Properties and Structure of Bible Lexical Network

Clark Brown, Brigham Young University

Introduction

For this project, I examined the lexical network of proper nouns in common verses in the King James Version of the Holy Bible. I was led to this network from the Koblenz Network Collection.[1] The network data was procured from Linton Freeman of the U.C. Irvine sociology department.[2]

I was drawn to this network because of my love of scripture and my appreciation for the intricate references between various groups and people in histories, prophecies, and teachings in the Bible and in the Book of Mormon. I limited myself to the already collected data for the Bible, but I think that a similar network that contains data for the Book of Mormon or for all scriptural canon of the Church of Jesus Christ of Latter-day Saints could be of interest.

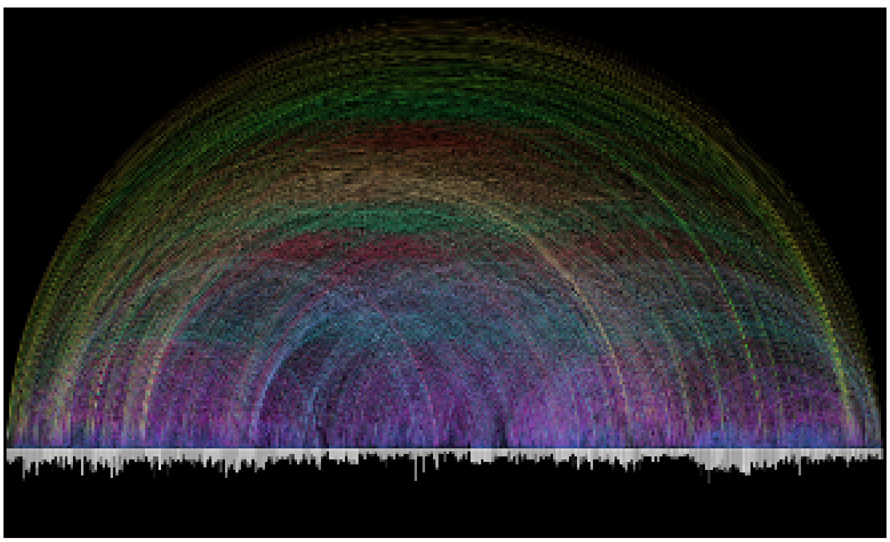

Figure 1
A visualization of the Bible cross references (colorful arcs) of proper nouns between the verses of the Bible (grey bars at the bottom).[3] The network is built out of these connections as explained in the following section.

Data and Potential Errors

Each node represents a proper noun (individuals, groups, nations, places) in the Bible. An edge between nodes represents that the two proper nouns appear together in a verse. The weight of the edge represents the number of verses in which the two nouns appear together.

The data for this network[4] was collected by Christoph Romhild and Chris Harrison. The method of tallying is unknown. There could be missing or extra edges from a mis-tally. There could also be missing nodes, but this is less likely. In a broader sense, this network was created from one

---

[1] The Koblenz Network Collection, Universität Koblenz-Landau, http://konect.uni-koblenz.de/
[2] Research Website of Dr. Linton Freeman, http://moreno.ss.uci.edu/data.html#bible
[3] Chris Harrison, https://www.chrisharrison.net/index.php/Visualizations/BibleViz
[4] Bible Names, Christoph Romhild and Chris Harrison, http://moreno.ss.uci.edu/names.dat

translation of the Bible in English (KJV) which leads to differences from what we would expect if a lexical analysis of the “original” documents were possible. Some proper nouns that differed in Hebrew or Aramaic may have been translated to the same word in the KJV, and the reverse phenomenon could also have occured. We will ignore these possibilities for now and treat the network as data relevant mostly in the context of the King James Version of the Bible.

Structure

The network could be thought of as a social network, but in general the analysis that follows treats the network more as a lexical network or a co-occurrence/contact network. The network is a single-layer, weighted, undirected network with 1773 nodes and 9131 edges and no self-loops. Like most real-world networks, the network is not planar. The diameter is 8 edges which is a little more than ln(n), and the average shortest path is 3.38 edges.

There are no isolated nodes because the network only includes proper nouns (nodes) that are in a verse together with another proper noun. So every node is connected to at least one other node making the minimum degree 2. Israel has 364 neighbors which is the maximum degree.

The size of the network is somewhat larger than the optimal for running some of our network algorithms, but it is not unmanageable. I used Python for the analysis of this network. I relied on the efficiency of the NetworkX and NumPy libraries for network algorithms and linear algebra computations. The longest calculations that I was able to complete involved paths in the network. Calculating the average shortest path was computationally expensive and ran overnight.

Other difficult tasks included community detection algorithms. Trying to find all of the k-components in the network was a disaster and crashed my machine because of temporal and spatial complexity. K-cores were much easier to find by the percolation method.

The greedy Modularity Maximization algorithm found one giant community that resembles the largest connected component and several smaller communities of four or five nodes each. Running the asynchronous fluid communities algorithm to find 100 communities in the largest connected component yielded more interesting results. For instance, it placed Agrippa and his daughter, Bernice, in the same community. It also made communities that also line up with the Biblical narrative like {Adam, Eve, Seth, Cain, Abel}, {Abraham, Abram, Sarah, Sarai, Isaac, Jacob, Rebekah, Ur}, and {Isaiah, Hezekiah, Ahaz, Amoz, Uzziah, Rezin} from the Old Testament and {Bethany, Lazarus, Martha, Mary, Magdalene} and {Andrew, Bartholomew, Galilee, Judea, James, Judas, Matthew, Nazareth, Peter, Philip, Thomas} from the New Testament.

Measures

I ran several algorithms for centrality measure. HITS and PageRank both are designed for directed networks, so I first converted every undirected edge to a pair of corresponding directed edges before running the algorithm. This method gives the authority vs. hub distinction from

HITS less meaning, but each can give us a centrality measure for our undirected network. The top ten (most central) results for each of these algorithms are listed in Table 1 below.

Table 1. *Centrality Measure Rankings*

Rank	1	2	3	4	5	6	7	8	9	10
Betweenness	israel	jerusalem	judah	david	benjamin	egypt	manasseh	ephraim	jesus	joseph
Closeness	israel	judah	jerusalem	david	egypt	ephraim	manasseh	benjamin	joseph	moses
Degree	israel	judah	david	jerusalem	egypt	benjamin	manasseh	ephraim	saul	philistines
Eigenvector	israel	judah	david	jerusalem	egypt	benjamin	ephraim	manasseh	moses	jordan
HITS Authorities	israel	judah	david	jerusalem	moses	saul	egypt	aaron	benjamin	solomon
HITS Hubs	israel	judah	david	jerusalem	moses	saul	egypt	aaron	benjamin	solomon
Katz(a=.01)	israel	judah	david	jerusalem	egypt	benjamin	manasseh	ephraim	saul	moses
Katz(a=.001)	israel	judah	david	jerusalem	egypt	benjamin	manasseh	ephraim	saul	philistines
PageRank	israel	judah	david	jerusalem	moses	saul	egypt	jesus	manasseh	benjamin
Most Common	israel	judah	david	jerusalem	egypt	benjamin	manasseh	ephraim	saul	moses

For every measure, we have Israel as the most central node. Though exact placement varies, we see Jerusalem, Judah, and David as the next three candidates. Those familiar with the Bible will not be surprised to find that Israel plays a central role to the text. The same goes for Judah, David, and Jerusalem.

Though the Old Testament takes up a significant amount of the text of the Bible, it may seem strange to some that Jesus was only listed in the top ten for PageRank and Betweenness. This is where an understanding of the network structure is important. Edges represent a shared verse between proper nouns. Centrality in this network does not mean that the node is the most important figure or place in the Bible. Although it could be indicative of such importance, it means the figure or place is central in connection to other people and places in the Bible. The word Jesus appears mostly in verses with no other proper nouns because the verses describe his teachings and actions. Additionally, there are two separate nodes for Jesus and Christ which ultimately refer to the same person.

The assortativity by unweighted degree is -0.0488, and the assortativity by weighted degree is -0.0461. These values suggest that the network is somewhat disassortative by degree and by edge weight. This implies that proper nouns that show up in verses with many other proper nouns or that have many appearances with the same noun are not more likely to be in verses with other “well-connected” nouns.

Real World Features

The degree distribution of the network is power-law-like. The distribution is shown below. The line that is fit to the data is for a power law with exponent 2.289. In real world networks, we often see this same kind of distribution of the degree of nodes. Scale-free networks often have a degree distribution with a power-law exponent between 2 and 3. There are fewer and fewer highly connected nodes. However we do still see a cluster in the degree distribution. Another visualization of the weighted degree of the nodes is in Figure 3.

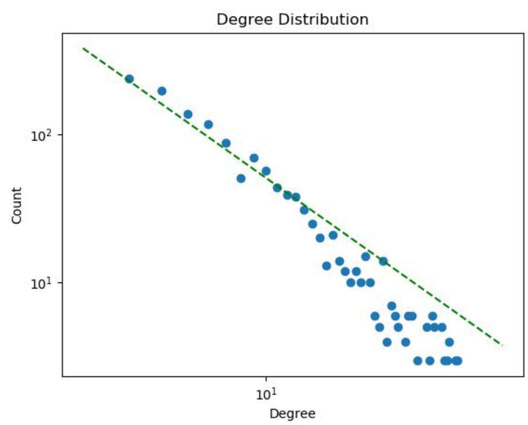

Figure 2. *The Degree Distribution and Best Fit Power Law Curve*

Figure 3. *Another visualization of the network with size corresponding to weighted degree[5].*

Another real world property of the network is the giant component. The largest connected component in the network contains 96% of the network (1707 of 1773 nodes). The network also appears to be sparse, with its density close to 0.0058.

The global clustering coefficient is approximately 0.721 suggesting that the network tends toward tight-knit groups in its structure. This high clustering coefficient is indicative of what we see in real world networks. Because the network is a fixed network barring any unforeseen changes to the Bible, it is difficult to know if the clustering coefficient will remain bounded away from zero as the number of nodes increases. However, with such a high clustering coefficient to begin with, it is not difficult to imagine that that would be the case as we have with real world networks.

If we wanted to try and model this network with a known model we would have to limit ourselves to models for undirected networks or adapt a model designed for directed networks. Preferential attachment or maybe Barabasi-Albert could give us the power law degree distribution that is similar to the Bible network. Modeling for different n would be difficult though because of the fixed nature of the network. It lacks dynamics. One way to introduce time or dynamics to the network would be to build the network chronologically by verse in the Bible. We would add the nodes as they appear, and add edges as the nodes appear together in verses moving from Genesis to the Revelation of St. John.

[5] Chris Harrison, https://www.chrisharrison.net/index.php/Visualizations/BibleViz

Conclusion

With its power-law like degree distribution, significant giant component, tight clustering, and sparsity, the Bible Lexical Network shows many of the features we see in real world networks as outlined in the previous section. Non-network-like models like Erdős–Rényi and others would not be able to capture the unique features of this network. It has the features that we look for to justify networks apart from the general study of graph theory.

The Biblical network is a fascinating one. I believe the data could be improved by better tallying or accounting for multiple people with the same name (Mary) or synonyms referring to the same place (Zion, Sion) or person (Jesus, Christ). Even then, I believe it would have these real world features. The findings of my study[6] into the network structure could possibly be helpful to those who study the Bible as a historic, literary, or sacred text. I would be interested to see what Bible scholars make of these findings.

[6] The code and data for this project are open source: https://github.com/clarkbyu/bible-network-project