Mining of Social Networks from Classic Books and Visualization

SeungJin Lim
Computer Science
Merrimack College
North Andover, MA 01845, U.S.A.
lims@merrimack.edu

Journal of Digital Information Management

ABSTRACT: *While modern online social networks offer unprecedented amount of data and opportunities for mining, we acknowledge that other forms of social networks have been around throughout human history. In particular, we recognize the classic books as valuable sources for the analysis of interpersonal interactions as they have had positive influence on human mind and continue to enlighten the human race.*

In this paper, we propose a formal framework for the mining of interpersonal interactions from classic books. Our approach employs a heuristic over a traditional named-entity recognition for extracting people names. We demonstrate the effectiveness of our approach by presenting the encouraging result from a network analysis conducted over the persons appearing in the book of Genesis. The result aids the understanding of the persons featured in the book and their interactions. We also present the merits and challenges of visualization techniques pertaining to our work.

Subject Categories and Descriptors
[I.6.8 Types of Simulation]: Visual; [H.2.8 Database Applications]: Data Mining; [H.5.3 Group and Organization Interfaces]

General Terms: Social Networks, Network mining, Text Mining, Information Visualization

Keywords: Interpersonal Interaction Network, Social Network Mining, Text Mining, Visualization

Received: 1 October 2017, Revised 15 November 2017, Accepted 27 November 2017

1. Introduction

Social networks have drawn increased attention from nearly all branches of science and humanity studies in recent decades. This increase in popularity has been largely propelled by the unprecedented scale of online social networks. The notion of social networks has been broadly applied to the relationships between any entities, both people and non-people, as exemplified by electrical power grids [1] and the Internet packet routing networks. One way researchers from computer science are successfully contributing to the advance of this topic is the development of efficient link analysis techniques thereby enabling automatic, large-scale analysis.

In this paper, we present a methodology for analyzing the social networks found in classic books in an attempt to understand the interaction between the people appearing in classic books. This paper is an extension of work originally presented in the proceedings of the 12th International Conference on Digital Information Management (ICDIM 2017). It includes a comparison of our heuristic to a state-of-the-art named-entity recognition tool.

In his extensive survey on complex networks [2], Newman recognized the analysis of personal connections between people is one of the mainstream activities in social network analysis. In this type of analysis, co-appearance of individuals in the same context forms the basis of the network between the individuals as commonly done on Web pages [3] and in newspaper articles [4], [5]. However we are not aware of a study on interpersonal interaction

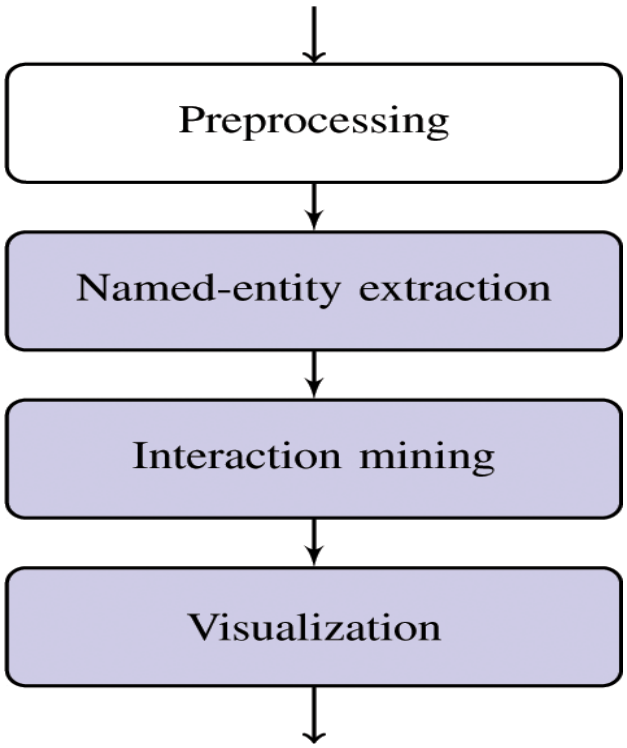


Figure 1. The three major steps in our analysis of the interpersonal interaction among the individuals appearing in the book of Genesis

networks from ancient literature.

In this paper, we propose a formal framework for the analysis of interpersonal interaction networks from classic books. In the proposed approach, we view classic books as the event logs of the people who are referenced in them. This notion of event logs is agreeable especially when the books are historical or biographical books. We demonstrate the effectiveness of this approach by presenting the result from a network analysis conducted over the people appearing in the book of Genesis. The result from the analysis is encouraging and provides additional insight to the understanding of the individuals and the interactions between them. The experimental result also demonstrates the merits and challenges of the visualization techniques used.

Our project is conducted in four stages as depicted in Figure 1. First, we transform the source text to event logs, using the proposed method, by dividing the book into the textual units suitable to capture interpersonal interaction events. Punctuation symbols and stop words are removed from the event logs at this stage. Second, the names of the people mentioned in each event are extracted using a light-weight yet effective string matching, rather than using a heavy-duty natural language processing tool or a complete lookup table of names. Third, the interactions between the names which appear in the same event are identified, and the statistical significance of each interaction is quantified. An analysis of the discovered patterns follow. Fourth, the result is analyzed and presented visually using 2D/3D visualization tools that are developed in-house for this purpose. The details of these steps are discussed in the Methods section below.

The rest of this paper is organized as follows: Selected topics related to our work are summarized briefly in Section 2. The details of our approach, the data used, and the steps of our experiment are discussed in Section 3. We then conclude our presentation with a remark in Section 4.

2. Related Works

An extensive review of complex networks such as the Internet, social networks and biological networks can be found in [2]. Properties of networks, the types of real-world networks, random graphs, and the small-world model [1] are also discussed in [2]. In addition to the social network analysis performed in the past such as studies of the patterns of friendships between individuals [6], [7], business relationships between companies, intermarriages between families and the patterns of human sexual contacts, more modern analyses [8] are introduced in the same paper.

Named-entity recognition (NER) aims at the discovery and categorization of named entities in text such as the names of persons and locations. NER offers a valuable advantage to the early stage of interpersonal interaction networks in which the recognition of the names of persons is essential. NER approaches range from statistical approaches over manually tagged data to supervised approaches over a large amount of training data. Conditional Random Fields (CRF) [9] is one of the most competitive NER algorithms. The authors of [10] reported the performance of 91.21% F1 measure for NER without requiring feature engineering or data processing. They employed CRF [9], convolutional neural networks, and bidirectional long-short term memory in their approach.

The role of visualization is generally complementary to text based analysis or information presentation. Visualization is also indispensable in some cases, especially when the amount of information is large. An effective visualization provides the overview of the data first, then a way to examine details on demand. When it comes to 2D or 3D visualization, the layout of visual information becomes one of the most critical design issues, if not the most critical.

One layout option is a free-form, graph layout. This technique has been used in math and science for a long time. Examples include the map of the Internet, artificial neural networks, and various forms of Web graphs. In this layout, graphs are characterized by the geometric shapes of the nodes and the edges decorated with various attributes, such as weight and direction, while node location on the screen is irrelevant to the information carried by the graph. Hence, nodes can be placed anywhere in the given space. If node locations become ever specific in

the graph, it is mainly for a cosmetic reason. For example, the location of a node in the Podevsef layout shown in Figure 2c is chosen mainly for the maximum visual comfort.

On the other hand, node layout can be more controlled. Figure 2 shows a few layout options in 2D. The objects may be arranged either along the circumference of a circle (Figure 2a), or on a single linear axis, either horizontal or vertical, (Figure 2b), or anywhere in the given plane (Figure 2c). In fact, the three visualization layouts shown in the figure carry the exactly same set of relationships between the same set of objects.

Another commonly-used layout in the literature is to arrange objects along both the horizontal and the vertical axes [12]. The underlying graphs of this visualization layout should maintain not only the identifiers of the nodes and the edges but also the geometric shapes and locations of the nodes as their attributes.

3. Methods

In this section, we present our formal framework for analyzing interpersonal interaction networks from classic books in general. For an experiment of the formal approach, we used the book of Genesis. We begin with a description of the data set. We then present the formal definitions of event, interaction, and significance of interaction which serve as the foundation for our network analysis.

3.1 Data

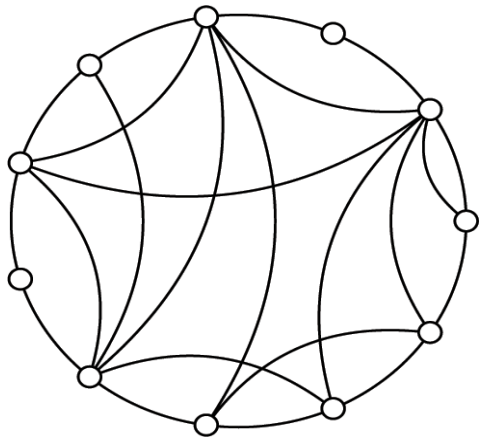
We selected the book of Genesis contained in the Bible as the data set for three reasons. First, classic literature offers to us living today the timeless wisdom that has been accumulated through the ages. For example, Bernholc noticed the importance family plays in Greek culture and concluded, "*Even today family remains as vital an aspect in Greek culture as it was in ancient times*." [13]. Second, the Bible is commonly listed as one of the most influential books ever written [14] in the history of humankind, along with other classic books from ancient Greece, Rome and China. It is also listed as the world's best-selling and most widely distributed book1. Third, the book of Genesis contains the accounts of about 400 individuals. They capture exciting, dynamically-changing and panoramic accounts of joy, hope, courage, despair and pain that can be experienced by every one of us. The analysis of interactions of these individuals may present interesting patterns to us.

3.2 Formal Approach

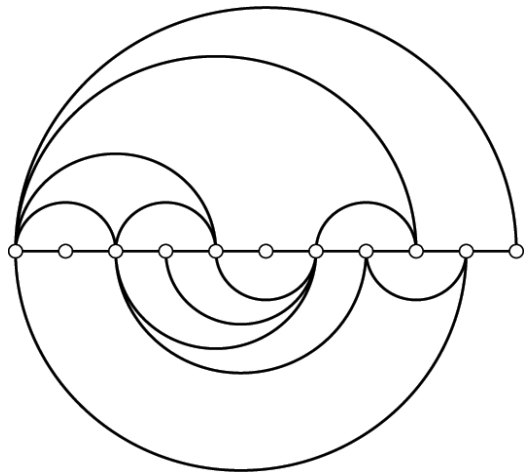
The concepts that are defined in this section form the foundation of our framework.

Definition 1 (Event): An event is a textual unit from the source text which contains one or more actors.

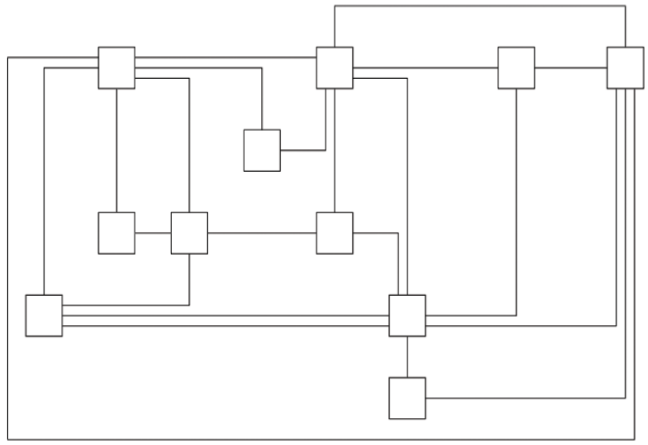
Having the concept of event defined, we transform the source text to a sequence of events where persons may interact with one another.



(a) Circular layout



(b) Kandinski layout



(c) Podevsef layout

Figure 2. Circular, Kandinski and Podevsef layouts [11]. Nodes are arranged along the circumference of the base circle (a) or along a single linear axis (b), or they can appear anywhere in the given visual area (c)

Notice that an event can be as narrow as a sentence or as wide as the entire volume. If events are narrow, the locality of interactions tends to improve. In other words, interactions become more specific and personal, which is desirable. On the other hand, the probability of having no interpersonal interactions in an event may increase. Interactions may be understood out of context.

In contrast, if events are wider, the number of interacting individuals tend to increase, but the locality of interactions may be compromised.

The book is hierarchically organized into chapters first and then into verses. We consider this division as the product of collective wisdom over time as it has been used for centuries. Between chapters and verses, we chose verses as textual units for events in our work because verses would provide a good locality as they normally consist of a few sentences in average. With this configuration, we have 1,533 events in the book of Genesis.

Definition 2 (Interpersonal interaction): An interpersonal interaction is a subset of the names of persons appearing in the same event.

In our model, the persons appearing in the same event are assumed to interact with some of the other individuals in the event for simplicity, whether true or not.

Notice also that interaction events in our model are of the simplest form, having no attributes. All interactions are treated the same. Loving and hateful relationships are accounted for the same level of significance. Both friendly and antagonistic relationships yield no difference in terms of statistical significance.

In addition, the cultural and geographical backdrop of an interaction is also irrelevant. We do not make distinction between family, political, military and religious events.

Definition 3 (Arity of interaction event): The arity of an interpersonal interaction event is determined by the number of distinct persons participating in the event.

The arities of interaction events range from 2 to any number. The arity of an interaction cannot be any larger than the number of distinct persons appearing in the event by definition. Each individual has only the name attribute in an interaction. Other probable attributes such as sex and title are irrelevant to our current work.

Definition 4 (Maximal interaction): A maximal interaction is an interaction which is not contained in another interaction from the same event log.

As an example, consider an interaction I. No subset of I *can be maximal as they are subsumed by* I. *However,* another interaction I' whose arity is smaller than that of I can be maximal if it is not a proper subset of I and no superset interaction of I' exists in the event log.

The concept of maximal interaction is useful when we visualize significant interactions. By visualizing maximal interactions only, we can effectively summarize interactions without repeating the same information which is already included in other interaction.

Definition 5 (Significance of interaction): The significance of an interpersonal interaction I is the fraction of the events appearing in the same event log.

According to the definition, the significance of an interaction ranges from 0 to 1 with 1 being the highest.

Definition 6 (Interpersonal interaction network): An interpersonal interaction network consists of a set of persons, P, and a set of interaction events, E, each of which is a subset of P.

Example 1: As mentioned earlier, the book of Genesis has a little over 1,500 events, in one of which we read, "*And Adam called his wife's name Eve; because she was the mother of all living*" (Genesis 3:20). This event has an interaction of arity 2 because Adam and Eve participate in a binary family relationship. Another event "And Noah was five hundred years old: and Noah begat Shem, Ham, and Japheth" (Genesis 5:32) presents a 4-ary interaction which occurred in Noah's family.

3.3 Preprocessing

We used an English copy of the King James Version of the Bible in our project. The book of Genesis comprises 1,533 verses in 50 chapters. The source text is divided into 1,533 textual units as events. Punctuation symbols and stop words are removed from these events.

Note, however, that some biblical names, such as 'Beth-el', contain a hyphen in it. These names are location names with the exception of 'Poti-pherah', 'Tubal-cain' and 'Zaphnathpaaneah'. Hence, preprocessing should be adapted to properly take this factor into account.

Table 1 gives a statistical overview of the source text.

Category	Total count	Distinct count
Chapters	50	–
Verses or events	1,533	–
Sentences	2,596	–
Words	38,262	2,617
Characters	196,870	–

Table 1. Statistics on the book of Genesis

3. 4 Named-entity Extraction

The primary object of this step is to obtain the names of the persons appearing in the source text.

$n = 2617$	Predicted		
	True	False	
Actual True	TP = 288	FN = 71	359
Actual False	FP = 89	TN = 2169	2258
	377	2140	

Table 2. True positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) of people names from NeuroNER

Category	Count	Ratio
People names	359	55.6%
Place names	107	16.6%
Stopwords	180	27.8%
Total	644†	100%

† Two words are both people and place names.

Table 3. Statistics on the distinct capitalized words in the book of Genesis

$n = 2617$	Predicted		
	True	False	
Actual True	TP = 359	FN = 0	359
Actual False	FP = 107	TN = 2151	2258
	466	2151	

Table 4. True positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) of people names from our heuristic

A robust named-entity recognition (NER) system would require a large amount of training data to which hand-crafted features, such as part-of-speech tags, are manually added while incorporating domain-specific knowledge [15]. However, building such a robust NER system is not the primary objective of our work. In our work, we first experimented with NeuroNER [16]. We also implemented a cost-effective heuristic.

NeuroNER is an artificial neural network-based NER tool. It requires a large volume of manually-tagged training dataset to make the model useful. This characteristic is common all learning-based algorithms. The tool comes with two publicly available datasets, CoNLL 2003 and i2b2 2014. Training the model using these datasets take hours on a descent personal computer.

NeuroNER also offers pre-trained models. We used this option as we can bypass a lengthy training time. With the pre-trained models, NeuroNER identified 376 people names, 51 place names, 107 organization names and 27 miscellaneous names. The confusion matrix for people names is shown in Table 2. As shown in the table, 71 (19.8%) of 359 true people names are not detected and 89 non-people names are recognized as people names by NeuroNER. We are concerned with the relatively high false positive count as they may introduce noise to the mining of interpersonal interaction. Theoretically, the performance of the model should improve if we train the system with properly tagged datasets. However, this entails overhead for manual tagging and training time. Because of this overhead, we were motivated to take an alternative approach.

As an alternative to NeuroNER, we developed a cost effective heuristic based on the following observation:

First, the name of a person always starts with an uppercase letter in an English version of the book of Genesis. These words can be easily extracted using the UNIX shell commands or by writing a relatively simple text processing program.

Second, not all capitalized words are the names of persons. Many of these words are capitalized because they are the first word of a sentence. These words include stop words such as 'and,' 'after,' 'for,' 'here,' 'in,' 'there,' 'while,' and the like, and several other verbs. These capitalized words do not produce an adversary effect on this heuristic because stop words are already removed during the preprocessing stage. Also, the capitalized words which are not stop words are statistically insignificant, and hence are pruned during the mining stage.

Third, another category of the capitalized words which are not the names of persons are place names. We rely on a dictionary lookup to remove these words in our approach.

Employing the heuristic approach described above, we found 5,531 occurrences of 644 distinct capitalized words from the book. Out of these 644 words, 180 words are stopwords, 107 words are place names, and the remaining 359 words are people names as shown in Table 3.

Table 4 shows the performance of the heuristic when a dictionary for places names is not used, i.e., all capital words are mistakenly considered as people names. The heuristic exhibits a slightly better false positives than NeuroNER with no false negatives. The result is promising with the benefit of requiring no training data nor time. The place name dictionary helps to reduce FP to zero in our approach.

At the end of this process, 1,064 events out of 1,533 events turned out to contain people names and are ready for interaction mining.

3.5 Interaction Mining

We have defined the significance of an interaction previously in Definition 5. Having the event log of interactions among the people appearing in the book of

Category	# of instances	Significance (%) Min	Max
binary interactions	1230	0.00093985	0.031955
maximal interactions of any arity	225	0.00093985	0.00845865
maximal interactions of arity 2	50	0.00093985	0.00845865
maximal interactions of arity 11	1	0.00093985	0.00093985

Table 5. Significance of interactions discovered in the book of Genesis. A maximal interaction is an interaction which is not contained in another interaction. The number of persons appearing in maximal interactions ranges from 2 to 11

Genesis, we now compute the significance of every interaction of which arity is two or large.

Definition 7: Given an interpersonal interaction network (P, E), let $P = \{p_1, p_2, \ldots, p_m\}$ be the set of persons and $E = \{e_1, e_2, \ldots, e_n\}$ be the set of interaction events in the event log where $e_i \subseteq P$. The significance of an interaction $I \subseteq e_i$ $(1 \leq i \leq n)$, $\sigma(I)$, is calculated as

$$\sigma(I) = |\{e_i | e_i \in E \wedge I \subseteq e_i\}| / |E|$$

The significance of each interaction can be conveniently calculated using the A priori algorithm with a low minimum support configuration. In our analysis, we used the Borgelt's implementation [17].

A statistical summary of the result of the Apriori algorithm is shown in Table V. We have found 1,230 binary interactions which occur in at least one event (i.e., one verse) in the book of Genesis. The top 12 most frequently appearing names in these events are God, Abraham, Jacob, Joseph, Isaac, Pharaoh, Esau, Laban, Sarah, Rachael, Israel, and Noah. Notice that Abraham is the father of Isaac, Isaac is the father of Jacob, and Jacob is the father of Joseph. These four generations of Abraham's family are statistically dominating individuals in the accounts of the 402 people in the book.

Next to Abraham, Isaac, Jacob and Joseph, Pharaoh an ancient Egyptian ruler, Abraham's wife Sarah, Jacob's brother Esau, Jacob's father-in-law Laban, Jacob's two wives Rachel and Leah, and Abraham's 10th ancestor Noah are also commonly appearing names in the book.

Out of the 1,230 binary interactions, 20 interactions appear at least 1% of significance. They are shown in Figure 3. The most significant binary interactions is between God and Abraham, followed by the interactions between Joseph and Pharaoh, and between Abraham and Sarah. Figure 4 shows the interaction network of the persons who participate in the top 16 significant binary interactions. Each edge is annotated with its statistical rank of the respective interaction with 1 being the highest. The rank of an interaction also determines the thickness of the link in the visualization. In addition, solid red edges denote family interactions and dashed blue edges denote interactions with a non-family member such as God or a ruler.

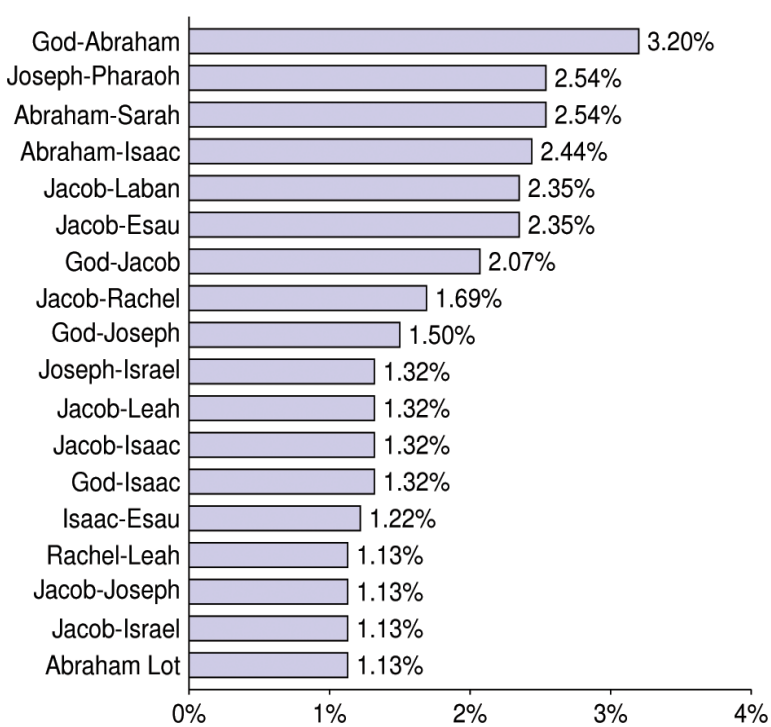


Figure 3. Binary interactions with a significance of 1.1% or higher found in the book of Genesis. The most significant binary interaction in the book is the interaction between God and Abraham, followed by Joseph and Pharaoh. Also, the interactions between the four generations of Abraham and their family members are noticeable

We notice two outstanding patterns from the analysis as follows: First, since the book of Genesis is a religious text, it is no surprise that the interaction of God with several prominent persons such as Abraham, Issac, Jacob and Joseph are common events in the book. However, when an individual life is concerned, the most common events in the book are interactions with family members such as loving wives, fathers, sons or brothers, with an exception of Joseph's interaction with his Egyptian ruler. In other words, the book of Genesis can be thought of largely as a family journal. Joseph's exception can be justified by the fact that he was given an extraordinary responsibility for the welfare of the kingdom. This pattern is illustrated in Figure 4.

Second, the interaction between Abraham and God is the most prominent relationship in the book of Genesis. This is congruent to the traditional view of Abraham as "father in faith" in the Christian world. The three major world faiths, namely, Judaism, Christianity and Islam, give Abraham a high position of respect.

We have also found 225 maximal interactions which occur in at least one event in the book (see Table V). The arity of these interactions ranges from 2 to 11. Because the statistical significance of interactions follows the anti-monotone property, we see that the occurrences of binary interactions are much higher than the occurrences of higher-arity interactions. In Table 5, the minimum significance 0.00093985 is equivalent to one event (1; 0640:00093985 = 1), and the maximum significance 0.031955 is equivalent to 51 events in our dataset. The largest interaction of arity 11 comprises Benjamin, Rosh, Naaman, Muppim, Becher, Belah, Ehi, Gera, Huppim, Ashbel and Ard. These interactions are presented in the Visualization section below.

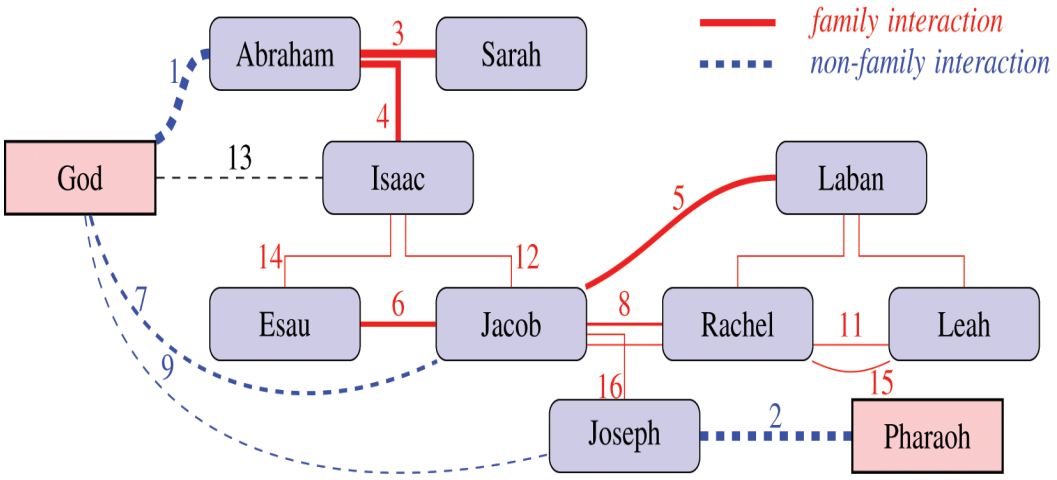


Figure 4. Genealogical relationships of Abraham and other important characters in the book of Genesis who participate in strong binary interactions. The number associated with a link between names represents the statistical rank of the respective interaction with 1 being the highest. Solid red lines denote family interactions between husband and wife, parent and child, brothers, or father-in-law and son-in-law. Blue dashed lines denote interactions with the people outside family

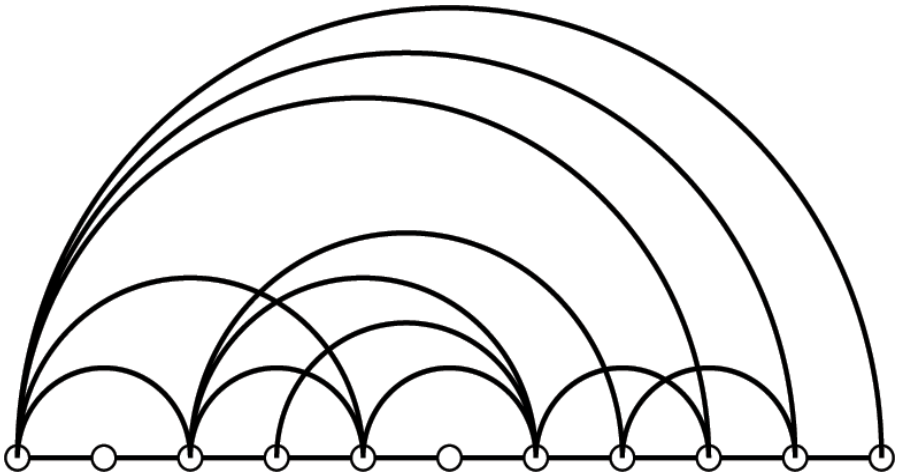


Figure 5. A variation of Kandinski layout where all the edges are rendered in the upper hemisphere. The information captured in this visualization is as same as the information in the Kandinski of Figure 2b

3.5 Visualization

While the textual output from the interpersonal interaction mining step provides all the discovered patterns in detail, visualization plays a unique role in analyzing the data visually as well as in the presentation of the results. In this section, we present the visualization of the results from our experiment using the 2D and 3D visualization tools developed in-house.

First, we experimented the results of the interaction mining by adapting the circular and the Kandinski layouts introduced in Figure 2b. In particular, we made a modification to the Kandinski layout such that all the edges are rendered on the northern hemisphere, as illustrated in Figure 5, without compromising the integrity of the information presented. Our motivation for the modification is to avoid misinterpretations that may occur when some of the edges appear on the north and some others on the south of the axis.

Figure 6 shows the visualization of the binary interpersonal interactions between the 406 names from the book of Genesis using the modified Kandinski layout where the names are ordered alphabetically from left to right on the single linear axis. Readers must be careful when they read the information presented using the Kandinski layout. Here, the significance of an interaction is represented only by the color and thickness of the corresponding edge. The height or length of the arc is irrelevant. In the figure, the red-orange-yellow-green-blue-gray color scheme is used where the red color represents the highest significance. Using the color scheme, only a small fraction of the binary interactions of higher significance is outstanding while many other insignificant interactions are suppressed as the gray backdrop. Abraham's interactions with God, Issac and Sarah stand out clearly in the figure. Another noticeable interaction which is not mentioned before is the one between Abraham and his cousin Lot.

One potential issue we noticed in Kandinski is that if strongly interacting names appear in a very close proximity, their interactions may not be effectively visualized. It is because the radius of the arc between the two nodes will become extremely small, and subsequently the visual effect of the relationship will become unnoticeable. It will vanish away from the reader's attention.

Figure 7 shows the binary interactions rendered in a circular layout. All the 406 names are arranged along the circumference of a circle counterclockwise, starting from the east end of the circle. The same color scheme is used as before with red being the highest significance. One advantage of the circular layout over Kandinski is that names have more space in between as they are arranged on the circumference of the circle which is 3.14 times longer than the linear axis of Kandinski.

2D visualization is useful when used with a reasonable number of binary interactions to capture the overall pattern of the interactions or when there is a small fraction of outstanding interactions. However, it is challenging to visualize interactions of higher arity using 2D. For this purpose, we developed a geometric shape suitable for higher-arity interactions in a 3D visualization tool. Figure 8 shows a 3D rendering of the same information shown in Figure 7. Abraham is located at the north end with the red edge to God. Figure 9 shows examples of higher-arity interactions in 3D.

Figure 10 shows the maximal interaction of arity 11 found

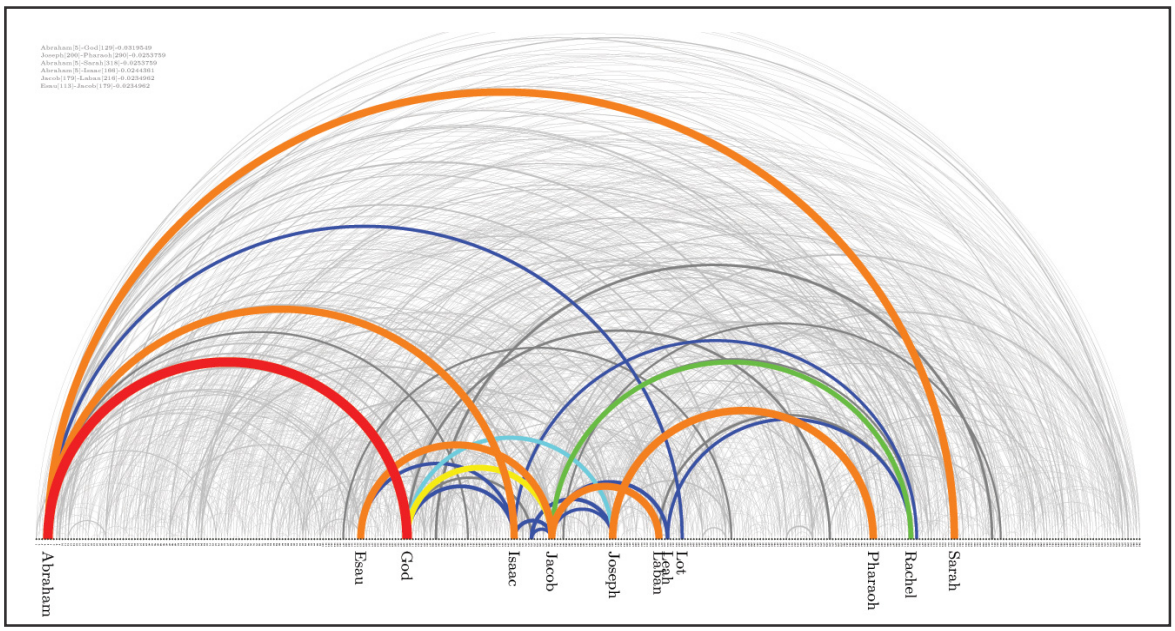


Figure 6. Binary interpersonal interactions from the book of Genesis shown in a Kandinski layout. The strength of an interaction is represented by the color and thickness of the corresponding edge. The red-orange-yellow-green-blue-gray color scheme is used with red being the highest significance

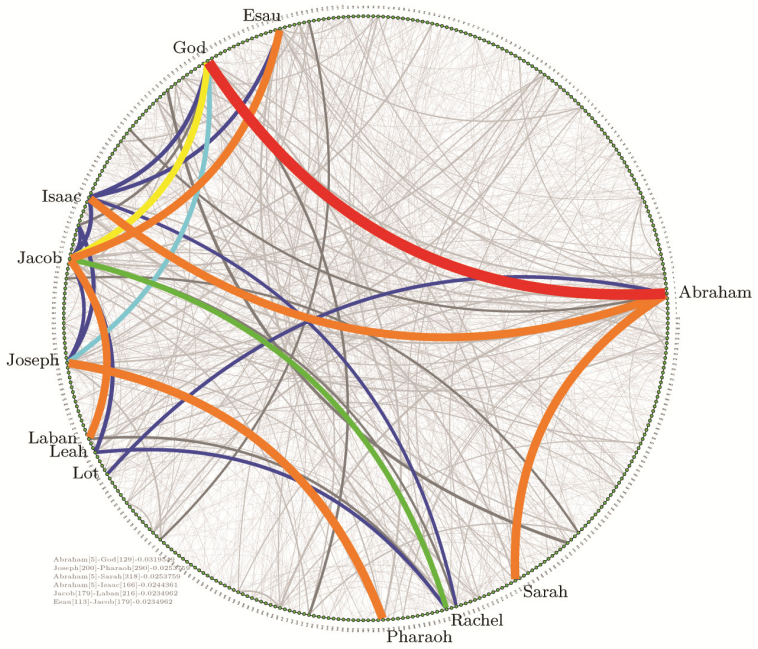


Figure 7. Binary interactions from the book of Genesis shown in a circular layout. Interactions are rendered using the same red-orange-yellow-green-blue-gray color scheme. The high interaction between God and Abraham is noticeable along with his interaction with Sarah and Issac. Pharaoh holds a unique position in the book

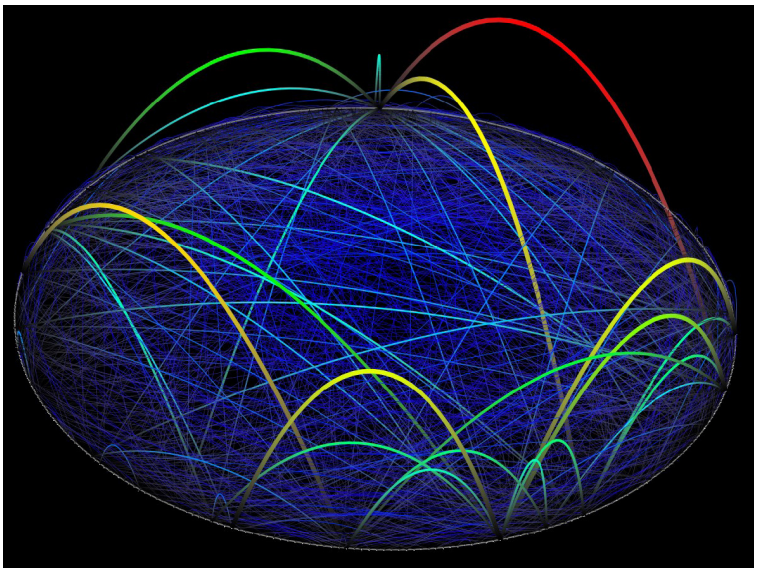


Figure 8. A 3D visualization equivalent to Figure 7

in the book of Genesis which consists of Benjamin and his 10 sons.

Our tools provide zoom, pan and rotate functions as well as edge picking by which the user can inspect any interaction dynamically.

4. Conclusion

In this paper, we developed a formal framework for interpersonal interaction mining from classic books. We derived three observations from an experiment conducted over the book of Genesis using the proposed approach.

First, classic books have a potential for social network mining. As they have been treasured for ages, they may offer some hidden wisdom that may be beneficial to us.

Second, the result of the experiment confirms the traditional view of Abraham. Joseph's interaction with Pharaoh is also an important substance. At the same time, the analysis offers a fresh insight on Abraham's four generations. They are well engaged in family interactions. It may sound counterintuitive, but the major substance of the book of Genesis is indeed the interactions among these Abraham's families.

Third, we found that a cost-effective heuristic works well

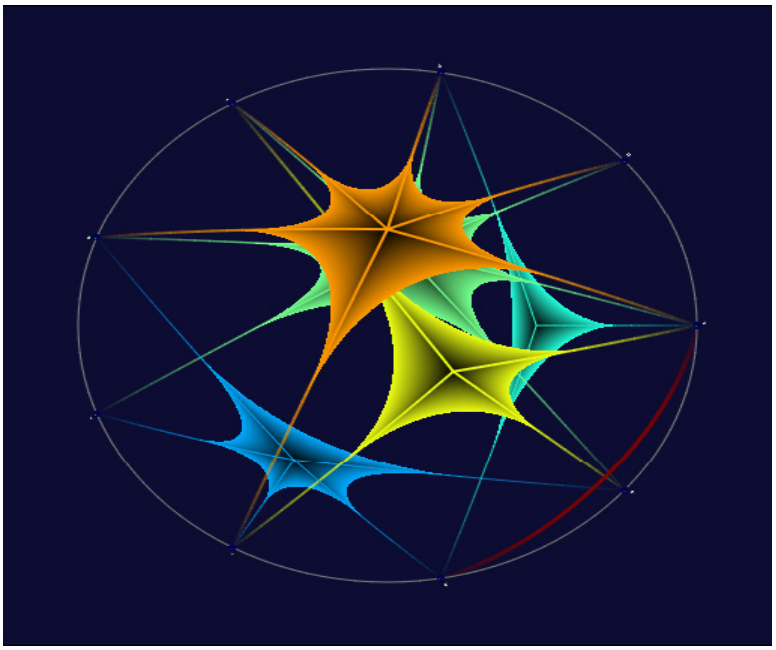


Figure 9. High-arity interactions visualized using the 3D geometric shapes. The number of edges represents the arity of the corresponding relationship

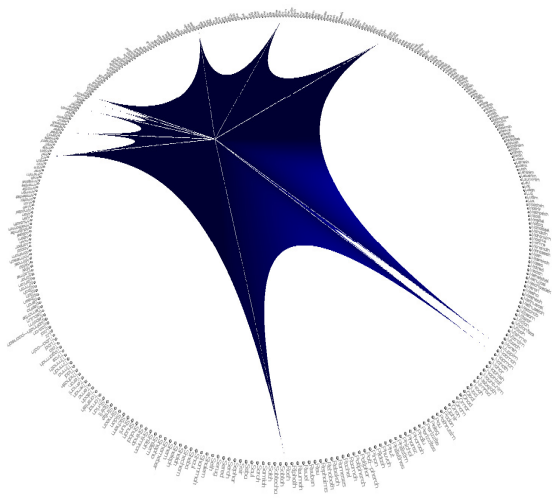


Figure 10. The interaction among 11 people is rendered as a 3D hypergraph node in this figure

without using a robust but heavy-weight named-entity recognition tool for extracting people names from the book of Genesis.

We will examine if this benefit holds for other classic books.

Fourth, 2D and 3D visualization approaches complement each other as each approach has own merits and challenges.

Generally, when the amount of the visual information grows large, the interactive drill-down capability of 3D becomes valuable.

We plan to extend our work in two ways in the near future. First, we plan to perform a lager scale analysis using more volumes from the Bible and other classic books from ancient Greece and Rome. Second, we plan to introduce bipolar attributes to interactions such as love and hate interactions.

References

[1] Duncan, J. (1998). Watts and Steven H. Strogatz. Collective dynamics of 'smallworld' networks. *Nature*, 393. 440–442,

[2] Newman. M. E. J. (2003). The structure and function of complex networks. *SIAM Review*, 45 (2)167–256.

[3] Lada, A. (2003). Adamic and Eytan Adar. Friends and Neighbors on the Web. *Social Networks*, 25 (3) 211–230,

July.

[4] Rothenberg, Richard. (2002). From Whole Cloth: Making up the terrorist network. *Connections*, 24 (3) 36–42.

[5] Ressler, Steve. (2006). Social Network Analysis as an Approach to Combat Terrorism: Past, Present, and Future Research. *Homeland Security Affairs*, 2 (8) July.

[6] Moreno, J. L. (1934). Who shall survive? A new approach to the problem of human interrelations. *Nervous and Mental Disease Pusblishing Co*.

[7] Rapoport, Anatol., Horvath, William J.. (1961). A study of a large sociogram. Systems Research and Behavioral Science, 6 (4) 279–291, October.

[8] Fredrik Liljeros., Christofer R.. Edling., Luis A. Nunes Amaral, H., Eugene Stanley., Yvonne Aberg. (2001). The web of human sexual contacts. *Nature*, 411. 907–908.

[9] John D. Lafferty., Andrew McCallum., Fernando C. N. Pereira. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. *In:* Proceedings the 18th ICML, p. 282–289.

[10] Xuezhe Ma., Eduard Hovy. (2016). End-to-end sequence labeling via bidirectional lstm-cnns-crf. In Proc. of the 54th Annual Meeting of the Assoc. for Computational Linguistics (1: Long Papers), August.

[11] Bekos, Michael A., Michael Kaufmann., Stephen, G. Kobourov., Antonios Symvonis. (2013). *Smooth Orthogonal Layouts*. 17 (5) 575–595.

[12] Friendly, Michael. (2002). Corrgrams: Exploratory displays for correlation matrices. *The American Statistician*, 56 (4) 316–324, Nov 2002.

[13] Eric, D., Bernholc. (2011). The Value of Family in Ancient Greek Literature. http://www.ancient.eu /article/254/, October 7.

[14] Seymour-Smith, Martin. (1998). The 100 most influential books ever written: the history of thought from ancient times to today. Carol Publ. Group, September.

[15] Lample, Guillaume., Ballesteros,Miguel., Subramanian, Sandeep., Kawakami, Kazuya., Dyer, Chris. (2016). Neural architectures for named entity recognition. CoRR, abs/1603.01360.

[16] Dernoncourt, Franck., Lee, Ji Young., Szolovits, Peter. (2017). NeuroNER: an easy-to-use program for named-entity recognition based on neural networks. Conference on Empirical Methods on Natural Language Processing (EMNLP).

[17] Christian Borgelt. (2012). Frequent Item Set Mining. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2 (6) 437–456.