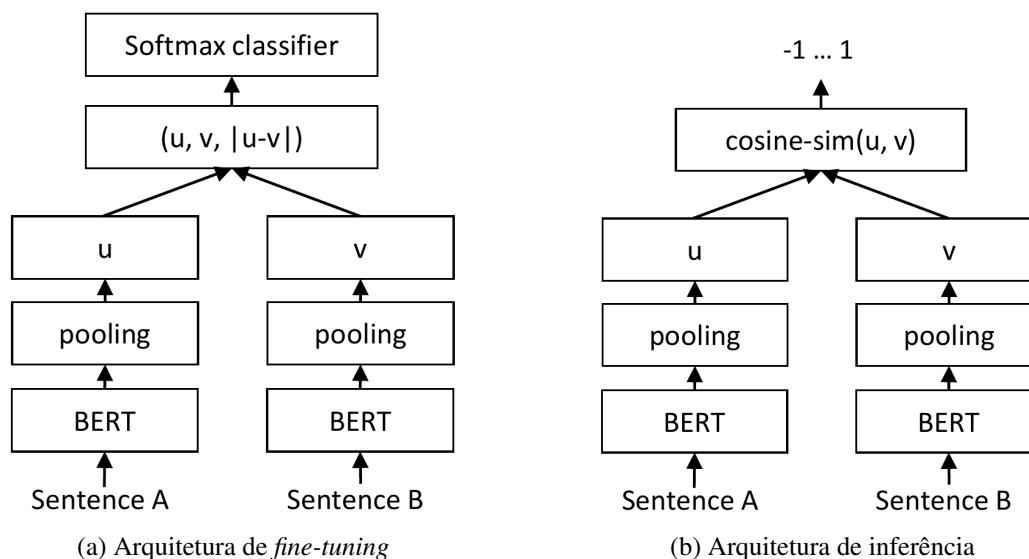
1. *Masked Language Model*: descreve a tarefa de selecionar uma porcentagem de *tokens* que deve ser mascarados, demarcando *tokens* a serem preditos pelo modelo. Isso permite obter um modelo de linguagem bidirecional.
2. *Next Sentence Prediction*: desenvolvido para capturar as relações entre duas sentenças, o processo de treino é realizado pela tarefa binária de prever se uma sentença é a continuação da sentença analisada ou não. Esse treinamento revela-se importante para tarefas de *Question Answering* e *Natural Language Inference*.

Após esse pré-treinamento, processo custoso computacionalmente, o modelo pré-treinado pode ser *fine-tuned* para outras tarefas de PLN com custo computacional reduzido.

A técnica de Sentence-BERT (SBERT) visa modificar o BERT para possibilitar extrair *embeddings* contextuais de sentenças que sejam significantes semanticamente, o que significa que sentenças com significados similares estarão próximas no espaço vetorial (REIMERS; GUREVYCH, 2019). Essa característica permite o uso de métricas como similaridade por cosseno e viabiliza o modelo para tarefas de *semantic textual similar*, *semantic search* e *paraphrase mining* em larga escala devido ao custo computacional de inferência reduzido. Entretanto, o número de *tokens* é limitado a 128 por sentença.

Para obter tais *embeddings*, as duas sentenças em análise são computadas pelo modelo BERT, onde é feito o *polling* da saída da rede para derivar as *embeddings*. Posteriormente, é empregada uma estrutura de rede siamesa para atualizar os pesos dos vetores e obter um *embedding* semanticamente significativa. Rede similar também é usada na fase de inferência para possibilitar a comparação das sentenças, conforme a Figura 3.2.

Figura 3.2 – Arquitetura do SBERT



Fonte: Reimers e Gurevych (2019)

3.4 Redução de Dimensionalidade

O objetivo principal das técnicas de redução de dimensionalidade é diminuir o número de *features* de um conjunto de dados, preservando o máximo possível de suas características fundamentais a um determinado problema. Uma aplicação possível dessa categoria de técnicas no presente trabalho seria reduzir a dimensionalidade dos vetores que representam os documentos analisados a ponto de ser exibidos em gráficos 2D, preservando as relações dos escritos examinados.

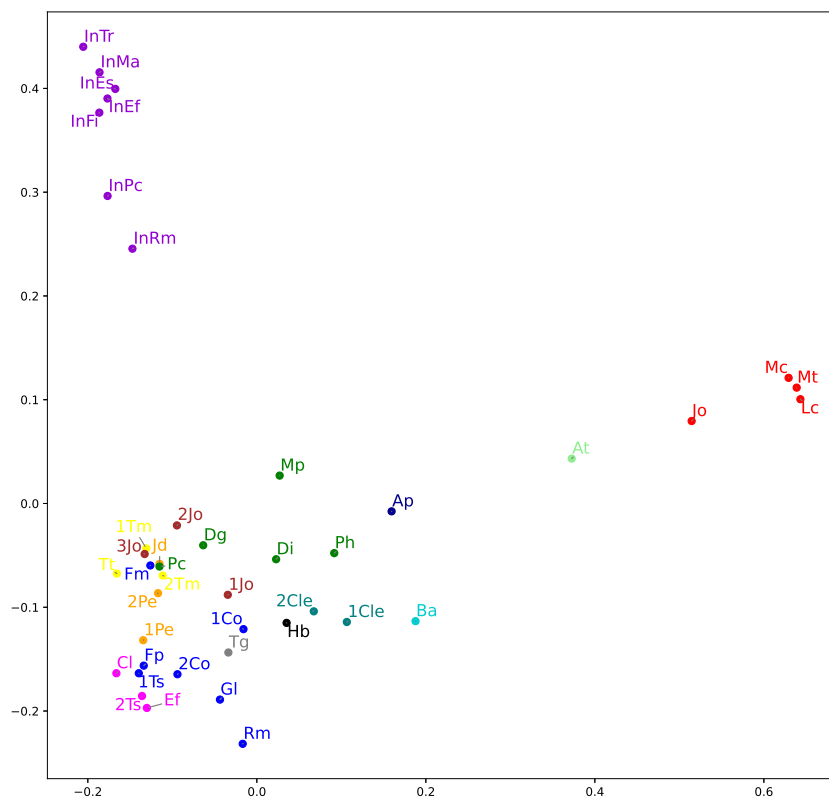
Para resolver essa categoria de problema, foram propostos diversos algoritmos, cada um com enfoques, objetivos e complexidades diferentes. Um dos algoritmos mais empregados nessa tarefa é o PCA (Principal Component Analysis), técnica de redução de dimensionalidade linear que visa preservar a máxima variância possível, utilizando o conceito de componentes principais (LEE; VERLEYSSEN, 2007, p. 24–37). Existem outras abordagens possíveis, como os métodos não-lineares, como o Isomap, que emprega métricas não-lineares para obter uma melhor aproximação no problema de preservação de distância. Existem métodos ainda que almejam realizar a preservação de topologia, como a técnica LLE (Locally Linear Embedding) (LEE; VERLEYSSEN, 2007, p. 234).

Foram realizadas análises experimentais utilizando tais algoritmos com os dados do presente trabalho, entretanto, foram notadas imprecisões relevantes, onde as relações entre os livros estudados foram impactadas significativamente, pois vetores com centenas ou milhares de dimensões tiveram que ser reduzidos para vetores bidimensionais. A Fi-

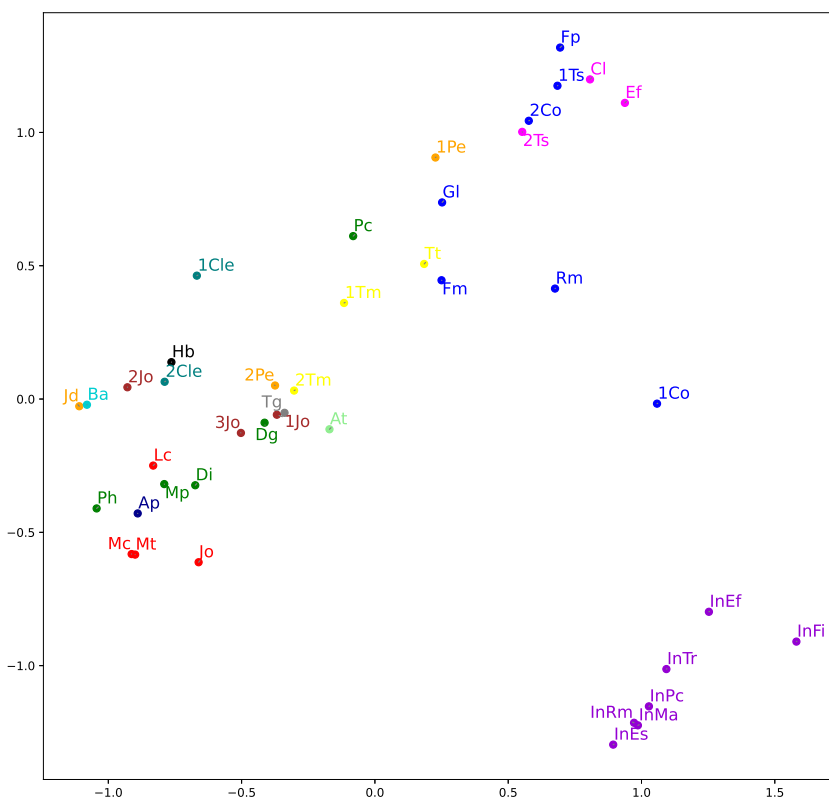
gura 3.3 exibe os resultados desses experimentos, onde foram aplicadas as técnicas PCA e Isomap. De modo geral, essas representações apresentam boas aproximações dos resultados, principalmente dos agrupamentos de textos. Entretanto, há erros visíveis como, por exemplo, os livros mais similares em relação a Atos dos Apóstolos são os evangelhos sinóticos, contudo, em nenhuma das análises isso se confirma. Outra distorção notável é exibida em 1 João, sendo que 1 João tem o Evangelho de João como o livro mais similar, entretanto, na Figura 3.3 é possível verificar haver um afastamento significativo entre esses dois livros nas análises.

Durante o desenvolvimento do presente trabalho, foi notada uma natureza assimétrica intrínseca nas comparações dos livros analisados, onde um livro X pode ter o livro Y como mais similar, porém, o livro Y pode ter o livro Z como mais similar. Esse tipo de relação pode ser exemplificada pelos resultados obtidos entre os livros de 1 e 2 Clemente na Seção 6.3.12. Portanto, com os dados levantados, foi constatada a inadequação do uso de técnicas de redução de dimensionalidade para obter dados passíveis de serem representados em forma de mapas 2D para serem analisados qualitativamente, segundo o objetivo do presente trabalho.

Figura 3.3 – Ilustração das distorções dos resultados da análise TF-IDF utilizando técnicas de redução de dimensionalidade



(a) PCA



(b) Isomap

Fonte: o autor

Legenda disponível na Tabela 6.3

4 TRABALHOS RELACIONADOS

Neste Capítulo serão examinados trabalhos com objetivos similares ao presente trabalho, além de realizar uma análise comparativa dos diferentes aspectos empregados.

O trabalho de Royal (2012) explora o uso de modelos Rasch, modelos oriundos da psicometria e baseados na teoria de resposta ao item para realizar análises de dados categóricos ou quantitativos. A importância dos modelos Rasch para as ciências humanas, bem como um aprofundamento teórico neste tópico pode ser encontrado em Bond et al. (2020). A análise dos textos do Novo Testamento recorre à técnica PCA sobre as correlações residuais do modelo Rasch, utilizando o texto baseado na concordância bíblica de Strong. A concordância bíblica de Strong¹ foi criada utilizando o texto da bíblia *King James*, tradução popular para o inglês que data do ano 1611. Entretanto, essa tradução não incorpora inúmeros avanços da crítica textual, como a descoberta do *Codex Sinaiticus* no ano 1844, além de trazer os vieses de seus tradutores, bem como prováveis imprecisões nas escolhas de tradução de palavras e remove as variações intrínsecas ao texto, o que poderia ser útil no trabalho apresentado. A frequência das palavras mais utilizadas nos textos foram modificadas em um esquema similar a uma escala logarítmica. Com o resultado obtido do PCA residual, foi elaborado um mapa 2D que exhibe as correlações obtidas.

O trabalho de Putniņš et al. (2006) estuda três métodos que podem ser usados para realizar atribuição de autoria, entretanto, somente a terceira técnica foi avaliada com textos bíblicos. A terceira análise utiliza Multiple Discriminant Analysis (MDA), uma técnica de redução de dimensionalidade, em conjunto com uma classificação baseada em distâncias de Mahalanobis. O trabalho utiliza marcadores de estilo como *features* analisadas, utilizando somente a frequência bruta de 70 palavras funcionais da língua analisada como marcadores de estilo em sua análise. Foi realizada uma análise com os textos em grego, entretanto, em uma fase de validação, obtendo nessa fase somente dados inconclusivos sobre a não-autoria de Paulo em relação a Epístola aos Hebreus. Uma segunda análise explora a possível similaridade do texto extra-bíblico da Epístola de Barnabé para inferir a autoria da Epístola aos Hebreus, contudo, esse trabalho não considerou uma maior gama de textos da Patrística para ter uma análise mais completa deste livro e utiliza os textos em inglês da bíblia *King James*.

¹Esta obra visava criar um mapeamento entre as palavras e uma numeração única. Essa elaboração também realiza o agrupamento de variações de palavras similares como, por exemplo, conjugações verbais de um mesmo verbo.

O trabalho de Barr (2003) inicia analisando os resultados de duas publicações que empregam análises multivariadas para obter dados estilométricos dos livros do Novo Testamento. Em sua análise, o autor questiona tais publicações e investiga a hipótese da adoção de diferentes estilos nas epístolas paulinas para se adaptar ao público receptor das obras. Foi utilizada uma abordagem escalométrica para exibir padrões que demonstram variações de estilo em obras de vários autores. Técnicas escalométricas foram mais profundamente examinadas em Barr (2002). Posteriormente, são encaminhadas análises qualitativas das epístolas paulinas, petrinhas e de Hebreus, utilizando nove *features*, compostas por palavras e categorias de palavras, onde são classificadas como formal ou informal dependendo da frequência de utilização de tais *features* nos textos.

Em Mealand (2012), é realizada uma análise estilométrica utilizando os textos do Novo Testamento em contraste aos textos da LXX e de diversos textos de autores gregos da Standard Hellenistic Greek (SHG). A análise considera uma seleção de 30 *features*, palavras funcionais e sufixos comuns, escolhidas visando capturar o estilo do autor, evitando influência pelo conteúdo e gênero do texto. A principal análise adotada foi a *Correspondence Analysis* (CA), técnica similar ao PCA, utilizando como base as frequência brutas das *features*.

Tabela 4.1 – Comparação dos trabalhos relacionados

Trabalho	Variiedade de Documentos	Versão do Texto	Utilização dos Textos	Metodologia	Dados Utilizados
Royal (2012)	NT	Strong	Texto Integral	PCA	Freq. Mod. + Rasch
Putnínš et al. (2006)	Seleção NT + Barnabé	King James	70 Palavras	MDA + Mahalanobis	Freq. Bruta
Barr (2003)	Seleção NT	Grego	9 Features	Análise Qualitativa	Freq. Relativa
Mealand (2012)	NT + LXX + SHG	Grego	30 Features	CA	Freq. Bruta

Com as informações levantadas, é possível elaborar a Tabela 4.1 que compara as principais características dos estudos analisados. É visto como ideal que uma análise adequada demonstre as seguintes características:

1. *Variiedade de Documentos*: é ideal que todos os textos que podem trazer informações úteis para o estudo sejam examinados.
2. *Versão do Texto*: o texto utilizado deve ser o mais próximo da obra original possível.
3. *Utilização dos Textos*: deve ser utilizado o máximo do texto possível, visando analisar o documento integralmente, minimizando o viés da análise.
4. *Metodologia*: empregar diferentes técnicas, tendo seus resultados comparados com *baselines* adequados.
5. *Dados Utilizados*: os dados utilizados devem ser adequados para elaborar a análise, processados e analisados sem distorções ou perda de informações significantes.

Conforme pode ser verificado na Tabela 4.1, os trabalhos examinados possuem poucas ou nenhuma das características ideais descritas acima: utilizam apenas os textos do Novo Testamento, ou ainda, somente um subconjunto destes textos; recorrem a traduções ou concordâncias; utilizam meramente uma fração do texto em suas análises; recorrem a técnicas de redução de dimensionalidade para possibilitar análises visuais, além de não apresentar uma diversidade maior de experimentos; bem como usam dados de frequência brutas, suscetíveis a distorções relativas ao tamanho das obras analisados.

O presente trabalho almeja conter o máximo dessas características ideais: realizar a análise com três técnicas diferentes com base em similaridade por cosseno, tendo seus resultados comparados aos dados obtidos por especialistas em crítica textual, utilizando os textos do Novo Testamento e dos escritos dos Pais Apostólicos em sua língua original em sua integridade, evitando ao máximo alterações e distorções nas análises.

5 METODOLOGIA

Neste Capítulo será detalhado a metodologia, pré-processamento e *pipelines* necessários para alcançar os objetivos do presente trabalho.

5.1 Proposta

Nossa proposta consiste em utilizar diferentes técnicas de PLN para quantificar a similaridade por cosseno dos textos-alvos, para assim, possibilitar a comparação com as relações relatadas por especialistas em crítica textual e comentaristas bíblicos. As relações expostas pelos especialistas, além de embasar os resultados obtidos, também tem papel de distinguir o motivo da similaridade apresentado nos textos, motivos estes que podem corresponder a dependência textual, similaridade semântica, autoria compartilhada, dentre outros. Para isso, foram escolhidas três técnicas que exploram diferentes aspectos dos textos: TF-IDF, Doc2Vec e SBERT. Pela comparação dos resultados obtidos pelas diferentes técnicas, é esperado obter resultados robustos que embasem as conclusões do presente trabalho.

Podemos resumir os procedimentos dos diferentes algoritmos em três passos: (i) *Encoding* para uma representação adequada dos textos para uso no algoritmo, (ii) geração da representação vetorial do documento, e (iii) quantificação da similaridade entre os documentos por meio da métrica de similaridade por cosseno. Um documento denota um livro/texto do Novo Testamento ou dos Pais Apostólicos. A seguir será detalhado o procedimento de pré-processamento dos documentos e o pipeline utilizado por cada algoritmo.

5.2 Pré-processamento

A partir dos textos brutos obtidos, foram realizados diversos processamentos para adequar os textos ao propósito deste trabalho. Primeiramente, foi realizada a adequação de diversos símbolos ambíguos nos textos. As edições do Novo Testamento e dos textos dos Pais Apostólicos utilizadas divergiam nas escolhas de pontuação e acentuação adotadas baseadas em Unicode. Posteriormente, foi realizada a remoção completa das notações de versículos, notações editoriais e pontuações. Ao final, os caracteres dos textos foram

convertidos para minúsculas, evitando que a capitalização crie entradas duplicadas nas representações posteriores. O resultado do pré-processamento pode ser exemplificado pela Figura 5.1.

Figura 5.1 – Comparação do resultado do pré-processamento de um versículo

Jn 14:6 λέγει αὐτῷ ὁ Ἰησοῦς· Ἐγώ εἰμι ἡ ὁδὸς καὶ ἡ ἀλήθεια καὶ ἡ ζωὴ· οὐδεὶς ἔρχεται πρὸς τὸν πατέρα εἰ μὴ δι' ἐμοῦ.
λέγει αὐτῷ ὁ ἰησοῦς ἐγώ εἰμι ἡ ὁδὸς καὶ ἡ ἀλήθεια καὶ ἡ ζωὴ οὐδεὶς ἔρχεται πρὸς τὸν πατέρα εἰ μὴ δι' ἐμοῦ

Fonte: o autor

Nos textos dos Pais Apostólicos, foi necessária a remoção adicional de capítulos e versículos introdutórios, alheios aos textos originais, bem como de porções das obras Epístola de Policarpo e Pastor de Hermas escritas em latim, evitando assim distorções nos resultados pelo uso de vocabulário estranho ao grego empregado nas demais obras do *corpus*.

Após a realização do pré-processamento descrito, ainda foi conduzida uma validação visual dos textos, onde foi empregado um *script* para extrair os caracteres contidos nos textos, para verificar se os caracteres restantes correspondem ao resultado esperado.

5.3 Pipeline TF-IDF

1. *Tokenização*: A tokenização foi realizada extraindo *tokens* separados por espaço, pois todos os caracteres passíveis de interferir nesse processo foram removidos na fase de pré-processamento, como caracteres de pontuação e notação editorial.
2. *Geração de Dicionário*: Nessa etapa, os *tokens* são codificados em uma representação eficiente que mapeia os *tokens* para inteiros.
3. *Geração de BOWs*: Os BOWs são construídos relacionando a contagem de ocorrências dos termos em um documento com as entradas do dicionário.
4. *Ponderação TF-IDF*: É gerada a representação vetorial utilizando os pesos dados pela fórmula do TF-IDF, conforme detalhado na Seção 3.1.
5. *Resultado*: Com os vetores TF-IDF prontos, é possível criar uma matriz de similaridade por cosseno dos documentos.

5.4 Pipeline Doc2Vec

1. *Tokenização*: Foram extraídas listas de *tokens* separados por espaço, oriundas dos textos pré-processados, assim como no modelo TF-IDF.
2. *Documentação*: As listas de *tokens* derivadas dos textos são agrupadas em documentos identificados para serem processados.
3. *Geração de Vocabulário*: Os documentos são utilizados para verificar o vocabulário empregado nos textos para possibilitar o uso do algoritmo Doc2Vec.
4. *Treinamento*: Os documentos são transformados em representações vetoriais que serão utilizadas pelo algoritmo. O processo de treinamento é realizado em um esquema *leave-one-out cross-validation*.
5. *Inferência*: Com os vetores resultantes do treinamento do modelo, é realizada a inferência de uma representação vetorial do documento em análise.
6. *Resultado*: Utilizando o vetor inferido, calcula-se a similaridade do cosseno entre este vetor e os vetores dos demais documentos.

5.5 Pipeline SBERT

1. *Encoding*: Nesta etapa, os versículos de um texto são codificados em *tokens* do modelo pré-treinado. Em seguida, os *tokens* são processados, resultando em uma *embedding* contextual mapeada em um espaço vetorial para cada versículo.
2. *Agregação*: Para obter a representação de um livro, é realizada a agregação das *embeddings* de cada versículo do livro através de uma operação de média, obtendo-se assim, uma única *embedding* que representa o documento em sua integridade, evitando o limite de *tokens* por sentença, conforme relatado na Seção 3.3.
3. *Resultado*: Com a *embedding* resultante, é possível calcular a similaridade por cosseno em relação às *embeddings* dos outros documentos, resultando em uma matriz de similaridade.

5.6 Sumário

Com os procedimentos e algoritmos descritos, agora é possível viabilizar a implementação e a análise experimental comparativa dos textos desejados.

6 ANÁLISE EXPERIMENTAL

Neste Capítulo são descritos todos os recursos e procedimentos necessários para implementar a metodologia descrita no Capítulo 5.

6.1 Configuração Experimental

Nesta Seção são descritos os *corpora*, ferramentas e parametrizações utilizadas para materializar a análise experimental. Os códigos utilizados nos experimentos estão disponíveis em: <<https://github.com/cassianobruckhoff/new-testament-and-apostolic-fathers-analysis>>.

6.1.1 Corpora

Foram utilizados dois *corpora* para realizar todas as análises do presente trabalho:

1. *SBL Greek New Testament*¹: editada por Michael W. Holmes, autor de Holmes (2006), a SBLGNT tem como objetivo disponibilizar o texto do Novo Testamento e aparato crítico em diversos formatos digitais para uso livre em trabalhos acadêmicos e educacionais, recorrendo a diversas outras edições do Novo Testamento para estabelecer seu texto.
2. *Open Apostolic Fathers*²: editada por Tauber e Macdonald (2019), Esta compilação contém todas as principais obras dos Pais Apostólicos, visando ser uma versão livre, verificada e corrigida da obra original de Lake (1914).

6.1.2 Ferramentas

As análises foram implementadas na linguagem *Python*, um dos ambientes mais utilizados no desenvolvimento de projetos de PLN e ML. Os gráficos gerados utilizam a biblioteca *Matplotlib*³ (HUNTER, 2007), capaz de elaborar diversos tipos de visualizações. Na Figura 3.3 foi utilizada a biblioteca *adjustText*⁴ (FLYAMER et al., 2020) para

¹<<https://sblgnt.com/about/introduction>>

²<<https://github.com/jtauber/apostolic-fathers>>

³<<https://matplotlib.org>>

⁴<<https://github.com/Phlya/adjustText>>

realizar o posicionamento automático dos *labels*.

A biblioteca *Gensim*⁵ serviu como base para as análises TF-IDF e Doc2Vec. A biblioteca *Gensim* é uma suíte que conta com diversos modelos de PLN implementados eficientemente, como Word2Vec (MIKOLOV et al., 2013) e FastText (BOJANOWSKI et al., 2017), implementações estas que são amplamente utilizadas na indústria⁶.

A implementação do modelo SBERT recorre à biblioteca *sentence_transformers*⁷, a implementação dos idealizadores desta técnica (REIMERS; GUREVYCH, 2019). Baseada nas bibliotecas *PyTorch* e *Transformers*, a biblioteca *sentence_transformers* possibilita a utilização de inúmeros modelos pré-treinados estado da arte hospedados pela HuggingFace⁸.

6.1.3 Parametrização

Foram ajustados cinco dos principais parâmetros do Doc2Vec: *vector_size*, *dm*, *window*, *epochs* e *min_count*. Somente os parâmetros *dm*, e *epochs*, tiveram efeitos significativos nos resultados. O parâmetro *dm* seleciona o método de operação adotado, conforme discutido na Seção 3.2. O presente trabalho irá utilizar os resultados dos dois métodos, pois fornecem dados complementares que serão úteis para efetuar a validação dos resultados.

Tabela 6.1 – Parâmetros do Doc2Vec

<i>vector_size</i>	<i>window</i>	<i>epochs</i>	<i>min_count</i>
100 (d)	5 (d)	100	1 (d)

Os parâmetros denotados com (d) são valores *default* da implementação utilizada.

O modelo SBERT utiliza o modelo pré-treinado *lighteternal/stsb-xml-r-greek-transfer*⁹ da *Hellenic Army Academy* e *Technical University of Crete*, criado utilizando a técnica de *knowledge distillation* descrita em Reimers e Gurevych (2020).

⁵<<https://radimrehurek.com/gensim>>

⁶<<https://github.com/RaRe-Technologies/gensim#adopters>>

⁷<<https://www.sbert.net>>

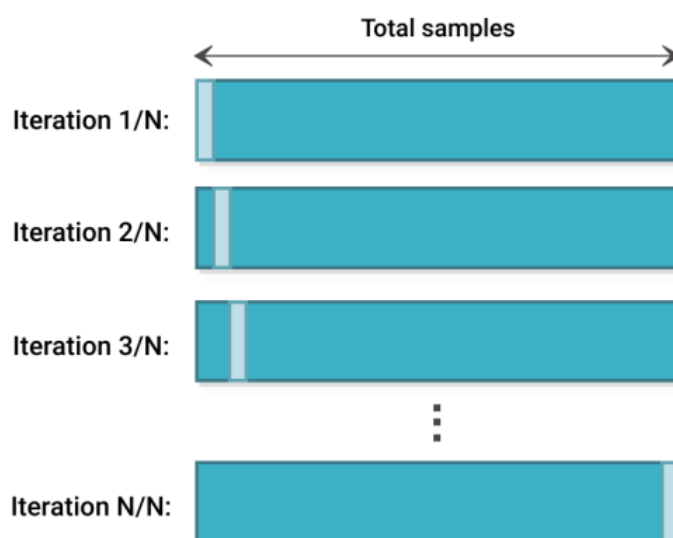
⁸<<https://huggingface.co>>

⁹<<https://huggingface.co/lighteternal/stsb-xml-r-greek-transfer>>

6.2 Validação

A validação foi realizada através de uma avaliação experimental, tomando os textos em blocos de 200 palavras em um esquema similar ao *leave-one-out cross-validation*, conforme a Figura 6.1, onde o bloco é comparado a todo o restante do livro em análise, bem como todos outros textos utilizado no presente trabalho. A análise SBERT não foi avaliada nesta validação devido à limitação de *tokens* inerente ao algoritmo e tempo de computação necessário.

Figura 6.1 – Ilustração do processo de *leave-one-out cross-validation*



Fonte: Neptune.ai <<https://bit.ly/3GwFdOF>>

O resultado da avaliação pode ser observado na Tabela 6.2, sendo possível verificar que, na grande maioria dos resultados, o resultado mais frequente trata-se do respectivo livro em análise ou textos, provavelmente, de mesma autoria. Isso ajuda a revelar a robustez dos métodos em um cenário que utiliza um número limitado de palavras. É esperado que com a utilização dos textos completos, além do uso comparado dos resultados das diferentes análises acarrete uma confiabilidade maior nos resultados.

Tabela 6.2 – Avaliação comparativa das análises utilizando amostras em blocos de palavras

Livro	Bl.	TF-IDF	Doc2Vec (DM)	Doc2Vec (DBOW)
Mt	91	Mt: 50, Mc: 21, Lc: 18, Di: 2	Mt: 45, Mc: 22, Lc: 16, Jd: 2	Mt: 41, Mc: 30, Lc: 16, At: 2
Mc	56	Mc: 31, Mt: 13, Lc: 10, Jo: 2	Mc: 27, Mt: 22, Lc: 3, Jo: 2	Mc: 23, Mt: 20, Lc: 7, At: 3
Lc	97	Lc: 67, Mt: 18, Mc: 11, Di: 1	Lc: 41, Mt: 28, Mc: 17, At: 4	Lc: 35, Mc: 23, Mt: 20, At: 15
Jo	77	Jo: 75, Mc: 2	Jo: 71, Mc: 2, 2Jo: 2, Jd: 1	Jo: 74, Mc: 3
At	92	At: 85, Lc: 5, Mt: 1, Ba: 1	At: 83, Mp: 2, Dg: 1, 1Cle: 1	At: 89, 1Cle: 1, 2Pe: 1, Ef: 1
Rm	35	Rm: 22, Ba: 4, 1Co: 3, Gl: 2	Rm: 15, Fm: 3, Gl: 3, 1Pe: 2	Rm: 17, Gl: 5, 1Co: 3, 1Pe: 2
1Co	34	1Co: 26, Gl: 2, Rm: 2, Fp: 1	1Co: 23, 1Ts: 3, Fp: 3, Fm: 1	1Co: 29, 2Co: 2, 1Tm: 1, Rm: 1
2Co	22	2Co: 16, 1Co: 2, 1Ts: 1, Ba: 1	2Co: 11, Fp: 3, 1Ts: 3, 1Pe: 1	2Co: 16, 1Ts: 2, 1Co: 2, Gl: 1
Gl	11	Rm: 6, At: 2, Gl: 2, 1Co: 1	Gl: 6, 1Co: 2, Rm: 2, Fp: 1	Rm: 7, At: 2, InRm: 1, Gl: 1
Ef	12	Cl: 7, Ef: 5	Cl: 8, Gl: 2, Ef: 1, 1Pe: 1	Cl: 7, Ef: 2, Rm: 1, Gl: 1
Fp	8	2Co: 3, Fp: 2, 1Co: 1, Rm: 1	Fp: 4, Fm: 1, 1Ts: 1, 2Co: 1	Fp: 5, Rm: 1, Gl: 1, Cl: 1
Cl	7	Ef: 6, 1Co: 1	Ef: 4, 2Ts: 1, Cl: 1, 1Co: 1	Ef: 6, Dg: 1
1Ts	7	1Ts: 6, 2Ts: 1	1Ts: 4, 2Ts: 3	1Ts: 6, 2Ts: 1
2Ts	4	1Ts: 4	1Ts: 3, 2Ts: 1	1Ts: 3, InEs: 1
1Tm	7	1Tm: 5, Tt: 2	1Tm: 5, 2Tm: 1, 3Jo: 1	1Tm: 4, Tt: 2, InEs: 1
2Tm	6	1Tm: 2, Tt: 2, 2Tm: 2	1Tm: 4, Tt: 1, 3Jo: 1	1Tm: 2, 2Tm: 2, Ph: 2
Tt	3	Tt: 2, 1Tm: 1	1Tm: 2, Tt: 1	Tt: 2, 1Tm: 1
Fm	1	Fm: 1	Fm: 1	Ef: 1
Hb	24	Hb: 18, 1Cle: 3, Lc: 1, Ba: 1	Hb: 16, 1Pe: 3, Fp: 2, Ba: 1	Hb: 19, Ba: 1, 1Tm: 1, At: 1
Tg	8	1Co: 3, Rm: 3, Gl: 1, Mt: 1	Tg: 3, 1Cle: 2, 1Co: 1, 1Tm: 1	2Pe: 2, InFi: 1, Gl: 1, 1Cle: 1
1Pe	8	Pc: 2, Rm: 1, Ba: 1, 1Co: 1	1Pe: 7, Tt: 1	1Pe: 4, 2Ts: 1, Ba: 1, Pc: 1
2Pe	5	Jd: 2, 1Tm: 1, 2Pe: 1, 1Cle: 1	2Pe: 3, Jd: 2	Jd: 2, Tt: 1, 2Pe: 1, Dg: 1
1Jo	10	1Jo: 9, Jo: 1	1Jo: 9, Jo: 1	1Jo: 10
2Jo	1	1Jo: 1	3Jo: 1	1Jo: 1
3Jo	1	2Jo: 1	InPc: 1	2Jo: 1
Jd	2	2Pe: 2	Jd: 2	Jd: 1, 2Pe: 1
Ap	49	Ap: 49	Ap: 49	Ap: 49
1Cle	49	1Cle: 36, Ba: 5, Lc: 4, At: 1	1Cle: 27, Jd: 4, Pc: 4, Rm: 2	1Cle: 40, Jd: 1, Di: 1, At: 1
2Cle	15	2Cle: 12, Lc: 1, Jo: 1, Mt: 1	2Cle: 12, Dg: 1, Tg: 1, Jd: 1	2Cle: 12, 1Cle: 2, Mt: 1
InEf	8	InTr: 2, InMa: 1, InEs: 1, 1Ts: 1	InEs: 5, InMa: 1, InRm: 1, 2Tm: 1	InEs: 3, InMa: 1, Ef: 1, InRm: 1
InMa	5	InEf: 1, InFi: 1, InMa: 1, Jo: 1	InTr: 3, InFi: 1, InEf: 1	InTr: 3, InFi: 1, InEf: 1
InTr	4	InMa: 2, InRm: 1, InEs: 1	InMa: 3, InFi: 1	InMa: 3, InEs: 1
InRm	5	InRm: 3, InTr: 2	InRm: 4, InMa: 1	InRm: 3, InMa: 1, InEs: 1
InFi	5	InFi: 2, InEf: 1, InMa: 1, InEs: 1	InEs: 3, InTr: 1, InMa: 1	InEs: 2, InTr: 1, InFi: 1, InMa: 1
InEs	5	InFi: 2, InTr: 1, Mc: 1, InEf: 1	InTr: 3, InEf: 2	InTr: 2, InEf: 2, InFi: 1
InPc	3	InMa: 2, InEf: 1	InPc: 2, InEf: 1	Fm: 1, Tt: 1, InFi: 1
Pc	5	Pc: 3, Mp: 1, 1Cle: 1	1Pe: 2, InEf: 1, 1Tm: 1, Pc: 1	Pc: 2, Fp: 1, 1Pe: 1, Ba: 1
Di	10	Di: 8, Ba: 2	Di: 6, Ba: 2, Tt: 1, InPc: 1	Di: 6, Ba: 2, InPc: 1, 1Jo: 1
Ba	33	Ba: 28, Di: 4, Lc: 1	Ba: 28, Di: 3, 2Pe: 2	Ba: 27, Di: 3, Tg: 1, Mt: 1
Ph	135	Ph: 133, Ba: 1, 1Cle: 1	Ph: 134, Jd: 1	Ph: 130, InTr: 2, 1Pe: 1, At: 1
Mp	12	Mp: 9, At: 2, Lc: 1	Mp: 10, Dg: 1, 2Pe: 1	Mp: 7, At: 2, Pc: 1, Tt: 1
Dg	13	Dg: 9, Rm: 2, Mt: 1, Ba: 1	Dg: 9, 2Pe: 2, Mp: 1, 1Tm: 1	Dg: 10, Rm: 1, Cl: 1, 1Pe: 1

6.3 Resultados

Nesta Seção serão expostos os resultados obtidos e realizada a discussão comparativa com as relações documentadas nas Seções 2.1 e 2.2.

A avaliação dos resultados foi realizada qualitativamente, pela observação dos resultados de similaridade por cosseno geradas pelos três modelos utilizados. A partir dos resultados tabelados, foram realizadas comparações entre os resultados obtidos e as relações encontradas por especialistas em crítica textual e comentaristas bíblicos.

Durante a elaboração dos experimentos, os modelos utilizados exibiram níveis de resultados diferentes, bem como afinidade com certas categorias de tarefas. Devido a isso, os resultados das análises foram utilizados de forma a maximizar a qualidade geral do trabalho, balizando-se pelas relações bem estabelecidas documentadas pelos críticos textuais e comentaristas bíblicos.

O modelo Doc2Vec obteve o melhor desempenho global, exibindo ótimos resultados em tarefas que dependem tanto de elementos sintáticos quanto semânticos. O modelo TF-IDF exibiu bom desempenho em tarefas que dependiam, predominantemente, de componentes sintáticos, sendo utilizado extensivamente como complemento a análise Doc2Vec. O SBERT exibiu uma grande variabilidade nos resultados, sendo aproveitado, principalmente, como complemento em análises com forte similaridade semântica.

Para contribuir com a objetividade do texto, foi escolhido exibir somente os resultados do modelo Doc2Vec operando com o método DM. Também foram suprimidos os gráficos de algumas cartas de Paulo e de Inácio de Antioquia nos casos onde as informações foram julgadas menos importantes, para contribuir com a brevidade do trabalho. Os resultados foram exibidos em forma de gráficos ordenados pela similaridade por cosseno apresentada relativamente ao texto analisado, denotado no título do gráfico. As cores utilizadas agrupam os textos por categoria, onde os grupos são especificados pela legenda contida nos gráficos. As similaridades por cosseno negativas foram zeradas para evitar distorções nos gráficos, entretanto, o ordenamento foi mantido intacto. A lista de abreviações adotadas se encontra na Tabela 6.3. As abreviações dos livros com um asterisco (*) nos gráficos denotam que foi removida a passagem em análise do respectivo livro.

Tabela 6.3 – Lista de abreviações adotada

Abrev.	Livro	Abrev.	Livro	Abrev.	Livro
Mt	Evangelho de Mateus	1Tm	1 Timóteo	2Cle	2 Clemente
Mc	Evangelho de Marcos	2Tm	2 Timóteo	InEf	Inácio aos Efésios
Lc	Evangelho de Lucas	Tt	Tito	InMa	Inácio aos Magnésios
Jo	Evangelho de João	Fm	Filemom	InTr	Inácio aos Trálios
At	Atos dos Apóstolos	Hb	Hebreus	InRm	Inácio aos Romanos
Rm	Romanos	Tg	Tiago	InFi	Inácio aos Filadélfos
1Co	1 Coríntios	1Pe	1 Pedro	InEs	Inácio aos Esmirniotas
2Co	2 Coríntios	2Pe	2 Pedro	InPc	Inácio a Policarpo
Gl	Gálatas	1Jo	1 João	Pc	Policarpo aos Filipenses
Ef	Efésios	2Jo	2 João	Di	Didaquê
Fp	Filipenses	3Jo	3 João	Ba	Barnabé
Cl	Colossenses	Jd	Judas	Ph	Pastor de Hermas
1Ts	1 Tessalonicenses	Ap	Apocalipse	Mp	Martírio de Policarpo
2Ts	2 Tessalonicenses	1Cle	1 Clemente	Dg	Diogneto

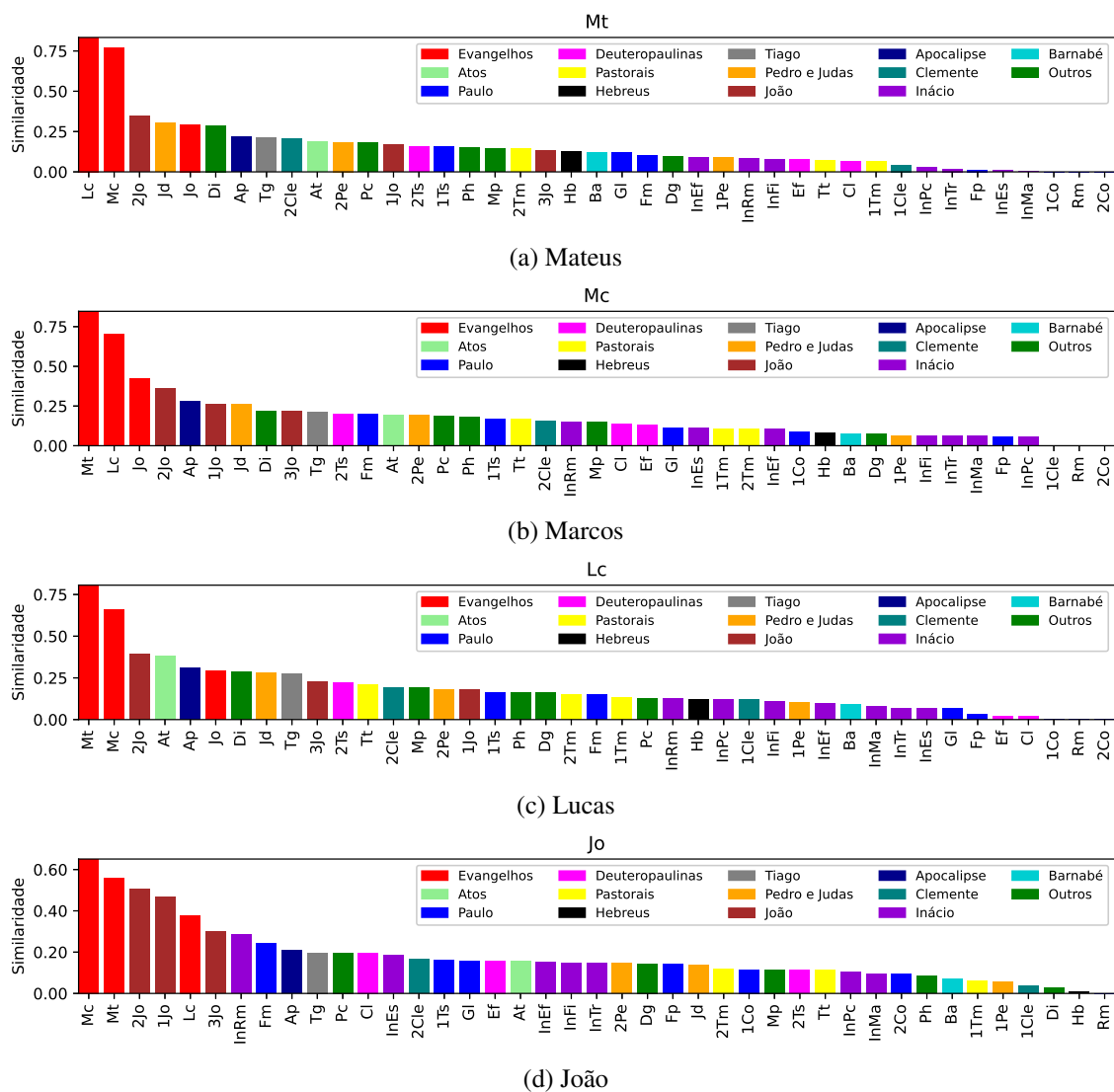
6.3.1 Evangelhos

O primeiro dado que chama atenção é a grande coesão exibida na maioria das análises dos evangelhos. Os quatro evangelhos exibiram os maiores valores TF-IDF dentre todos os textos analisados, mesmo se comparado com textos considerados de mesma autoria, como as cartas autênticas de Paulo ou de Inácio de Antioquia.

A coesão dos evangelhos sinóticos também foi demonstrada. Todas as análises de algum desses evangelhos sempre tiveram os demais evangelhos sinóticos como textos mais similares. Já o Evangelho de João foi, consistentemente, o evangelho com o menor nível de similaridade com os demais, inclusive, exibindo valores de similaridade inferiores a livros não relacionados aos evangelistas em algumas das análises do Doc2Vec e SBERT, como na Figura 6.2d. Essa informação pode ajudar a demonstrar a autoria do apóstolo João e a datação tardia do evangelho, alicerçando-se nas diferenças evidenciadas nos resultados e nas informações relatadas na Seção 2.1.1.

Em relação ao problema sinótico, o Evangelho de Mateus traz uma maior afinidade com o Evangelho de Lucas e vice-versa, especialmente na análise com Doc2Vec usando o método DM, conforme as Figuras 6.2a e 6.2c. Isso pode evidenciar uma maior similaridade sintática e semânticas entre seus textos. O Evangelho de Marcos apresenta uma maior similaridade com Mateus, vide Figura 6.2b, entretanto, em menor grau que a relação Mateus-Lucas. Utilizando o pressuposto da primazia de Marcos, conforme a Seção 2.1.1, e que o texto menos similar seja o mais afastado temporalmente, podemos levantar a suposição que o Evangelho de Marcos serviu como base para Mateus e que

Figura 6.2 – Comparações de similaridade dos evangelhos pela análise Doc2Vec



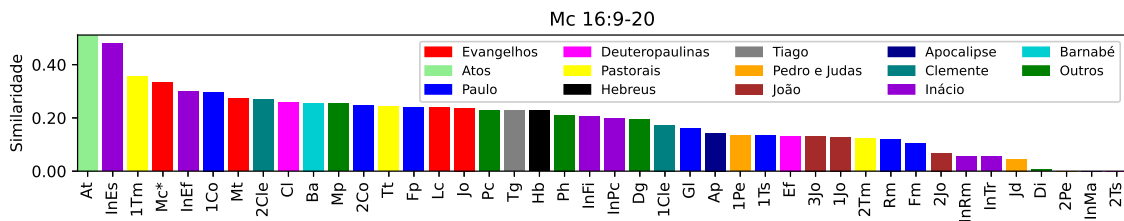
Fonte: o autor

Lucas utilizou pelo menos o Evangelho de Mateus como fonte de sua investigação, isso poderia explicar o ínfimo conteúdo exclusivamente compartilhado entre Marcos e Lucas. Entretanto, isso se trata de uma inferência e um estudo mais aprofundado nesse tema seria necessário para obter dados mais conclusivos.

6.3.2 Final Longo de Marcos

Esse texto exhibe irregularidade nos resultados, demonstrando relativa afinidade com os evangelhos e Atos nas análises TF-IDF e SBERT, entretanto, nas análises Doc2Vec, vide Figura 6.3, houve uma mudança de padrão, onde Atos foi seguido por outros textos

Figura 6.3 – Comparações de similaridade do Final Longo de Marcos pela análise Doc2Vec



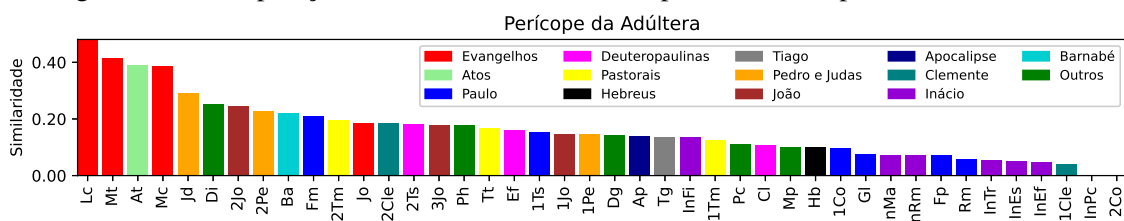
Fonte: o autor

diversos. Na análise TF-IDF, o texto mais similar ao texto em questão foi o Evangelho de Lucas, enquanto, na análise SBERT, o Evangelho de Marcos foi o mais similar. Contudo, nessas últimas duas análises, a margem de similaridade para outros textos foi baixíssima.

Com a exceção da análise SBERT, o Evangelho de Marcos manteve certo afastamento do texto aqui analisado. Conseqüentemente, dificilmente é possível relacionar o Final Longo de Marcos com o Evangelho de Marcos, o que pode confirmar a provável adição desse texto no final do evangelho, conforme discutido na Seção 2.1.1.

6.3.3 Perícopes da Adúltera

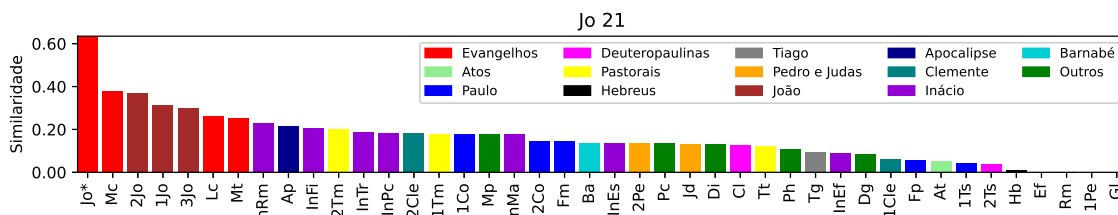
Figura 6.4 – Comparações de similaridade da Perícopes da Adúltera pela análise Doc2Vec



Fonte: o autor

A perícopes apresentou similaridade com o Evangelho de Lucas em todas as análises, com a exceção da análise SBERT, vide Figura 6.4. Além disso, em todas as análises, o Evangelho de João foi o evangelho com a menor similaridade com essa passagem. Isso parece confirmar a hipótese de uma possível realocação indevida do texto e mostra a importância das múltiplas atestações do texto bíblico e suas diversas famílias de manuscritos, mesmo que datados tardiamente, como no caso do grupo de manuscritos Ferrar.

Figura 6.5 – Comparações de similaridade do Capítulo 21 do Evangelho de João pela análise Doc2Vec



Fonte: o autor

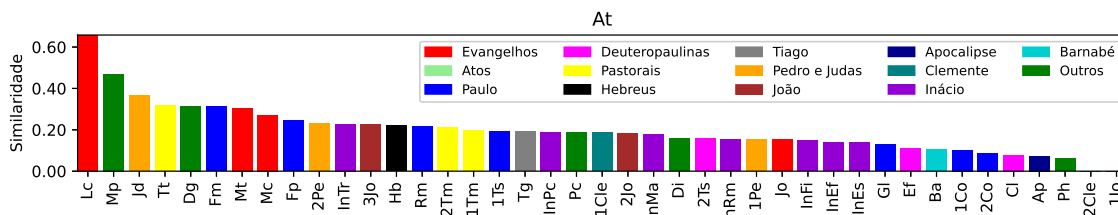
6.3.4 Capítulo 21 do Evangelho de João

O capítulo em questão apresenta uma grande similaridade com o Evangelho de João em todas as análises, exceto na análise SBERT, vide Figura 6.5. Entretanto, uma análise semântica desse texto, provavelmente, não seja a mais adequada, partindo do fato que esse capítulo narra uma história única no texto bíblico e não mostra relações óbvias com outros textos.

Com base nos dados e nas evidências textuais levantadas na Seção 2.1.1, é possível inferir a autoria compartilhada desse capítulo com o restante do Evangelho de João.

6.3.5 Atos dos Apóstolos

Figura 6.6 – Comparações de similaridade de Atos dos Apóstolos pela análise Doc2Vec

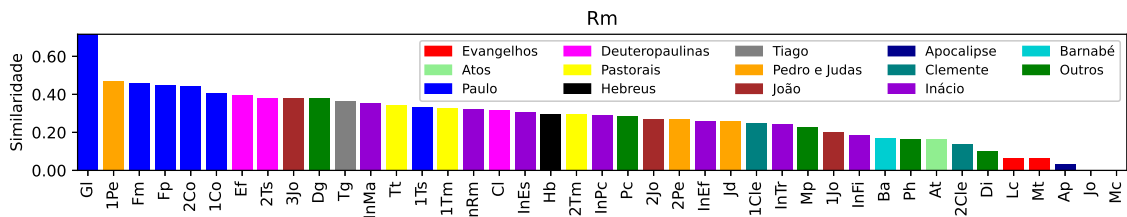


Fonte: o autor

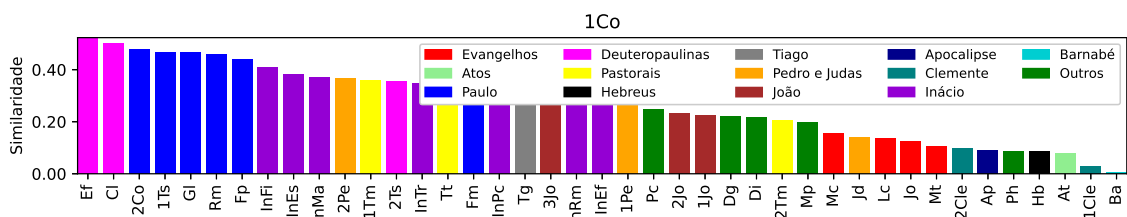
Os resultados exemplificados pela Figura 6.6 confirmam a relação de Atos dos Apóstolos com a redação do Evangelho de Lucas em todas as análises. A relação com os outros evangelhos se revela principalmente nas análises TF-IDF e SBERT, sendo bem menos evidente nas análises Doc2Vec. Esse conjunto de dados se mostra relevante para afirmar a autoria em comum com o Evangelho de Lucas e sua relação com os outros evangelhos.

6.3.6 Epístolas Paulinas

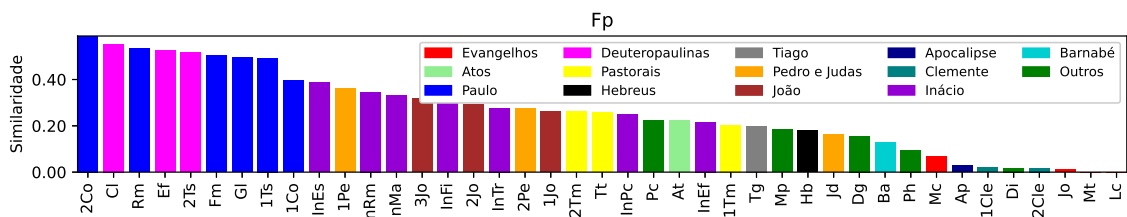
Figura 6.7 – Comparações de similaridade das epístolas autênticas de Paulo pela análise Doc2Vec



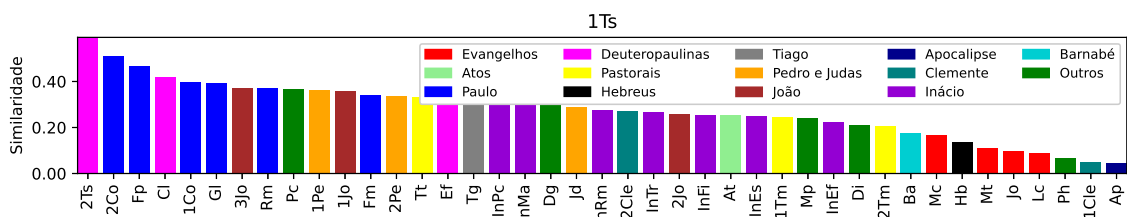
(a) Romanos



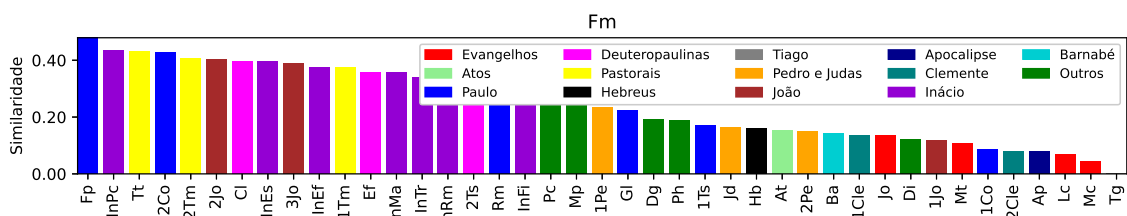
(b) 1 Coríntios



(c) Filipenses



(d) 1 Tessalonicenses



(e) Filemom

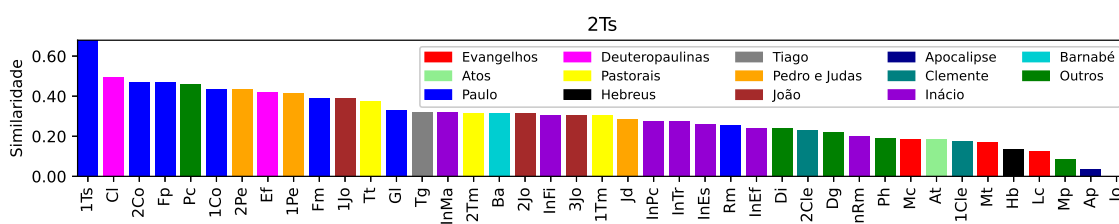
Fonte: o autor

As epístolas autênticas apresentam, geralmente, uma similaridade relevante com uma seleção de cartas, porém não com a coleção completa. Por exemplo, na Figura 6.7a, podemos verificar a grande similaridade de Romanos com Gálatas. Na Figura 6.7b, a carta

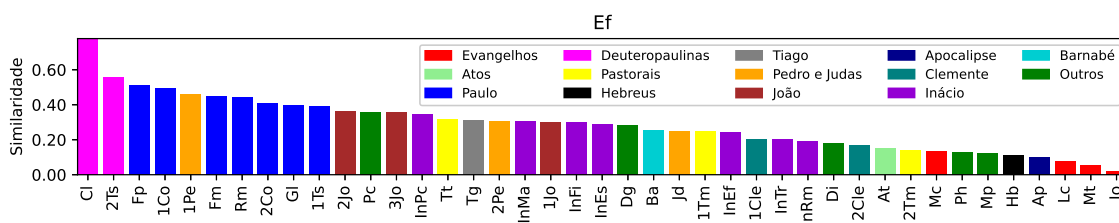
de 1 Coríntios apresenta similaridade com Efésios, Colossenses e 2 Coríntios, entretanto, com maior variabilidade dos resultados nas outras análises. A Figura 6.7c exibe a forte coesão de Filipenses com todas as demais epístolas autênticas de Paulo e deuteropaulinas. A carta de 1 Tessalonicenses exibe similaridade com diversas cartas de Paulo, conforme a Figura 6.7d.

A exceção a essas análises se trata da Epístola a Filemon, que apresenta nível de similaridade significativamente menor na maioria das análises, se mostrando maior afinidade com as epístolas pastorais, apesar de sua análise individual apontar a autoria paulina, conforme exemplificado pela Figura 6.7e. A maior variabilidade dos resultados de 1 Coríntios talvez pode ser explicada pela coautoria de Sóstenes¹⁰, sendo sua única contribuição relatada no texto bíblico.

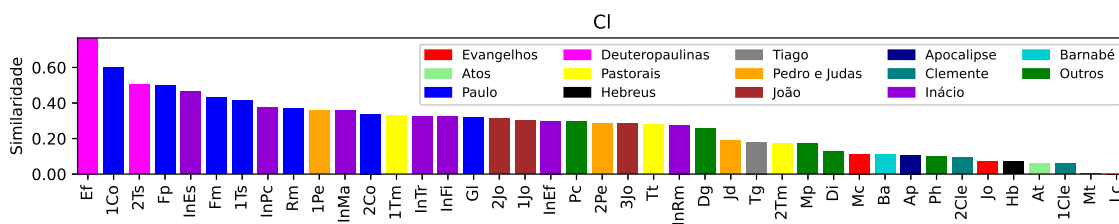
Figura 6.8 – Comparações de similaridade das epístolas deuteropaulinas pela análise Doc2Vec



(a) 2 Tessalonicenses



(b) Efésios



(c) Colossenses

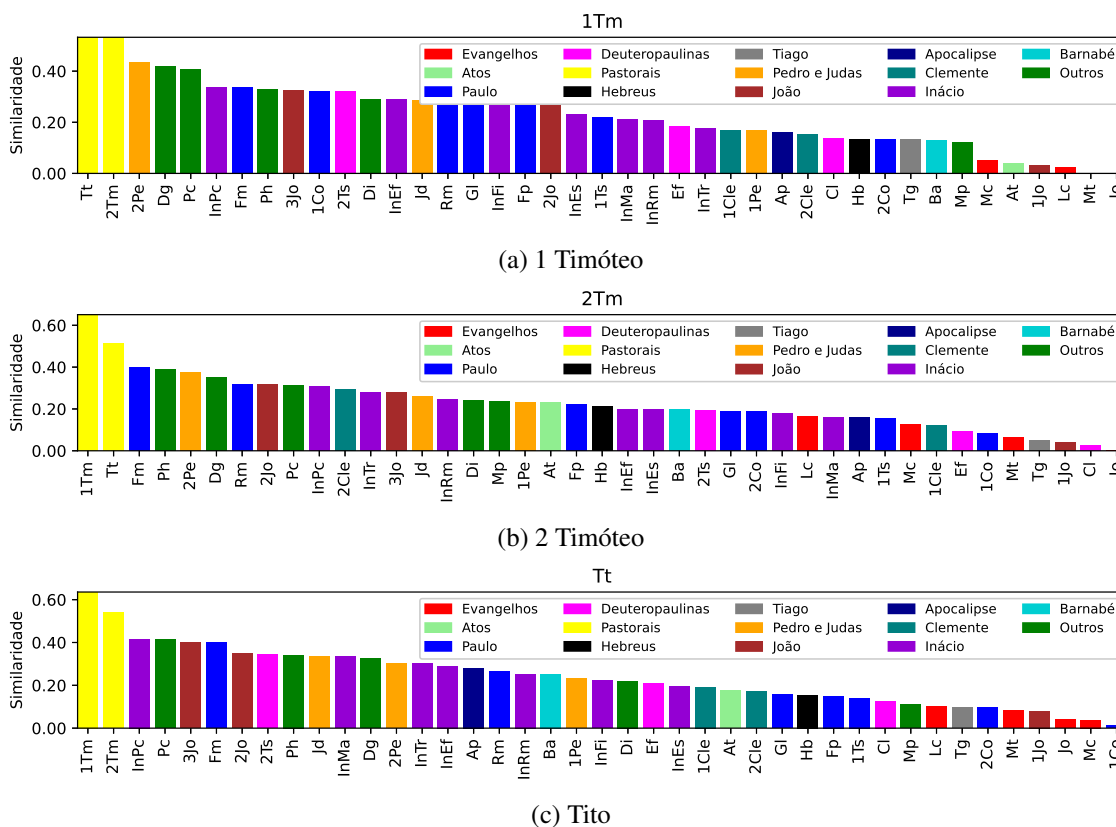
Fonte: o autor

No grupo de epístolas deuteropaulinas, tanto a relação entre 1 e 2 Tessalonicenses, vide Figura 6.8a, quanto a relação entre Colossenses e Efésios, conforme as Figuras 6.8b e 6.8c se confirmou em todas as análises e de forma significativa. Também demonstram, recorrentemente, similaridade com as epístolas autênticas de Paulo, que pode ajudar a

¹⁰1Co 1:1

concluir que derivam do mesmo autor. A similaridade acentuada nas duas relações aqui expostas mostram semelhanças com a relação Romanos-Gálatas, o que pode indicar o reuso de material por parte do autor.

Figura 6.9 – Comparações de similaridade das epístolas pastorais pela análise Doc2Vec



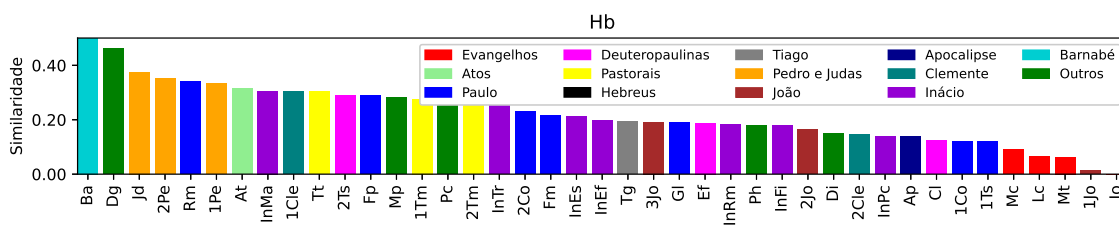
Fonte: o autor

As três epístolas pastorais constituem um grupo bem definido devido as suas múltiplas similaridades em todas as análises, conforme a Figura 6.9, inclusive com o SBERT, porém em menor nível nesse último caso. Contudo, apresenta baixa afinidade com o restante das epístolas paulinas, evidenciando assim que, dificilmente, se trata de uma autoria em comum ou há interferência de terceiros, como um secretário ou coautor.

6.3.7 Epístola aos Hebreus

A atribuição de autoria de Hebreus se mostrou um desafio considerável devido à variabilidade dos resultados. A mesma autoria da Epístola de Barnabé seria a opção mais provável, dado os escritos disponíveis devido aos resultados mais consistentes nas análises, conforme exibido na Figura 6.10, entretanto, não exibem uma afinidade definitiva

Figura 6.10 – Comparações de similaridade da Epístola aos Hebreus pela análise Doc2Vec



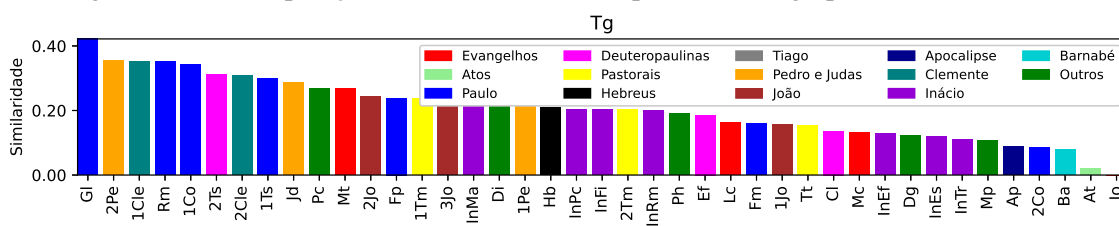
Fonte: o autor

que permita ser categórico na atribuição de autoria. A relação com os escritos de Paulo se mostrou relativamente fraca e demonstram a improbabilidade da autoria paulina.

O texto mostra alguma relação com os trabalhos de Lucas e de 1 Clemente, o que pode evidenciar uma similaridade com uma escrita mais helenística do texto. Com todos esses pontos levantados, se mostra mais prudente a atribuição de autoria de Hebreus a um judeu helenizado que não teve outras obras sobreviventes.

6.3.8 Epístola de Tiago

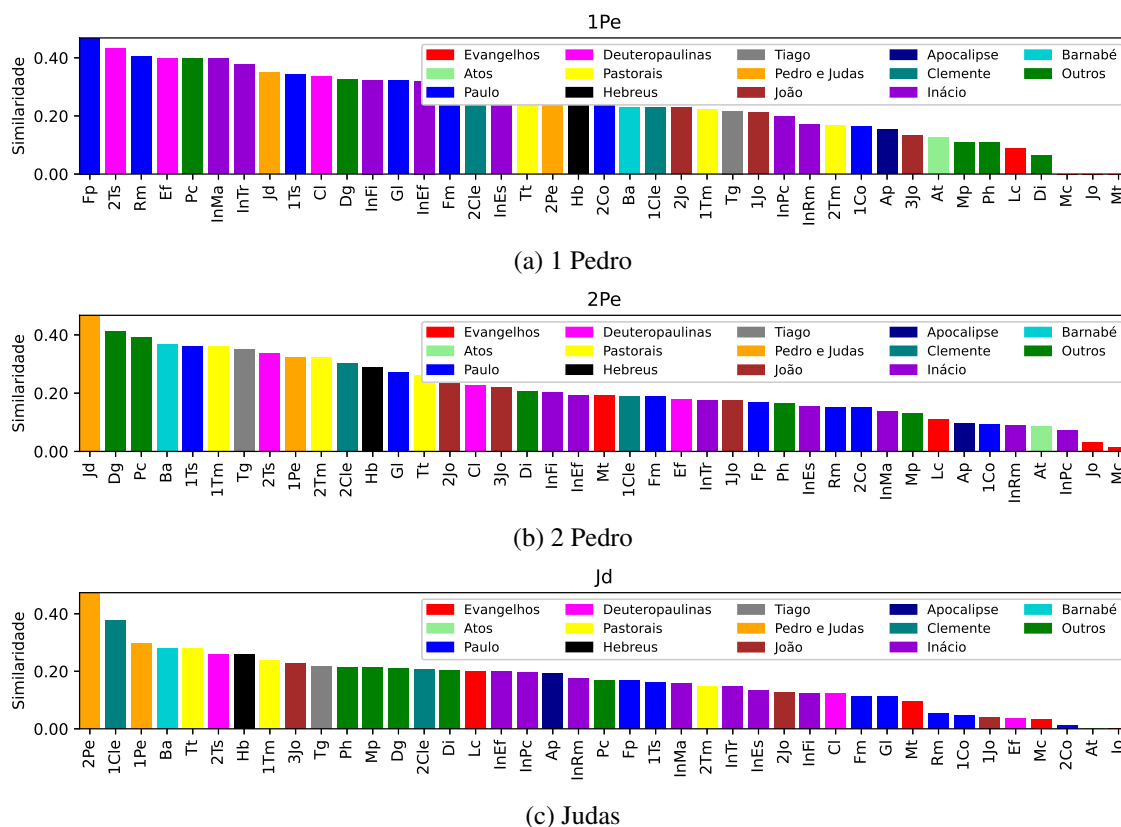
Figura 6.11 – Comparações de similaridade da Epístola de Tiago pela análise Doc2Vec



Fonte: o autor

A carta de Tiago mostra pouca relação com as demais obras analisadas nesse trabalho. Isso se evidencia pela variabilidade dos resultados e baixa similaridade com outros textos. A única relação com algum destaque seria a similaridade com Romanos-Gálatas que, entretanto, não se mostra de forma enfática entre as diferentes análises, exemplificado pela Figura 6.11. Os possíveis paralelos relatadas na Seção 2.1.5 não se mostraram relevantes nas análises. Parece razoável concluir que esse trabalho seja a única obra desse autor, podendo servir para afirmar a autoria de Tiago.

Figura 6.12 – Comparações de similaridade das epístolas petrinhas e Judas pela análise Doc2Vec



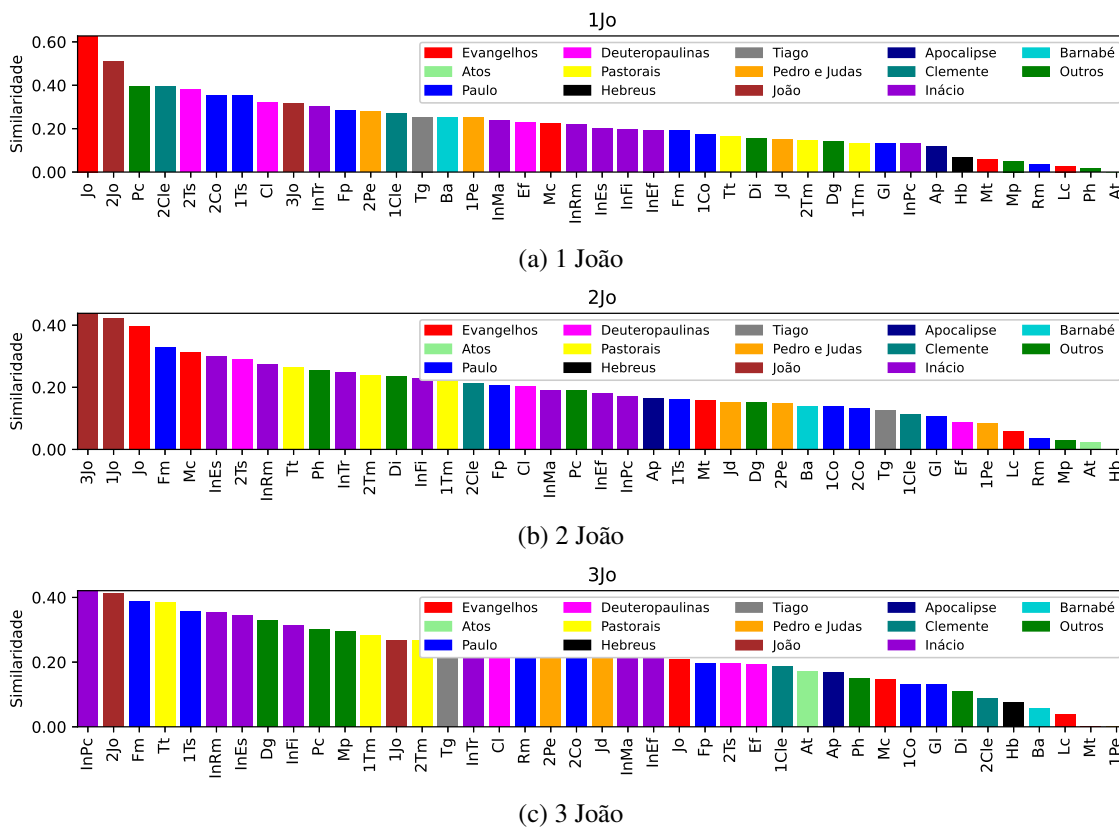
Fonte: o autor

6.3.9 Epístolas Petrinhas e Judas

A relação entre Judas e 2 Pedro se confirma em todas as análises e com similaridade relevante. Judas ainda apresenta alguma relação com 1 Clemente e 1 Pedro, porém, em menor grau com esta última epístola, conforme mostrado na Figura 6.12c. Já 2 Pedro apresenta grande variabilidade nos resultados, sendo difícil apontar alguma relação adicional, vide Figura 6.12b.

A epístola de 1 Pedro também mostra significativa variabilidade nos resultados, podendo evidenciar uma autoria exclusiva e, principalmente, dificilmente podendo ser relacionada com 2 Pedro, conforme a Figura 6.12a. Esta epístola também não se relaciona consistentemente com 2 Coríntios, 1 e 2 Tessalonicenses, epístolas com coautoria de Silvano.

Figura 6.13 – Comparações de similaridade das epístolas joaninas pela análise Doc2Vec



Fonte: o autor

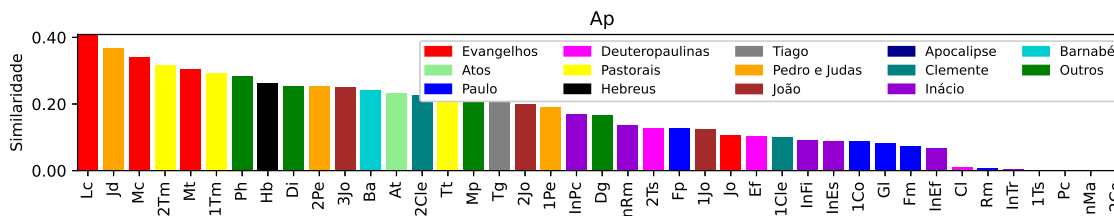
6.3.10 Epístolas Joaninas

As epístolas de João apresentam um padrão similar as cartas autênticas de Paulo, relações fortes com alguns textos e similaridade inferior se considerado a totalidade do conjunto. Na Figura 6.13a, a epístola de 1 João apresenta grande similaridade com o Evangelho de João e, em menor grau, com 2 João. A epístola de 2 João já mostra relação com 1 João e em seguida com 3 João na Figura 6.13b. A carta de 3 João já exibe uma relação somente com 2 João, apresentando maior variabilidade nos resultados, conforme a Figura 6.13c.

Parece razoável afirmar a autoria de 1 João pelo mesmo autor do Evangelho de João. As demais duas cartas também podem ser atribuídas ao mesmo autor, contando que nenhuma outra obra analisada apresentou uma afinidade significativa com esse conjunto de epístolas e há um padrão de similaridade nesse grupo de epístolas.

6.3.11 Apocalipse

Figura 6.14 – Comparações de similaridade do Apocalipse pela análise Doc2Vec



Fonte: o autor

Apocalipse apresenta relações de similaridade com os evangelhos sinóticos e Atos, algo que, a primeira vista, parece ser algo pouco óbvio. Contudo, apresenta também uma variabilidade nas diferentes análises e índice de similaridade relativamente baixo, exceto na análise TF-IDF, conforme a Figura 6.14.

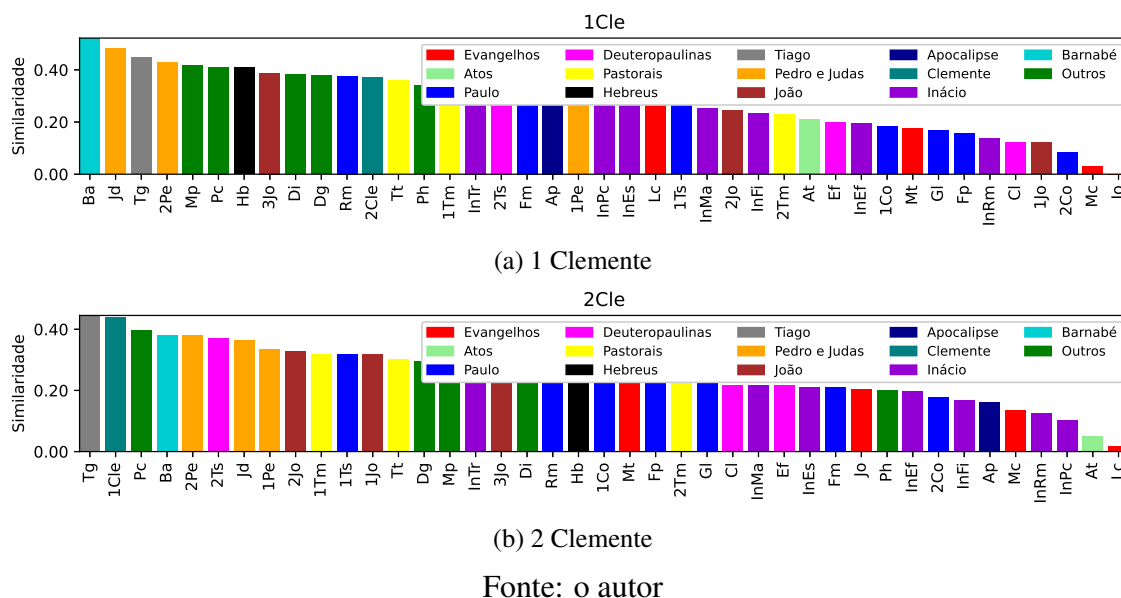
Com isso, a autoria de Apocalipse se trata de um tema difícil, restando apenas a possibilidade de se levantar hipóteses. A relativa proximidade com os evangelhos e baixa similaridade textual pode revelar a autoria exclusiva dessa obra, podendo ser especulado a autoria por João, apóstolo e evangelista, em seu possível exílio na ilha de Patmos, conforme relatado na Seção 2.1.8.

6.3.12 Epístolas de Clemente de Roma

A primeira epístola de Clemente apresenta variabilidade nos resultados, sendo a Epístola de Barnabé a única obra que aparece recorrentemente nas alternativas prováveis de similaridade e mostrando fraca relação com 2 Clemente, vide Figura 6.15a. Como na Figura 6.15b, a chamada segunda epístola de Clemente tem 1 Clemente como um dos textos mais similar, contudo, essa relação é instável através das diferentes análises.

Apesar de que, a primeira vista, esses resultados podem ser vistos como contraditórios, podemos levantar a hipótese que a relativa similaridade de 2 Clemente seja fruto de uma influência do trabalho de 1 Clemente e que o autor dos dois textos difiram, conforme as análises dos críticos textuais e comentaristas bíblicos. Podemos embasar essa hipótese no fato de 1 Clemente não se assemelhar com 2 Clemente e que 1 Clemente ter mais de três vezes o comprimento da carta de 2 Clemente, podendo 2 Clemente ter porções significativas de seu conteúdo em comum com 1 Clemente, porém, a recíproca pode não ser verdadeira.

Figura 6.15 – Comparações de similaridade das epístolas de Clemente de Roma pela análise Doc2Vec



Fonte: o autor

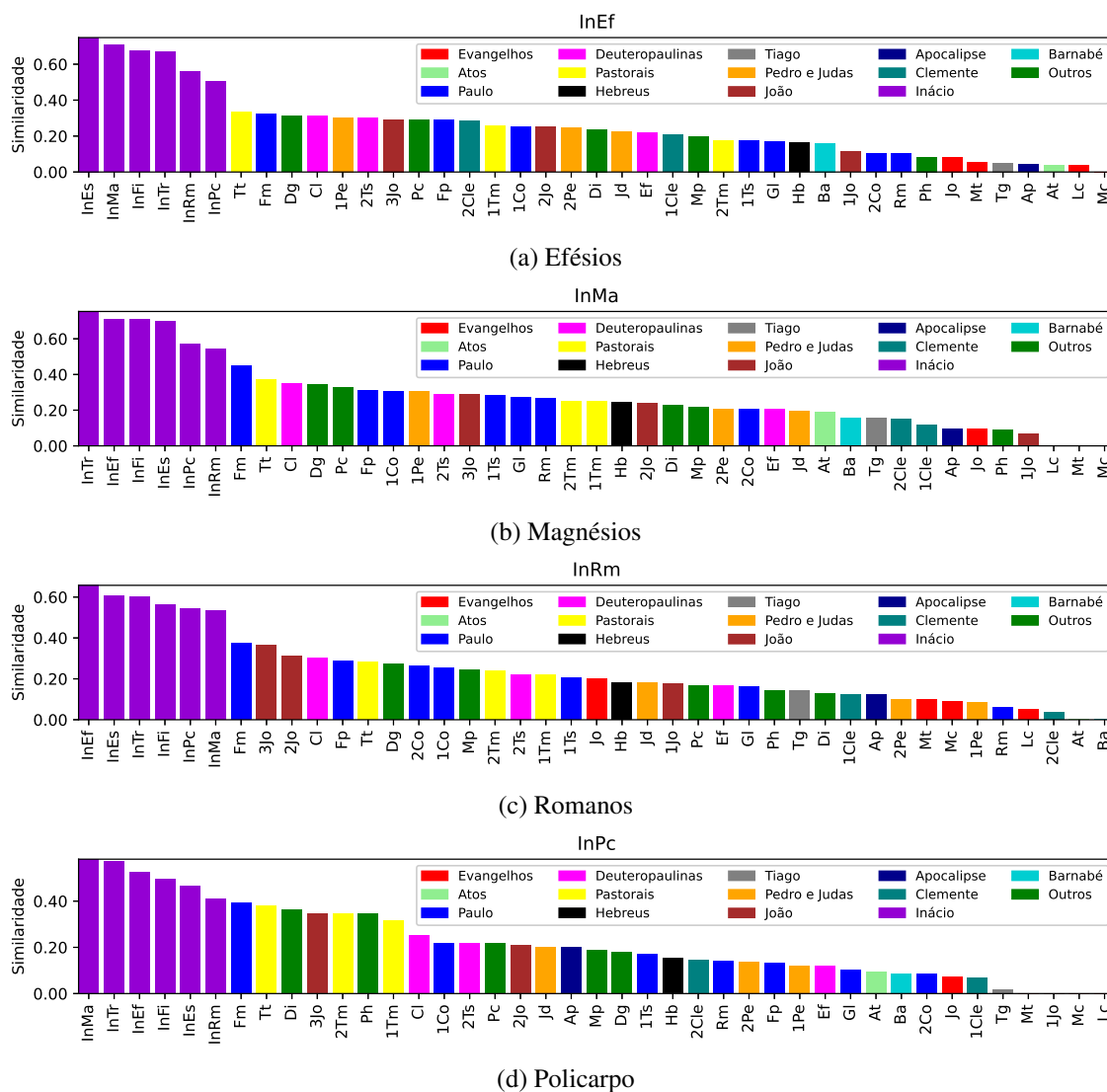
6.3.13 Epístolas de Inácio de Antioquia

As Epístolas de Inácio constituem, de longe, o grupo de cartas mais coeso e uniforme. As cartas desse grupo em, praticamente, todas as análises sempre despontam no topo de similaridade mutualmente, conforme a Figura 6.16. Após esse núcleo de epístolas, é possível verificar similaridades recorrentes com algumas epístolas paulinas e outras obras do Novo Testamento, principalmente na análise TF-IDF. Isso pode ser explicado pelo uso de citações de textos do Novo Testamento, como relatado na Seção 2.2.2.

6.3.14 Epístola de Policarpo aos Filipenses

A Epístola de Policarpo não apresenta alguma relação bem definida com outros escritos analisado. Na Figura 6.17, é possível notar que as obras mais similares não compartilham similaridade óbvias com a obra analisada nem entre si. A análise SBERT apresenta alguma similaridade com o grupo de cartas de Inácio de Antioquia, talvez refletindo a interação entre Policarpo e Inácio, conforme relatada na Seção 2.2.2. Esses dados parecem corroborar para afirmar uma autoria exclusiva dessa obra.

Figura 6.16 – Comparações de similaridade das epístolas de Inácio de Antioquia pela análise Doc2Vec



Fonte: o autor

6.3.15 Martírio de Policarpo

A única relação clara nos dados das diferentes análises é a similaridade com os textos de Lucas, principalmente Atos dos Apóstolos, conforme a Figura 6.18. Utilizando-se dos paralelos relatados entre esta carta e diversos outros textos bíblicos, é possível inferir que essa relação de similaridade seja devido a influências das narrativas de martírio do Novo Testamento. Atos dos Apóstolos é o livro que descreve os martírios de Estevão¹¹, considerado o primeiro mártir cristão, e de Tiago¹², irmão do apóstolo João.

¹¹ At 7:54-60

¹² At 12:1-2

Figura 6.17 – Comparações de similaridade da Epístola de Policarpo aos Filipenses pela análise Doc2Vec

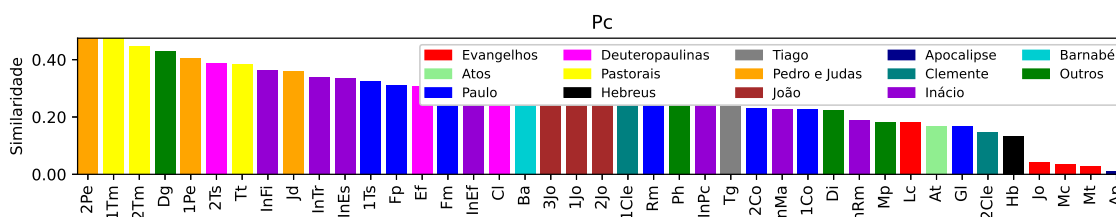
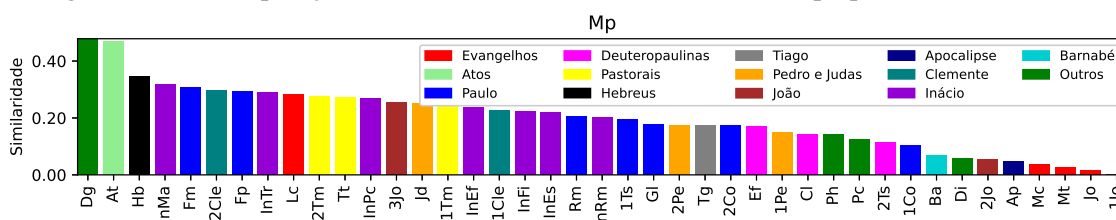
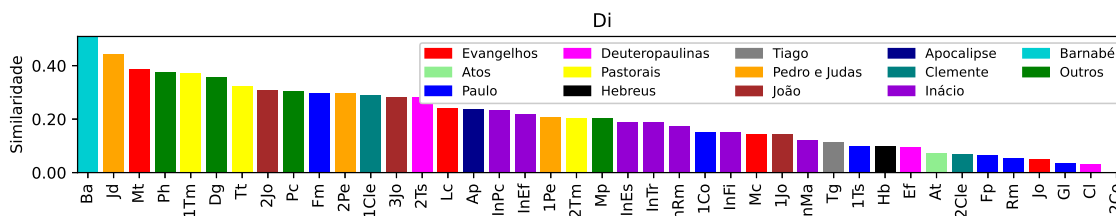


Figura 6.18 – Comparações de similaridade do Martírio de Policarpo pela análise Doc2Vec



6.3.16 Didaquê

Figura 6.19 – Comparações de similaridade do Didaquê pela análise Doc2Vec

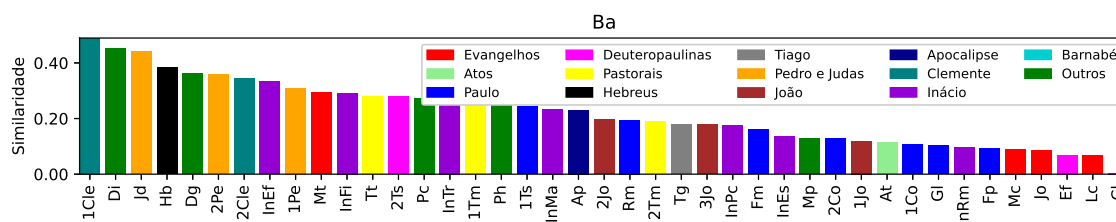


O Didaquê exibe uma relação relevante com a Epístola de Barnabé, principalmente nas análises TF-IDF e Doc2Vec, vide Figura 6.19. Outras relações menos bem definidas foram notadas entre as diferentes análises, como a similaridade com o Evangelho de Mateus, Pastor de Hermas e 1 Timóteo. Paralelos entre o Didaquê e os textos do Evangelho de Mateus e a Epístola de Barnabé são conhecidos, conforme relatados na Seção 2.2.5.

6.3.17 Epístola de Barnabé

A Epístola de Barnabé apresenta variabilidade em seus resultados ao longo das diferentes análises. Nas análises Doc2Vec, as obras mais semelhantes foram o Didaquê, Hebreus, Judas e 1 Clemente, porém, sem uma ordem definida entre esses textos, vide

Figura 6.20 – Comparações de similaridade da Epístola de Barnabé pela análise Doc2Vec

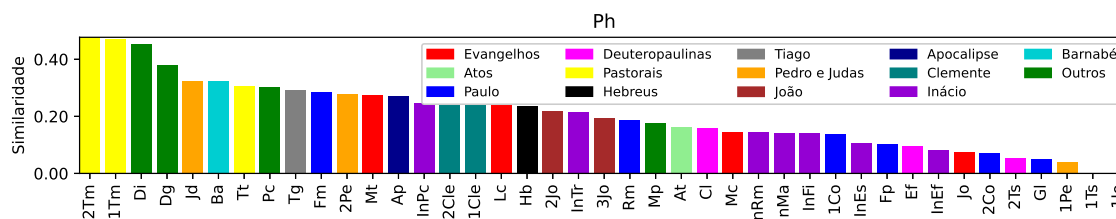


Fonte: o autor

Figura 6.20. No modelo TF-IDF, os destaques ficam para o Evangelho de João, seguido por 1 Clemente e os demais evangelhos sinóticos. Utilizando o SBERT, o texto mais similar segue sendo 1 Clemente, seguido por Pastor de Hermas e as epístolas de Inácio de Antioquia. Conforme mencionado na Seção 2.2.6, esta obra recorre a citações do Novo Testamento, foi citada por outras obras dos Pais Apostólicos e discorre sobre temas comuns a tradição cristã, o que pode ajudar a explicar as similaridades esporádicas relatadas nas análises, bem como ajudar a afirmar a autoria exclusiva desta obra.

6.3.18 Pastor de Hermas

Figura 6.21 – Comparações de similaridade do Pastor de Hermas pela análise Doc2Vec



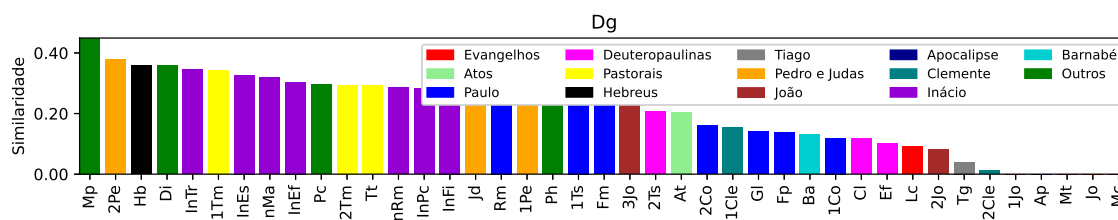
Fonte: o autor

Esta obra exibe fraca similaridade com livro de Apocalipse e grande variância nos resultados obtidos, conforme mostrado na Figura 6.21. Provavelmente, a melhor hipótese que se pode levantar com os dados obtidos é a viabilidade da autoria exclusiva desta obra, algo que reforça a hipótese da autoria de Hermas, conforme relatado na Seção 2.2.7.

6.3.19 Epístola a Diogneto

Em todas as análises, essa obra revela baixos índices de similaridade com as demais obras e alto grau de instabilidade nos resultados, vide Figura 6.22. As possíveis relações com os escritos de 1 Clemente e Policarpo não se confirmaram relevantes nas

Figura 6.22 – Comparações de similaridade da Epístola a Diogneto pela análise Doc2Vec



Fonte: o autor

diversas análises. Isso pode ajudar a reforçar a hipótese levantada na Seção 2.2.8, que essa obra seria de Quadratus de Atenas devido à diferença notória para com as demais obras que compõem o presente trabalho.

6.4 Sumário

Nesse Capítulo foi exemplificada uma implementação possível da metodologia descrita no Capítulo 5, relatado o processo de validação adotado e os resultados obtidos. Importantes resultados foram exibidos e discutidos, realizando paralelos com as relações descritas no Capítulo 2, finalizando a análise experimental.

7 CONCLUSÃO

Este trabalho utilizou diferentes técnicas de PLN para quantificar a similaridade dos textos do Novo Testamento e dos Pais Apostólicos, com uma análise comparativa com as relações descritas por críticos textuais e comentaristas bíblicos. O uso de quatro análises distintas se mostrou efetiva para diferenciar relações ocasionais de relações reais e recorrentes, contribuindo com a robustez geral do trabalho. Com os resultados obtidos, o presente trabalho visa fomentar e demonstrar a possibilidade do uso de técnicas computacionais na área de atribuição de autoria e estilometria de textos diversos.

Diversas das hipóteses levantadas foram testadas, onde importantes resultados foram obtidos. No Novo Testamento podemos destacar algumas relações como: a coesão dos evangelhos e Atos dos Apóstolos, similaridades significativas entre os grupos de cartas autênticas de Paulo e as cartas deuteropaulinas, a dissimilaridade das epístolas pastorais com o restante do *corpus* paulino, a dissimilaridade entre as epístolas petrinas, a relação de similaridade de 2 Pedro e Judas, bem como das epístolas joaninas. Também é possível verificar, de maneira geral, um afastamento dos escritos do Novo Testamento em relação aos escritos dos Pais Apostólicos, o que pode ajudar a demonstrar a preservação do cânon bíblico em relação a obras não canônicas.

Dentre os quatro problemas textuais específicos mencionados na Introdução, os resultados obtidos apontam para: 1) o afastamento da Epístola aos Hebreus em relação ao *corpus* paulino, bem como a possível autoria exclusiva dessa obra; 2) a plausibilidade da realocação da Perícope da Adúltera do Evangelho de Lucas para o Evangelho de João; 3) a dissimilaridade do Final Longo de Marcos com o restante do texto, e, por fim, 4) a coesão textual do Capítulo 21 do Evangelho de João com as demais partes da obra.

Nas análises dos Pais Apostólicos, foram exibidos resultados relevantes, entretanto, devido à escassez de obras do período e a provável autoria exclusiva de diversas obras, foram limitadas as possibilidades de se obter dados mais conclusivos em algumas análises. Podemos destacar: a relação de dissimilaridade e assimetria entre as epístolas de Clemente de Roma, a unidade das cartas de Inácio de Antioquia e as relações do Didaquê com a Epístola de Barnabé e o Evangelho de Mateus. As relações exibidas no estudo do Didaquê se mostram importante por demonstrar a capacidade das análises de capturar relações mais sutis.

O uso de técnicas de redução de dimensionalidade nas análises foram discutidas, sendo apresentado seus benefícios e limitações. O presente trabalho exibiu resultados

empíricos que mostram a dificuldade na aplicação dessas técnicas em tarefas de análises visuais qualitativas sem serem reveladas distorções importantes nos resultados apresentados, como no estudo de relações de similaridade de documentos, onde vetores com centenas ou milhares de dimensões são reduzidas para vetores bidimensionais.

A agregação de *embeddings* contextuais do SBERT se mostrou uma alternativa viável para possibilitar a aplicação desse importante modelo de PLN a textos maiores que sua limitação normalmente permite. Apesar de apresentar bons resultados em variados cenários, essa técnica de agregação de *embeddings* leva à diluição dos valores de similaridade. Outros trabalhos como Altakrori, Cheung e Fung (2021) e Custódio e Paraboni (2021) mostram que o BERT pode ter resultados similares ou piores que técnicas clássicas como *word N-grams* em tarefas de atribuição de autoria.

Existe espaço para uma vasta gama de trabalhos futuros na área. A extensão natural do presente trabalho trata-se de realizar a análise do Antigo Testamento. Existem diversos problemas em aberto dentro dessa literatura, como verificar a coesão textual do Pentateuco e avaliar a teoria da elaboração em etapas, por diferentes autores, dos livros de Isaías e Daniel. Descobertas relativamente recentes, como a descoberta dos manuscritos do Mar Marto tem colocado em cheque a teoria da compilação tardia da bíblia hebraica, o que pode mudar completamente o entendimento da história do povo israelita e da linguística semita. Trabalhos como estes podem contribuir com a discussão desses temas.

Um aprofundamento do presente trabalho seria examinar características intratextuais do Novo Testamento, como a teoria de que o texto de Atos dos Apóstolos seria uma obra composta em duas partes. Outra possibilidade se trata de analisar as relações do Novo Testamento em contraste aos pseudoepígrafos contemporâneos, como o Apocalipse de Pedro e o Evangelho de Tomé. Muitos desses textos exibem características interessantes e, sendo analisados, podem resultar em uma compreensão melhor do contexto histórico e do desenvolvimento do cristianismo ao longo da história.

REFERÊNCIAS

- ALTAKRORI, M.; CHEUNG, J. C. K.; FUNG, B. C. The topic confusion task: A novel evaluation scenario for authorship attribution. In: **Findings of the Association for Computational Linguistics: EMNLP 2021**. [S.l.: s.n.], 2021. p. 4242–4256.
- BARR, G. K. The impact of scalometry on new testament letters. **The Expository Times**, Sage Publications Sage CA: Thousand Oaks, CA, v. 114, n. 1, p. 3–9, 2002.
- BARR, G. K. Two styles in the new testament epistles. **Literary and linguistic computing**, Oxford University Press, v. 18, n. 3, p. 235–248, 2003.
- BARTON, J.; MUDDIMAN, J. **The Oxford Bible Commentary**. OUP Oxford, 2001. ISBN 9780191567254. Disponível em: <<https://archive.org/details/oxfordbiblecomme000unse/page/830/mode/2up?view=theater>>.
- BEEL, J. et al. Research-paper recommender systems: a literature survey. **International Journal on Digital Libraries**, v. 17, n. 4, p. 305–338, Nov 2016. ISSN 1432-1300. Disponível em: <<https://doi.org/10.1007/s00799-015-0156-0>>.
- BOJANOWSKI, P. et al. Enriching word vectors with subword information. **Transactions of the association for computational linguistics**, MIT Press, v. 5, p. 135–146, 2017.
- BOND, T. et al. **Applying the Rasch Model: Fundamental Measurement in the Human Sciences**. Routledge, 2020. ISBN 9780367141424. Disponível em: <<https://books.google.com.br/books?id=Oz9jzQEACAAJ>>.
- BROWN, R.; SOARDS, M. **An Introduction to the New Testament**. Yale University Press, 2016. (Anchor Bible Reference). ISBN 9780300173123. Disponível em: <<https://books.google.com.br/books?id=MkLOCwAAQBAJ>>.
- CUSTÓDIO, J. E.; PARABONI, I. Stacked authorship attribution of digital texts. **Expert Systems with Applications**, v. 176, p. 114866, 2021. ISSN 0957-4174. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0957417421003079>>.
- DEVLIN, J. et al. Bert: Pre-training of deep bidirectional transformers for language understanding. **arXiv preprint arXiv:1810.04805**, 2018.
- EHRMAN, B. **Pauline Forgeries: 2 Thessalonians as a Test Case**. 2014. Disponível em: <<https://ehrmanblog.org/pauline-forgeries-2-thessalonians-as-a-test-case>>.
- FATHERS, A.; GALLI, M. **The Apostolic Fathers**. Moody Publishers, 2009. (Moody Classics). ISBN 9781575673318. Disponível em: <<https://books.google.com.br/books?id=8O9lZTqxI0oC>>.
- FLYAMER, I. et al. Phlya/adjusttext: 0.8 beta. Zenodo, Jun 2020.
- GADDIS, M. **There Is No Crime for Those Who Have Christ: Religious Violence in the Christian Roman Empire**. University of California Press, 2015. (Joan Palevsky imprint in classical literature). ISBN 9780520286245. Disponível em: <<https://books.google.com.br/books?id=iaowDwAAQBAJ>>.

GRUEN, D. **2 Peter and Jude**. NavPress, 2018. (Life change series). ISBN 9780891099949. Disponível em: <<https://books.google.com.br/books?id=xZp5DwAAQBAJ>>.

HARRINGTON, D. J. **Biblical Criticism**. 2011. Disponível em: <<https://www.oxfordbibliographies.com/view/document/obo-9780195393361/obo-9780195393361-0066.xml>>.

HARTOG, P. **Polycarp's Epistle to the Philippians and the Martyrdom of Polycarp: Introduction, Text, and Commentary**. University Press, 2013. (Oxford Apostolic Fathers). ISBN 9780199228393. Disponível em: <<https://books.google.com.br/books?id=eZkRAAAAQBAJ>>.

HOLMES, M. **The Apostolic Fathers in English**. Baker Publishing Group, 2006. ISBN 9780801031083. Disponível em: <<https://books.google.com.br/books?id=iHwnpAGffRMC>>.

HUNTER, J. D. Matplotlib: A 2d graphics environment. **Computing in Science & Engineering**, IEEE COMPUTER SOC, v. 9, n. 3, p. 90–95, 2007.

KOESTER, H. **Introduction to the New Testament**. [S.l.]: Walter de Gruyter, 1995. (Einführung in das Neue Testament).

LAKE, K. **The Apostolic Fathers**. [S.l.]: Heinemann, 1914. (The Loeb classical library).

LAWRENCE, D. **Apocalypse: And the Writings on Revelation**. RosettaBooks, 2019. (The Definitive Cambridge Editions of D.H. Lawrence). ISBN 9780795351426. Disponível em: <<https://books.google.com.br/books?id=NXLnDwAAQBAJ>>.

LE, Q.; MIKOLOV, T. Distributed representations of sentences and documents. In: PMLR. **International conference on machine learning**. [S.l.], 2014. p. 1188–1196.

LEE, J.; VERLEYSSEN, M. **Nonlinear Dimensionality Reduction**. Springer New York, 2007. (Information Science and Statistics). ISBN 9780387393513. Disponível em: <https://books.google.com.br/books?id=o_TIoyeO7AsC>.

MARSHALL, I.; TOWNER, P. **The Pastoral Epistles**. Bloomsbury Academic, 2004. (International Critical Commentary). ISBN 9780567084552. Disponível em: <<https://books.google.com.br/books?id=jA7UAwAAQBAJ>>.

MEALAND, D. L. Hellenistic greek and the new testament: A stylometric perspective. **Journal for the Study of the New Testament**, Sage Publications Sage UK: London, England, v. 34, n. 4, p. 323–345, 2012.

MIKOLOV, T. et al. Distributed representations of words and phrases and their compositionality. **Advances in neural information processing systems**, v. 26, 2013.

PUTNINŠ, T. et al. Advanced text authorship detection methods and their application to biblical texts. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. **Complex Systems**. [S.l.], 2006. v. 6039, p. 60390J.

REIMERS, N.; GUREVYCH, I. Sentence-bert: Sentence embeddings using siamese bert-networks. **arXiv preprint arXiv:1908.10084**, 2019.

REIMERS, N.; GUREVYCH, I. Making monolingual sentence embeddings multilingual using knowledge distillation. **arXiv preprint arXiv:2004.09813**, 2020.

RICHARDS, E. **Paul and First-Century Letter Writing: Secretaries, Composition and Collection**. InterVarsity Press, 2004. ISBN 9780830827886. Disponível em: <<https://books.google.com.br/books?id=8a-o6rsm-l4C>>.

ROYAL, K. Using objective stylometric techniques to evaluate new testament authorship. **Journal of MultiDisciplinary Evaluation**, v. 8, n. 19, p. 1–7, 2012.

SNAPP JR., J. E. **The Origin of Mark 16:9-20**. 2007. Disponível em: <https://web.archive.org/web/2021*/www.textexcavation.com/snapp/PDF/snapporiginmk.pdf>.

TAUBER, J.; MACDONALD, S. **Open Apostolic Fathers**. [S.l.]: Zenodo, 2019.

VASWANI, A. et al. Attention is all you need. **Advances in neural information processing systems**, v. 30, 2017.

WALLACE, D. Revisiting the corruption of the new testament: Manuscript, patristic, and apocryphal evidence. In: . Kregel Academic, 2011, (Text and canon of the New Testament). chp. 1. ISBN 9780825489068. Disponível em: <<https://books.google.com.br/books?id=838A8BDUI5kC>>.

WHITE, G. **The Net Bible Synopsis of the Four Gospels**. Biblical Studies Press, 2009. ISBN 9780737501537. Disponível em: <<https://books.google.com.br/books?id=j-swPWda70UC>>.