Scientific
Research

Unsupervised Learning of Two Bible Books: Proverbs and Psalms

Wei Hu

Department of Computer Science, Houghton College, New York, USA
Email: wei.hu@houghton.edu

The book of Proverbs teaches the wisdom of life that was relevant in the days of King Solomon, the principal author of this book and a son of King David, but more importantly it is still applicable and needed in the life of today. The book of Psalms is the longest book and perhaps the most widely read book of the Bible, which contains 150 songs and prayers with King David as its main composer. The proverbs imparts that the success in life depends on personal choices and actions that relate to other people, while the psalms are used in worship that relates to God. In this report, we apply unsupervised learning to the study of these two books because of their shared association with wisdom. The chapters in each of the two books are grouped by content. Similar chapters and verses between these two books, written by a father and a son, are also identified, allowing one book to illuminate the other. Our computational findings match those by Biblical scholars at large, but have made a few new discoveries that could not be accomplished by traditional methods.

Keywords: Bible; Proverbs; Psalms; Affinity Propagation; Clustering Algorithm; Topic Model; Machine Learning; Text Mining

Introduction

Proverbs and Psalms are two books in the Old Testament (OT) of the Bible. Both were written by several authors with one main author for each: King David for Psalms and King Solomon, a son of David, for Proverbs. Psalms is a collecton of songs and prayers that reveal human hearts in worshiping God, while Proverbs offers principles and instructions for wise daily living with fellow human beings.

David was the youngest of the eight sons of Jesse from the tribe of Juda. During his childhood, he tended his father's sheep in the fields around Bethlehem. With God on his side, David as a young boy fought and killed the Philistine giant Goliath in a battle. King Saul, Israel's first king, became jealous of David's popularity and tried to kill him a few times, although Saul was initially impressed by David's talents as a soldier and musician. As a result, David had to flee and hide out from Saul. During this period of severe trial in life, David demonstrated his faith and trust in God, as seen from the psalms he wrote during this time. After Saul's death, David became the king of Juda, and the king of all 12 tribes of Israel seven years later.

David was a man after God's own heart (I Samuel 13:14). Yet, he was not a perfect man. After his adultery with Bathsheba and the indirect murdering of her husband, David made her his wife. Prophet Nathan, sent by God, confronted him. David repented of his sin and God pardoned him. However, several penalties as consequences of his sin still followed including the death of his first child with Bathsheba. Their second child was Solomon, who succeeded his father as a king of Israel.

David was a shepherd, musician, singer, composer, warrior, and king. He composed many of the psalms collected in the book of Psalms, expressing praises of the wonderful creation, grace, and mercy of God, and the strugles, fears, and pains in life. The inspiring story of David vs. Goliath becomes a classic example of the weak unexpectedly defeating the strong.

King David was one of the great patriarchs of Israel. The Gospel of Matthew describes Christ as a son of David, because the promised Messiah was to be a descendant of David, proclaimed by the prophets during the Old Testament times.

When Solomon became the third king of Israel, he was young and inexperienced to rule the nation. God appeared to him, offering him anything he requested, but he asked for wisdom to be a wise king. God was very pleased with Solomon's answer and promised him great riches and honor in addition to wisdom. He wrote 1005 songs and 3000 proverbs (1 Kings 4:32). He was the builder of the first temple in Jerusalem, although his father King David engaged in collecting building materials for the construction of this temple. Israel enjoyed 40 years of peace and prosperity during the reign of King Solomon, the best years in the history of this nation. He was the last king of the united kingdom of Israel before its division into the northern and southern kingdoms. Unfortunately, Solomon squandered his wisdom in the later stage of his life, which reminds us that wisdom must be used in a reverent relationship with God.

The rich life experience of David prepared him to compose many psalms in the book of Psalms. His son Solomon, with God's inspired wisdom, wrote many proverbs. Solomon was also the author of the other two books of the Bible: Ecclesiastes and Song of Songs. The OT could be divided into three divisions: Law, Prophets, and Writings. The book of Psalms and Proverbs belong to the third category.

The psalms were written from the time of Moses to the post-exilic period that spans almost one thousand years, which played a critical role in the public and private worship of ancient Israel. No book in the Old Testament has been read so much in the history of mankind as the book of Psalms (Mowinckel, 2004). It is the longest book in the Bible and covers the most diverse topics. It includes the creation, judgment, and

salvation of God, the history of Israel, the law of life, Davidic messiah, warning against wickedness and exhortation to righteousness, the glory, mystery, and misery of human conditions, and the kingdom of God. Psalms, along with Isaiah, are the two Old Testament books most cited in the New Testament (Mays, 1994).

The book of Proverbs is a special type of literature, composed of a collection of maxims covering a wide range of subjects in human life. As stated clearly in the beginning of this book, the main purpose of Proverbs is to teach people to be wise through the fear of the Lord. The phrase "the fear of the Lord" appears so many times in this book than any books else in the Bible, indicating its central role in Proverbs. The commandments expressed in the Law of Moses can never cover every possible situation in the life of a man. Therefore, the only way to live a life is to develop wisdom from the fear of God.

The aim of Proverbs is to provide readers with principles as well as practical and executable instructions on applying the fear of the Lord to their lives, i.e., divine wisdom for daily life. One feature of this book is that it imparts wisdom by means of personification and discourse. It shows how the free will of man can be best exercised under and harmonized with the sovereignty of God. The major conclusion of this book is that the ultimate wisdom is to be right with God. From this perspective, a knowledgeable person may be deficient in wisdom.

Only Proverbs and Psalms in the Old Testament demonstrate composite authorship, and yet their main contributor is David (father) and Solomon (son) respectively. Both books display similar Hebrew poetic form, and both elaborate on "the fear of the Lord". The father-son relationship could also be seen in these two books.

Wisdom is the central theme of Proverbs, but a portion of Psalms also deals with wisdom. Both books describe two distinctive pathways. The way of the righteous is guided by God and therefore leads to salvation, in contrast, the way of the wicked leads to destruction. These books demonstrate that wisdom and obedience to God are closely bound, and both pathways lead to their respective destinations.

Identifying groups of chapters or verses of similar content from Proverbs and Psalms could enhance the learning of these two books. To this end, computational techniques, topic model and clustering, are employed in this study. We first cluster the chapters of each book, and then find the correlation between these two books at chapter level and at verse level, allowing one book to illuminate the other.

Materials and Methods

Materials

The text of the Bible used in this study is from the King James version (1611 authorized version), downloaded from http://printkjv.ifbweb.com. Because Proverbs and Psalms are used as data in our study, this section provides some background information about them.

Proverbs

The book of Proverbs offers the wisdom of daily living, and explains how life, as God designed it, works in general. The first six verses in Proverbs serve as an introduction of this book. They state the purpose and theme of the whole book. Selected and streamlined, the proverbs are given for attaining wisdom, acquiring a prudent life, knowing what is right and wrong. The seventh verse and first proverb of this book teaches the fear of the Lord as the very beginning of knowledge and all wisdom, the foremost principle repeated throughout the whole book.

Solomon is the chief author as the majority of the chapters are under his name, including chapters 10:1 - 22:16 and 25:1 - 29:27. Because the name of Solomon shows in the start of chapter 10, it is reasonable to believe the writings of chapters 1 - 9 do not belong to him. Proverbs can be divided into several collections. There is a general title (Proverbs 1:1 - 7), introductory to the whole book. The first collection is made of chapters 1:8 - 9:18, composed not by Solomon, but by other wise men. The second collection is "The proverbs of Solomon" (Proverbs 10:1 - 22:16). The third is "The sayings of the wise" (Proverbs 22:17 - 24:22) and "These also are sayings of the wise" (Proverbs 24:23 - 34). The fourth is "These are other proverbs of Solomon that the officials of king Hezekiah of Judah copied" (Proverbs 25 - 29). The fifth is "The words of Agur" (Proverbs 30). The seventh is "The words of king Lemuel of Massa, which his mother taught him" (Proverbs 31:1 - 31:9). The eighth is "The ideal wise woman" (Proverbs 31:10 - 31:31) (Stuart, 1852). The second collection of Solomon's proverbs contains proverbs extended to several verses, a clear contrast to his first collection. However, as in the first collection, the second also contains repetition of words or phrases (Stuart, 1852).

The topics covered in Proverbs can be generally divided into three parts. The first part deals with ourselves that includes our emotions, our relationships, our tongues and bodies, godly traits, ungodly traits, and gluttony and drinking. The second talks about our world including economic matters, political matters, world of nature, cities. The third part presents principles (Pippert, 2003).

The structure of this book is very distinct. In the verses from chapter 1:8 to chapter 9:18, almost all of them are simple parallelisms, comprising two members or clauses in each verse. Yet, there are only 11 triples. Further, the distribution of these parallelisms is not even. There are 209 synonymous parallelisms, 36 synthetic, and four antithetic (Stuart, 1852).

Psalms

The word "psalms" in Hebrew means praises or hymns. This book comprises meditations and reflections expressed in various poetic forms. They include the prayer for help of an individual, corporate prayer for help, thanksgiving songs of an individual, hymns, and psalms of instruction (Mays, 1994). Although some psalms exhibit more than one genre, in general they can be viewed as psalms of lament, psalms of gratitude, wisdom psalms, and psalms of praise (Campbell, 2012). Of the 150 psalms, 73 are attributed to David, 2 to Solomon, 12 to the sons of Korah, 12 to Asaph, 1 to Heman, 1 to Ethan, and 1 to Moses. The Davidic psalms reflect his life and faith in God. Many of the psalms are based on the events in Hebrew history. For example, Psalm 51 is based on the story recorded in 2 Samuel 11 - 12.

The book of Psalms has an introduction (Psalms 1 and 2) and a conclusion (Psalms 150). In imitation of Pentateuch (the five books of Moses), the 150 psalms in this book could be further divided into five smaller books, Psalms 1 - 41 (the book of personal experience); 42 - 72 (the book of Elohim); 73 - 89 (the dark book); 90 - 106 (the book of the King); 107 - 150 (the book of praise) (Constable, 2012). This book could also be

logically grouped as: first Davidic collection (Psalms 3 - 41), first Korahite collection (Psalms 42 - 49), second Davidic collection (Psalms 51 - 70), Asaphite collection (Psalms 73 - 83), second Korahite collection (Psalms 84 - 88), third Davidic collection (Psalms 108 - 11), Egyptian hallel (Psalms 113 - 118), songs of ascents (Psalms 120 - 134), fourth Davidic collection (Psalms 138 - 145), and final hallel (Psalms 146 - 150) (Coulter, 2011).

The name of David appears in the superscriptions of 73 psalms out of the 150 in the book of Psalms. There are 13 psalms describing an event in David's life (Collins, 2007). Of the 49 psalms with no author ascription in the book of Psalms, two are designated as Davidic in other books of the Bible (Acts 4:25 refers to Psalm 2 and Hebrew 4:7 points to Psalm 95).

In book 1, there are three psalms lacking a superscript, and all the others carry a Davidic superscript. Textually, several psalms in book 1 are similar to those in book 2 including Psalms 14 and 53, Psalms 40 and 70, and Psalms 31:1 - 4 and 71:1 - 3, implying these two books were compiled around the same time (Anderson, 1994). Of interest is the fact that Psalm 31 has a Davidic superscript, but Psalm 71 does not have a superscript.

There are various methodologies of defining genres of the psalms. Genre refers to a group of texts that share similar mood, content, structure or phraseology. The interpretation of a text is mainly determined by the identification of its genre. Although all psalms are in the genre of poetry, a psalm can also be viewed on a narrower level such as a prayer, hymn, lament, thanksgiving, wisdom, kingship, meditation, or exhortation. The psalms in the hymn category share many common traits while each has its own character (Longman, 2009).

German scholar Hermann Gunkel (Gunkel & Horner, 1967; Gunkel, 1998) employed the form-critical approach to the psalms, classifying each psalm by its form and type and then attempting to identify the particular situation in the life of ancient Israel in which such a form might have been composed and used. Recognizing the recurring patterns of structure, theme, and compositional techniques of many psalms, he argued that psalms are not spontaneous prayers of individuals, but represent the fixed forms that communicated from generation to generation. Many scholars followed his approach in their study of the Psalms as well as in other portions of the Old Testament (Constable, 2012). The study by Gunkel identified several literary types of psalms, hymn, individual songs of thanksgiving, communal lament, individual lament, royal psalms, lesser types (pilgrimage songs, torah liturgies, and wisdom poetry). Based on these categories, Dennis Bratcher classified all 150 psalms (Bratcher, 2011) in the book of Psalms (**Table 1**).

Table 1.
Classifying the Psalms by Genre according to (Bratcher, 2011).

Lament Psalms	
Community	12, 44, 58, 60, 74, 79, 80, 83, 85, 89*, 90, 94, 123, 126, 129
Individual	3, 4, 5, 7, 9 - 10, 13, 14, 17, 22, 25, 26, 27*, 28, 31, 36*, 39, 40:12 - 17, 41, 42 - 43, 52*, 53, 54, 55, 56, 57, 59, 61, 64, 70, 71, 77, 86, 89*, 120, 139, 141, 142
Specialized Lament Psalms	
Penitential	6, 32*, 38, 51, 102, 130, 143
Imprecatory	35, 69, 83, 88, 109, 137, 140
Thanksgiving Psalms	
Community	65*, 67*, 75, 107, 124, 136*
Individual	18, 21, 30, 32*, 34, 40:1 - 11, 66:13 - 20, 92, 108*, 116, 118, 138
Specialized Thanksgiving Psalms	
Salvation History	8*, 105 - 106, 135, 136
Songs of Trust	11, 16, 23, 27*, 62, 63, 91, 121, 125, 131
Hymnic Psalms	
Hymn and Doxology	8*, 19: 1 - 6, 33, 66: 1 - 12, 67*, 95, 100, 103, 104, 111, 113, 114, 117, 145, 146, 147, 148, 149, 150
Liturgical Psalms (for Public Worship)	
Covenant Songs	50, 78, 81, 89*, 132
Royal / Enthronement	2, 18, 20, 21, 29, 45, 47, 72, 93, 95*, 96, 97, 98, 99, 101, 110, 144
Songs of Zion	46, 48, 76, 84, 87, 122
Temple Liturgies	15, 24, 68*, 82, 95*, 115, 134
Community Psalms	
Wisdom Psalms	1*, 36*, 37, 49, 73, 112, 127, 128, 133
Torah Poems	1*, 19: 7 - 14, 119

*These Psalms are difficult to classify because they could fit into more than one group or are mixed types.

The largest group of psalms is under the type of individual lament, which deal with prayers of supplication of a single person in unfortunate situations and trust in God in times of difficulty (Seybold, 2000). Many of Davidic psalms are in this category. Although there are a few psalms of community laments arising in national disaster, but most laments are individual. Hymns used in worship are descriptive praises of who God is. They comprise the second largest group of psalms and appear mostly in the second half of the book.

Thanksgiving could be used in general terms as in some contexts, but it is used typically for what God had done in Psalms, such as deliverance from the enemies or public celebration. Although thanksgiving is one type of forms listed in **Table 1**, thanksgiving and gratitude can also appear in laments.

Psalms 1 starts the whole book with a teaching of "two ways of life" as a general truth: the way of righteousness that leads to prosperity and the way of wickedness to ruin, suggesting this book ought to be studied as a guide to a blessed life. There are several psalms in the style of teaching, including Psalms 1, 37, 49, 78, and 112 (Mays, 1994). Therefore, there is rich wisdom in Psalms, and further it begins through the voice of wisdom.

Studying Proverbs and Psalms together enables us to discover the psalms that are similar to some proverbs. Some of the wisdom psalms are Psalms 1, 36, 37, 49, 73, 112, 127, 128, 133, as listed in **Table 1**, and moreover Psalms 1, 10, 12, 15, 19, 32, 34, 36, 37, 49, 50, 52, 53, 73, 78, 82, 91, 92, 94, 111, 112, 119, 127, 128, 139 are considered wisdom psalms in (Campbell, 2012).

Top 30 Mostly Used Words in Proverbs and Psalms

To gather the usage of various words in Proverbs and Psalms, we conduct a word count. The top 30 mostly used words in these two books are displayed in **Table 2**. The word "fear" appears 23 times in Proverbs and 64 times in Psalms. The ex-

pression of "fear of the Lord" appears in many books of the Bible, from Genesis, Exodus, Leviticus, Deuteronomy, to Revelation. To fear God is to know Him, to obey His commands with reverence and awe, to praise Him, to enjoy His blessings, and to care what God desires of us, but at the same time to fear His wrath or judgment. In short, it simply means to live a life with an awareness of the divine presence all the time, which should then be complemented by an awareness of the love of God.

Of the top 30 most frequently used words, there are 12 shared between these two books: fear, hand, heart, lord, man, men, mouth, righteous, righteousness, soul, wicked, word(s), illustrating the commonality of these two different books. On the other hand, the unique words in Proverbs are: wise, wisdom, understanding, evil, son, fool, knowledge, lips, life, poor, eyes, ways, house, instruction, woman, father, king, abomination, and in Psalms: God, praise, earth, people, mercy, David, enemies, made, strength, give, sing, make, salvation, Israel, chief, day, hear, deliver. These different key words in Proverbs and Psalms represent the distinctive focus of each book.

In order to appreciate how the word "fear" is used in different contexts, the verses containing the word "fear" are collected from Proverbs and Psalms respectively, from which word clouds images are generated (**Figure 1**). It seems that the word "fear" is combined with words, Lord, knowledge, wisdom, instruction, evil, wicked, and depart in Proverbs, and with Lord, God, praise, mercy, great, and earth in Psalms, revealing characteristic emphasis on the phrase "fear of the Lord" in these two books.

Methods

Topic Model

Different from keyword searching in traditional text data mining, topic models have been developed recently to discover

Table 2.

Top 30 mostly used words and their counts in Proverbs and Psalms.

Proverbs				Psalms			
Word	Count	Word	Count	Word	Count	Word	Count
Man	171	Life	38	Lord	793	Give	72
Wicked	89	Poor	38	God	440	Righteousness	71
Lord	87	Soul	38	Praise	160	Sing	70
Heart	81	Men	34	Earth	141	Make	69
Wise	66	Eyes	30	Soul	132	Fear	64
Righteous	54	Ways	29	People	130	Mouth	64
Understanding	54	House	28	Heart	125	Salvation	63
Wisdom	54	Hand	25	Man	101	Israel	62
Mouth	52	Instruction	25	Mercy	100	Men	62
Evil	50	Woman	25	Hand	99	Righteous	60
Son	45	Father	23	Wicked	90	Word	60
Fool	42	Fear	23	David	88	Chief	58
Knowledge	42	King	22	Enemies	81	Day	58
Lips	42	Righteousness	21	Made	76	Hear	57
Words	40	Abomination	20	Strength	74	Deliver	55

Figure 1.
Word clouds from the verses containing the word "fear" in Psalms (left) and Proverbs (right).

the hidden thematic structure in a collection of documents (Steyvers & Griffiths, 2007; Griffiths, Steyvers, & Tenenbaum, 2007; Griffiths & Steyvers, 2004). They employ statistical analysis on the words (observed variables) in the documents to infer their themes (hidden variables). The simplest topic model is latent Dirichlet allocation (LDA) (Blei, Ng, & Jordan, 2003), which builds on the assumption that a single article or document generally contains multiple topics and the topic distribution is a Dirichlet prior. In topic models, a topic is defined as a distribution over words. For instance, the school topic may contain words such as teachers, students, homework, curriculum, and GPA. The mixture of topics could be found through the discovering of the hidden thematic structure in the document, in which words are observed whereas the topic distribution needs to be inferred from them. The computational task of LDA is therefore to compute the posterior distribution, the conditional distribution of the hidden variables given the documents.

In this study, the topic proportion identified by LDA from each document is used to define a correlation to measure the similarity between two documents. The software we use in this report is MALLET (McCallum, 2002) that has an implementation of LDA. The standard Pearson correlation formula for two vectors in n-dimensional Euclidean space is defined as follows:

$$r_1 = \frac{\sum_{i=1}^{n}\left(\bar{X}-X_i\right)\left(\bar{Y}-Y_i\right)}{(n-1)S_X S_Y}$$

where $X = \left(X_1, X_2, X_3, \cdots, X_n\right)$, $Y = \left(Y_1, Y_2, Y_3, \cdots, Y_n\right)$. $\bar{X}$ and $\bar{Y}$ represent the average of the two vectors, S_X and S_Y stand for the standard deviation of the two vectors.

In the calculation of the correlation between two topic distributions from two documents, the X_i represents the probability of topic i, and the sum of all these X_i is one. Since the contributions from different X_i to the correlation are not equal, we assign a weight to each expression $\left(\bar{X}-X_i\right)$ and $\left(\bar{Y}-Y_i\right)$ as shown below:

$$r_2 = \frac{\sum_{i=1}^{n} X_i\left(\bar{X}-X_i\right)\left(\bar{Y}-Y_i\right)Y_i}{(n-1)S_X S_Y}$$

which suggests that the higher the probability of a topic is, the greater its contribution to the correlation. Throughout the current study, we use this weighted Pearson correlation r_2 for comparing two topic distributions.

Affinity Propagation

At present, common clustering algorithms such as K-means use data to find centers of clusters. The novelty of affinity propagation clustering algorithm (Frey & Dueck, 2007) lies in the search of exemplars from data to represent clusters through message passing. The data points in a cluster connect to the exemplar that best represents it.

Affinity propagation (AP) was designed to searches for a set of exemplars from n data points so the minima of the following energy function could be achieved

$$E(e) = -\sum_{i=1}^{n} s(i, e_i) \qquad s(i, e_i) \le 0$$

where e_i represents the exemplar of data point i, and $s(i, e_i)$ is the similarity between data point i and its exemplar e_i. In general, finding an optimal set of exemplars from data is a computationally intensive endeavor. However, the identification of these exemplars in AP was accomplished with an efficient algorithm that employs real-valued messages exchanged among data points. There were two types of messages, responsibility and availability. The responsibility $r(i, k)$, sent from data point i to candidate exemplar point k, reflects the suitability for point k to be the exemplar for point i. The availability $a(i, k)$, sent from candidate exemplar point k to point i, represents the propensity for point i to choose point k as its exemplar. In this study, the similarity used as input to AP is the Hamming distance between two sequences. At the beginning of affinity propagation the availability matrix is set to zero, and then repeats the following iteration until a stopping criterion is satisfied:

$$r(i, k) \leftarrow s(i, k) - \max_{k' s.t. k' \ne k}\left\{a(i, k') + s(i, k')\right\}$$

$$a(i, k) \leftarrow \min\left\{0, r(k, k) + \sum_{i'} s.t. i' \ne i, k \max\left\{0, r(i', k)\right\}\right\}$$

$$a(k, k) \leftarrow \sum_{i' s.t. i' \ne k} \max\left\{0, r(i', k)\right\}$$

At each iteration, the messages indicate the current affinity between data points and their exemplars.

In the present study, this clustering algorithm is used to group chapters in Proverbs and Psalms, where each chapter is represented by its topic distribution as a vector in Euclidean space.

Results and Discussion

The topic model and affinity propagation clustering algorithm are utilized to group the chapters of Proverbs (**Figure 2** and **Table 3**) and Psalms (**Figure 3** and **Table 4**) respectively, thus illustrating the internal order of these chapters by content. The chapters in each cluster are presented in **Tables 3** and **4**, while the closeness among the clusters is shown in **Figures 2** and **3**. To refine the clustering of the chapters and to understand the connection of these two books, the correlations of the chapters and verses between these books are analyzed (**Figures 4** and **5**). We also compare our computational results with those by Biblical scholars.

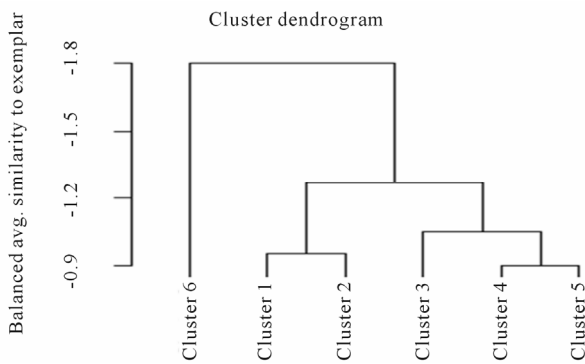## Clustering Chapters of Proverbs by Content

The usual division of the chapters in Proverbs is based on authorship and in sequence. However, our clustering algorithm does that based on content, and thus with nonlinearity (**Figure 2** and **Table 3**). **Figure 3** shows that there are three large groups of clusters: clusters 1 and 2; clusters 3, 4, 5; and cluster 6. The proverbs of Solomon are grouped together in clusters 3, 4 and 5 except chapters 20, 24, and 30, and those of non-Solomon in clusters 1, 2, and 7. Chapter 24 contains part of the 30 wise sayings, and the majority of the verses in chapter 20 are part of Solomon's proverbs and a small fraction of chapter 20 is part of the 30 wise sayings. Finally chapter 30 is the wisdom of

Table 3.
Chapters of Proverbs in different clusters and their exemplars.

Cluster Number	Exemplar Chapter Number	Chapters in Each Cluster
1	2	1 2 3 8 9
2	5	4 5 6 7 23
3	17	17 18 26 27
4	20	16 20 22 24 25 30
5	21	10 11 12 13 14 15 19 21 28 29
6	31	31

Table 4.
Chapters of Psalms in different clusters and their exemplars.

Cluster Number	Exemplar Chapter Number	Chapters in Each Cluster	Common Genre
1	8	2 8 50 72 104 144 148	Hymns
2	16	3 5 16 18 20 21 22 23 26 27 63 91 110 116 120 124 131 132 139 142	Individual laments, songs of trust, thanksgivings
3	36	1 10 11 12 15 19 32 34 36 39 49 52 58 64 73 92 94 101 112 125 140	Wisdom
4	37	37	Wisdom
5	40	25 35 40 69 70 71 109	Individual laments, imprecatory laments
6	43	42 43 57 60 62 68 108	Individual laments, Songs of trust
7	53	14 53 81 83 114	Individual and community laments
8	54	4 6 7 13 17 31 38 41 51 54 55 56 59 61 88 141 143	Individual, penitential laments
9	74	44 74 77 79 80	Community laments
10	87	46 48 65 76 84 87 137	Songs of Zion
11	90	90 102	Special laments
12	93	24 29 45 89 93 97 99	Royal/enthronement
13	98	47 96 98 149	Royal/enthronement
14	105	78 105 106	Salvation history
15	107	107	Community thanksgiving
16	117	9 30 33 66 67 75 82 85 86 95 100 111 113 117 123 138 145 146 147 150	Hymns, thanksgivings
17	119	119	Torah Poems
18	134	28 103 115 118 121 122 126 127 128 129 130 133 134 135	Songs of ascents
19	136	136	Salvation history

Figure 2.
Cluster dendrogram for chapters of Proverbs.
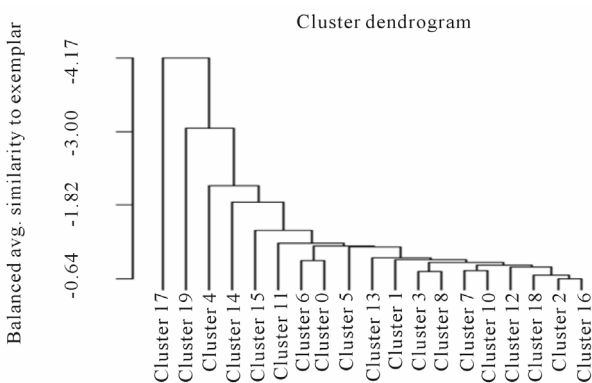
Figure 3.
Cluster dendrogram for chapters of Psalms.

Agur.

Structurally the first nine chapters are all long wisdom dialogs between a father and his sons. Our clustering algorithm groups chapters 1 through 9 into clusters 1 and 2, where cluster 2 selects the chapters that focus on lady adulteress and cluster 1 chooses the chapters on lady wisdom. However, a few verses in chapter 2 of cluster 1 also deal with the adulteress, and chapter 4 on lady wisdom is part of cluster 2. Cluster 3 contains verses about a fool, as a common feature, in all the four chapters of this cluster.

It is amazing to have the same subject of lady adulteress being brought up so frequently in chapters 5, 6, 7. It is suggested in (Stuart, 1852) that one single compiler brought these various writings of several authors into one piece. The same reasoning that leads to a diversity of authorship could also be applied to the repetition of wisdom observed in all the chapters in cluster 1.

The chapters in cluster 3 share some verses, such as 18:8 and 26:22, 17:3 and 27:21, 17:17 and 27:10, which gives another support of the same authorship of the two collections of Solomon's proverbs and two different compilers of these two collections. There are 6 times of using the word "fool" and one time for "folly" and one time for "foolish" in chapter 17. In chapter 18, there are 3 times of "fool" and one time of "folly". In chapter 26, 9 times of "fool", 3 times of "folly". In chapter 27, 2 times of "fool". There are similar verses in chapters 17 and 18: 17:15 and 18:5, 17:22 and 18:14.

Cluster 4 contains the 30 wise sayings of the wise (verses 22:17 - 24:22) that have the signature phrase "my son" as in chapters 1 - 9 and the further sayings of the wise (24:23 - 34)

that lack the signature phrase "my son", which exhibit a composing style different that of Solomon (say, sentences are completed in one verse are the exception, and "my son" is not prefixed to a paragraph). However, they were grouped with several chapters of Solomon because of their shared content. The sayings of King Lemuel (chapter 30), as part of cluster 4, are different from those of Solomon not only in style but also in contents. The phrase "my son" in Proverbs suggests that some instructions are particular relevant to a young person.

Cluster 4 has chapters containing several verses on kings. It was enlightening to see three such verses in chapter 30, which are: A servant who become a king, a fool who is full of food (verse 22); Locusts have no king, yet they advance together in ranks (verse 27); A strutting rooster, a he-goat, and a king with his army around him (verse 31).

Cluster 5 contains chapters 10 - 15 that deal with wise living through instructive contrasts, thus this cluster also contains chapters 19, 21, 28, and 29. Chapter 28 starts with verse: the wicked man flees though no one pursues, but the righteous are as bold as a lion, and ends with verse: when the wicked rise to power, people go into hiding; but when the wicked perish, the righteous thrive. Similarly chapter 29 starts and ends with a contrast between the righteous and the wicked.

Apart from the common features of the chapters in cluster 5, chapters 28 and 29 have their unique verses that are mainly for prospective kings/rulers to have kingly wisdom, a feature shared in cluster 4.

Chapters 10 through 15 exhibit strong patterns of comparison, but right in chapter 16 this pattern stops. Our clustering algorithm detects this shift in chapter 16 and includes it in another cluster.

Chapter 23 is included in cluster 2 with chapters 4, 5,6, and 7, because it contains the signature address "my son" as in many verses in the first collection and two verses that deal with prostitutes that is the major theme in chapters 5, 6, and 7. For this reason, the exemplar of this cluster is chapter 5. Therefore, it is surprising to have chapter 4 in cluster 2, rather than in cluster 1. Two verses in chapter 4 are similar to those in chapters 5 and 7: 4:4 and 7:2, 4:20 and 5:1, but more verses in this chapter are similar to those in chapters 1, 2, and 3 than in chapters 5, 6, and 7.

The verses in the start of Chapter 31 teach how to be a wise king, but the majority of this chapter describes the noble character of a wife, which is made of verses each starting a successive letter in the alphabet of Hebrew. Therefore, it makes sense to be grouped as a different cluster from cluster 4.

Clustering Chapters of Psalms by Content

The understanding of a psalm could be enhanced by classifying it according to its content such as historical, or wisdom. From **Figure 3** and **Table 4**, we can see that the grouping of the chapters in Psalms is largely in line with those in **Table 1**. **Figure 4** shows the overall grouping of different clusters, which, in particular, highlights the uniqueness of cluster 17 that contains Psalms 119 alone. Due the multi-faceted nature of psalms, some of them exhibit several types, such as lament and thanksgiving, within a single psalm. As a result, our clustering algorithm groups these in one cluster, as shown in clusters 2, 5, 6, 7, and 8. For example, Psalm 40 is in cluster 5 of individual and imprecatory laments. It has a thanksgiving feature in verses 1 - 10, but a lament flavor in verses 11 - 17.

At the same time, our clustering algorithm is able to detect the similarity between psalms. From the text, it is clear that Psalms 43 is a continuation of Psalms 42, also supported by our clustering. Other pairs of similar psalms are recognized by our algorithm as well: Psalms 53 and 14; Psalms 70 and 40: 13 - 17; Psalms 108 and 57: 7 - 11; and Psalms 108 and 60: 5 - 12, as also observed in (Driver, 1891).

Psalms 18, 20, and 21 are considered as royal/enthronement psalms in **Table 1**, but our clustering algorithm groups them into songs of trust, which is more appropriate according to their contents.

The songs of ascents (Psalms 120 - 134) were sung by pilgrims going to Jerusalem for annual festivals. Psalm 125 is a song of ascents, but our clustering considers it as a wisdom psalm, consistent with its content. For the same reason, songs of ascents 120, 124, 131, and 132 are grouped in cluster 2 of individual laments, songs of trust, and thanksgivings. Song of ascents 123 is in cluster 16 of hymns and thanksgivings. Interestingly Psalms 28, 103, 115, 118, and 135 are grouped with many songs of ascents in cluster 17 because of their contents. Taken together, the songs of ascents have different genres, but our clustering algorithm still detects those that are similar, and further identifies a few psalms such as 28, 103, 115, 118, 135 that are not part of songs of ascents but are still similar.

Psalm 73 is a wisdom psalm that sees the wicked prospering and the righteous suffering. It opens book three of Psalms and starts a new era after the kingdoms of David and Solomon. Psalm 11 is considered as a song of trust in **Table 1**, here it is treated as a wisdom psalm. Our correlation analysis in section 3.3 also connects this psalm with chapter 11 of Proverbs.

Hymns could be divided into the "imperative" and "participial" subgroups. The former is characterized with exhortation ("sing!", "play!", and "praise!", Psalms 100 and 136) and the latter typically contains a chain of participles ("wrapping yourself in light… stretching out the heavens… making… riding…", Psalms 104). These two subgroups have different functions. The imperative are suitable for public worship, while the participial focus on objective representation of theological contexts (Seybold, 2000). Our clustering results correctly identify these two subgroups, cluster 1 for the participial and cluster 16 for the imperative.

Psalm 136 is unique among the psalms because of its recurrent refrain, "His love endures forever". Each verse recites a great redemptive act of God, followed by a refrain with a total of 26 times. Because of its unique style, this psalm is itself a cluster (cluster 19). This distinct feature is also automatically detected by our pairwise correlation analysis conducted in Section 3.3 and displayed in **Figure 5**.

Correlation between Psalms and Proverbs

The purpose of this section is to find the similar chapters and verses between Proverbs and Psalms. To process the verses, we divide them in groups of three. The first number in **Figure 5** means the chapter number, and the second is the group number. The outcome of pairwise correlation analysis of chapters and verses between these two books is contained in **Figures 4** and **5**.

As seen from **Figure 4**, Proverbs 11 is highly correlated with Psalms 11, 37, and 112, and they share several key words. Thereby some verses in Psalms 11, 37, and 112 display the characteristics of proverbs. Two such verses in Psalms 11 are:
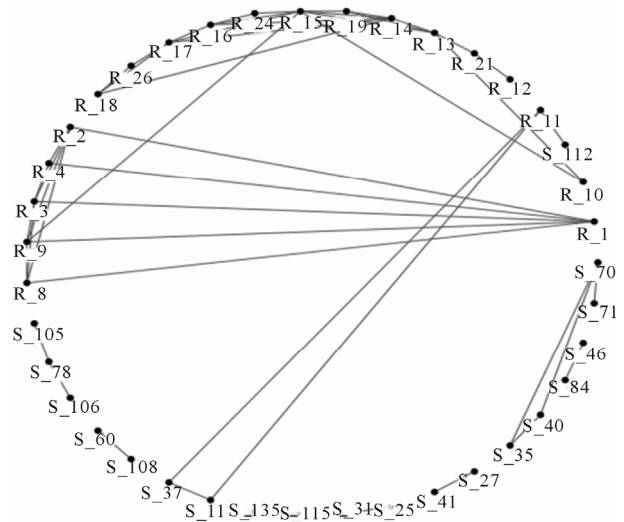
Figure 4.
Top 50 pairs of highly correlated chapters in Proverbs and Psalms, where letter "R" represents Proverbs and "S" for Psalms.
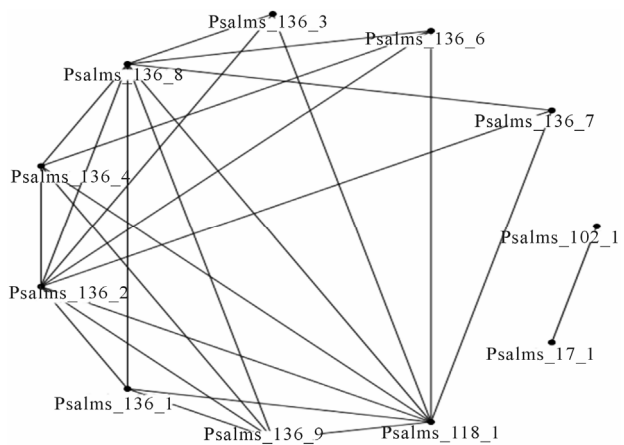
Figure 5.
Top 25 pairs of highly correlated verse segments in Proverbs and Psalms, but none is selected from Proverbs, where the first number represents the chapter number and the second is the segment number.

The Lord examines the righteous, but the wicked and those who love violence his soul hates (verse 5); For the Lord is righteous, he loves justice upright men will see his face (verse 7). In Psalms 37: Better the little that the righteous have than the wealth of many wicked (verse 16); for the power of the wicked will be broken, but the Lord upholds the righteous (verse 17). In Psalms 112: Blessed is the man who fears the Lord, who finds great delight in his commands (verse 1); Even in the darkness light dawns for the upright, for the gracious and compassionate and righteous man (verse 4); The wicked man will see and be vexed, he will gnash his teeth and waste away; the longings of the wicked will come to nothing (verse 10). Importantly, the fear of the Lord is part of Psalms 112:1.

Some verses in Psalms 11 have similar verses in Proverbs such as Psalms 11:4 and Proverbs 15:3. Although Psalm 11 is not listed as a wisdom psalm in **Table 1**, our correlation analysis connects it with Proverbs 11, and the text of this psalm reveals its wisdom nature.

There are many verses in Psalms that contain the concept of "fear of the Lord". Here we list some of them: 2:11, 19:9, 22:23, 25:12, 25:14, 33:8, 33:18, 34:7, 34:9, 40:3, 86:11, 96:4, 102:15, 103:13, 103:17, 111:10, 112:1, 115:11, 128:1, 135:20, and 147:11.

Comparing the two books of Proverbs and Psalms (**Figures 4** and **5**), Proverbs has a higher correlation between its chapters, while Psalms has a higher correlation between its verse segments. It was found in (Anderson, 1994) that no two psalms of the same textual material are found in any single book, but in two different books from the five books (Anderson, 1994). Furthermore here we identify a few similar chapters in book 1: Psalms 11 and 37; 25 and 31; 27 and 41 (**Figure 4**).

Psalms 11 and 37 in book 1 are highly correlated, and some similar verses between them are: Psalms 11:2 (For look, the wicked bend their bows; they set their arrows against the strings to shoot from the shadows at the upright in heart) and 37:14 (The wicked draw the sword and bend the bow to bring down the poor and needy, to slay those whose ways are upright); 11:5 (The Lord examines the righteous, but the wicked and those who love violence his soul hates) and 37:17 (for the power of the wicked will be broken, but the Lord upholds the righteous); 11:7 (For the Lord is righteous, he loves justice; upright men will see his face) and 37:28 (For the Lord loves the just and will not forsake his faithful ones).

The highly correlated verse segments (**Figure 5**) reveal that Psalms 136 has many similar verses within itself. The examination of the text of this chapter shows the repeated verses of the same sentence: His love endures forever. Although the two verse segments in Psalms 17 and 102 selected in Figure 5 do not contain same or similar verses, but do share many key words.

The clusters in **Tables 3** and **4** indicate an overall similarity of the chapters within the same cluster. The connected groups of highly correlated chapters in **Figure 5** are consistent with the clusters in **Tables 3** and **4**. For example, Psalms 35, 40, 70 and 71 forms a group in Figure 5, and they are also in the same cluster 5 in **Table 4**; Psalms 78, 105, and 106 formed a group in **Figure 5** are in the same cluster 14 in **Table 4**.

However, our pairwise correlation analysis of these chapters (**Figures 4** and **5**) provides more details in this regard. In Proverbs, chapters 17, 18, and 26 in cluster 3 in **Table 3** are mutually highly correlated, but chapter 27 in the same cluster is not strongly associated with them. Chapters 16 and 24 in cluster 4 are mutually highly correlated, but chapters 20, 22, 24, 25, and 30 in the same cluster are not. Chapters 10, 12, 13, 14, 15, 19, and 21 are mutually highly correlated, but chapters 11, 28 and 29 in the same cluster are not.

The connected group of chapters 1, 2, 3, 4, 8, and 9 of Proverbs in **Figure 4** suggests their mutual similarity, which corrects the inclusion of chapter 4 in cluster 2 in **Table 3**. In Proverbs, chapter 15 is highly correlated with chapters 9 and 10, which could not be seen from the clustering in **Table 3**. To summarize, the connected groups of chapters in **Figures 4** and **5** refine and detail the clusters in **Tables 3** and **4**.

Conclusion and Discussion

The aim of this study is to apply a data-driven approach to finding the similarity of the chapters in each of Proverbs and Psalms, and the similarity of chapters and verses between these two books, due to their shared wisdom emphasis. Meanwhile, we compare our computational results with those by Biblical scholars.

The word "wisdom" usually refers to knowledge, intelligence, education, and experience, however, the meaning of this word in Proverbs includes and surpasses them (Lennox, 1998). Wisdom in ancient Israel and early Judaism includes at least six critical components: knowledge, imagination, discipline, piety, order, and moral instructions (Perdue, Mays, & Miller, 2000). The book of Proverbs is designed to teach wisdom, instruction, justice, and judgment. It presents the main theme that true wisdom is a natural consequence of fearing God. The wisdom of a person is defined in terms of his relationship with God, rather than his academic knowledge and life experience. Fear of the Lord is the beginning of wisdom, which is the proverb of proverbs, and the foundation of all other proverbs.

The notion of fearing God was expressed even in Genesis (Genesis 22:12 and 42:18). Three wisdom books of the Bible, Job, Ecclesiastes, and Proverbs, contain similar verses like this one: The fear of the Lord—that is wisdom, and to shun evil is understanding (Job 28:28), Fear God and keep his commandments, for this is the whole duty of man (Ecclesiastes 12:13), the fear of the Lord is the beginning of knowledge, but fools despise wisdom and discipline (Proverbs 1:7), the fear of the Lord is the beginning of wisdom, and knowledge of the Holy One is understanding (Proverbs 9:10). All these verses highlight that God is the only source of wisdom.

Psalms are a collection of diverse poems, expressed as prayers and praises. It starts with a wisdom psalm, Psalm 1, as an introduction to the entire book. Proverbs leads indirectly to God through wisdom, whereas Psalms points directly to God through worship. In other words, Psalms deals with the relationship with God (in vertical direction), whereas Proverbs is about the association to other people (in horizontal direction). The book of Psalms is perhaps the most read book among all the books in the OT. It has the longest chapter of the Bible, Psalm 119, and the shortest chapter of the Bible, Psalm 117.

The poetic parallelism is widely employed in Psalms. Three types of parallelisms are commonly observed in this book: synonymous—the thought in the first line is repeated in the second line in different but equivalent words; antithetic—the second line gives a thought in the opposite of the first line; and synthetic—the second line completes the thought of the first without parallel meaning.

There is a great variation of character among the 150 psalms. Oscillating between lament and praise, many psalms, like Psalms 13 and 22, may begin with lament or prayer and then change into thanksgiving and praise, which makes it hard to classify them by content. Psalms tend to be grouped together with the same heading or with the same opening lines such as Psalms 3 - 6; 19 - 24, and more (Seybold, 2000). Several related psalms appear together such as Psalms 5 - 7, 54 - 57, 61 - 64, 69 - 71, 140 - 143 (laments of the individuals) and Psalms 103 - 105, 134 - 136, 145 - 150 (hymns). However, in general no clear internal relationship between neighboring psalms can be observed (Gunkel, 1998). Beyond the 150 psalms recorded in the book of Psalms, there are other psalms in the OT, such as Moses' "Song of the sea" (Exodus 15), the thanksgiving songs of Hanna (1 Sam 2), of David (2 Sam 22), of Jonah (Jonah 2), and of Hezekiah (Isa 38).

Scholars have proposed various ways of categorizing the psalms, i.e., grouping them according to their content and literary forms. Our work here is to apply the topic model and clus-

tering method to group the psalms together by content, and then see how well the outcome of this unsupervised learning matches their literary forms. Our findings demonstrate and reveal their internal arrangement in this diverse material. We use the same approach to study of Proverbs. The last task of our investigation is to find similar chapters and verses between Proverbs and Psalms, allowing one book to shed light onto the other to enhance the understanding. Our computational results are consistent with those by Biblical scholars, but have made several interesting new discoveries along the way.

Acknowledgements

REFERENCES

Anderson, R. D. (1994). The division and order of the psalms. *Westminster Theological Journal, 56,* 219-241.

Blei, D., Ng, A., & Jordan, M. (2003) Latent Dirichlet allocation. *Journal of Machine Learning Research, 3,* 993-1022.

Bratcher, D. (2011). Patterns for life: Structure, genre, and theology in psalms. URL (last checked 26 June 2012).
http://www.crivoice.org/psalmgenre.html

Campbell, L. (2012). Psalms with Lee Campbell. URL (last checked 26 June 2012). http://www.xenos.org/classes/psalms/index.htm

Collins, J. J. (2007). *A short introduction to the Hebrew Bible* (Abridged edition). Minneapolis, MN: Fortress Press.

Constable, T. L. (2012) Notes on Psalms, Sonic Light. URL (last checked 26 June 2012).
http://www.soniclight.com/constable/notes/pdf/psalms.pdf

Coulter, P. B. (2011). Psalms and wisdom literature: An introduction. URL (last checked 26 June 2012). www.paulcoulter.net

Dennis, & Kintsch. W. (Eds.), *Latent semantic analysis: A road to meaning.* Hillsdale, NJ: Laurence Erlbaum.

Driver, S. R. (1891). *An introduction to the literature of the Old Testament.* New York: C. Scribner's Sons.

Frey, B. J., & Dueck, D. (2007). Clustering by passing messages between data points. *Science, 315,* 972-976.

doi:10.1126/science.1136800

Griffiths, T., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences, 101,* 5228-5235.
doi:10.1073/pnas.0307752101

Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. T. (2007). Topics in semantic representation. *Psychological Review, 114,* 211-244.
doi:10.1037/0033-295X.114.2.211

Gunkel, H., & Horner, T. M. (1967). *The Psalms: A form-critical introduction.* Philadelphia: Fortress Press.

Gunkel, H. (1998). *Introduction to Psalms: The genres of the religious lyric of Israel (Mercer Library of Biblical Studies).* Macon: Mercer University Press.

Lennox, S. J. (1998). *Proverbs: A commentary for Bible students* (1st ed.). Indianapolis: Wesleyan Publishing House.

Longman, T. (2009). *How to read the Psalms (How to read series how to read).* Westmont, IL: Intervarsity Press.

Mays, J. L. (1994). *Psalms: Interpretation: A Bible commentary for teaching and preaching (Interpretation: A Bible commentary for teaching & preaching)* (1st ed.) Louisville: Westminster John Knox Press.

McCallum, A. K. (2002). MALLET: A machine learning for language toolkit. URL (last checked 26 June 2012). http://mallet.cs.umass.edu

Mowinckel, S. (2004). *The Psalms in Israel's Worship (Biblical Resource Series).* Grand Rapids: Wm. B. Eerdmans Publishing Company.

Perdue, L. G., Mays, J. L., & Miller, P. D. (2000). *Proverbs (Interpretation: A Bible commentary for teaching & preaching).* Louisville: Westminster John Knox Press.

Pippert, W. G. (2003). *Words from the wise.* Maitland, FL: Xulon Press.

Seybold, K. (2000). *Introducing the Psalms.* Edinburgh: T&T Clark Int'l.

Steyvers, M., & Griffiths, T. (2007). *Probabilistic topic models.* In T. Landauer, D. McNamara, S. Dennis, & W. Kintsch (Eds.), *Latent semantic analysis: A road to meaning.* Hillsdale, NJ: Laurence Erlbaum

Stuart, M. (1852). *A commentary on the book of Proverbs.* Rochester: M.W. Dodd.

Yuasa, K. (1891). *A Classification of the Solomonic Proverbs, the old and new Testament student* (pp. 147-153). Chicago, IL: The University of Chicago Press.