The Standard Works Reference Network

J. Nolan Faught

April 8, 2022

Abstract

I describe my methodology for creating the network of references between the books in the Standard Works and analyze the community structure and connectivity of the resulting network. I show that the network has several real-world properties and modify a few existing models to generate networks that mimic some of the unique properties of this network.

1 Methodology

The Standard Works is the collection of books accepted by the Church of Jesus Christ as canon, comprising four texts available on the Gospel Library[1] web app. One of the study aids provided by the church are footnotes, many of which provide cross-references between relevant portions of text. The Standard Works reference network is the weighted digraph generated by these references.

The church has made a valiant effort to make the Standard Works and related study resources available in as many languages as possible, with key portions of the texts available in 136 languages. However, the references are slightly different between languages due to the abiguity or lack thereof in different translations of the texts. For this project I focus only on the english translation.

To construct the network, I implemented part of the Gospel Library API in Julia and used it to generate a .csv file detailing the references found in the footnotes. I used Julia's robust graph, database, and linear algebra libraries to create the network and analyze its characteristics.

1.1 Possible Sources of Error

I was unable to devise tests to check false negatives, so there may be a few references that are absent from the weights. The *Topical Guide*, *Index of the Book of Mormon*, and *Bible Dictionary* are included in (english) print copies of the Standard Works but are not a part of the scriptural canon. The exclusion of these resources only constituted a very small portion of edges.

2 Structure

The Standard Works reference network contains 36250 total references, with 3692 edges and 101 vertices, including several references to parts of the Joseph Smith Translation (JST). The resulting graph is incredibly dense, with an edge proportion of 0.32, and highly connected. The strongly

[1] https://churchofjesuschrist.org/study/scriptures

connected component contains 87 vertices and 36194 references, which is 99.85% of the data in the network.

In the figure below, nodes are colored and positioned according to their core number, increasing as we move toward the center.

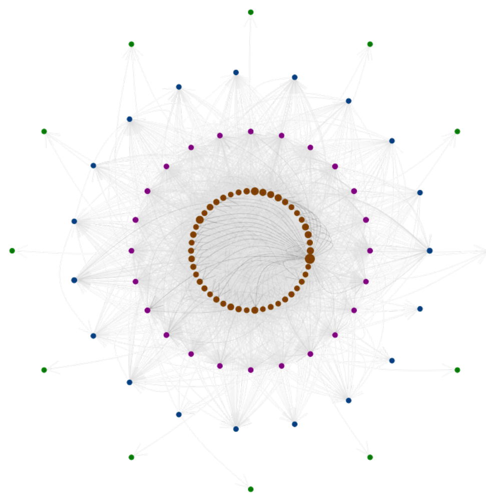


Figure 1: The Standard Works Reference Network

2.1 Components

The outermost group of nodes are the out-component of the graph, corresponding to books that are referenced but do not reference other texts. This group includes books such as JST Genesis, JST Leviticus, the Witness of the Eight, and the Introduction of the Doctrine and Covenants. All other vertices are in the strongly connected component. Unlike real-world graphs, the strongly connected component occupies much more than half of the graph and the weakly connected component is the entire graph.

2.2 Centralities

The centralities of the network deviate heavily from other real-world networks and provide interesting insight as to the most "important" books in Standard Works. Doctrine and Covenants had the highest degree by far, which resulted in it scoring high for nearly every centrality measure. Books that are typically not highly referenced, such as Ruth and Philemon, scored high for betweeness and closeness centrality.

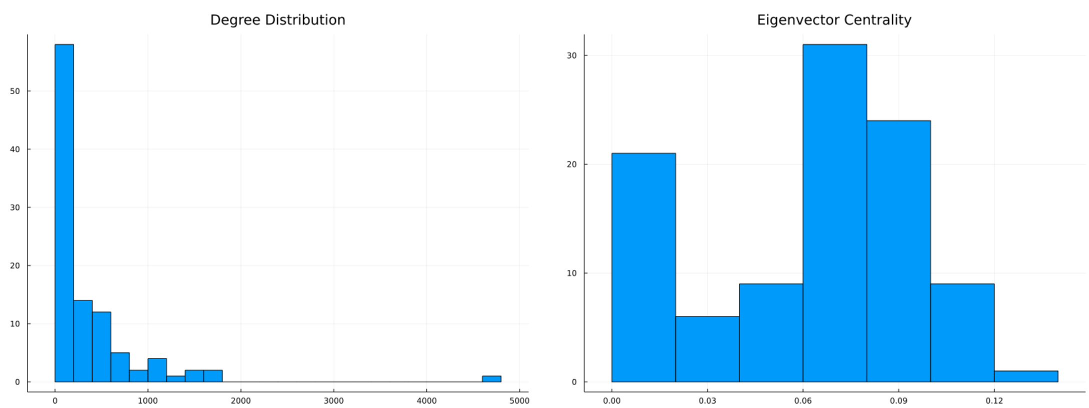
Top five books for common centrality measures

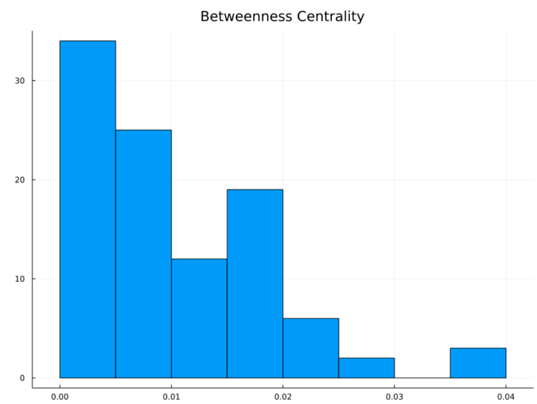
Degree	Eigenvector	Pagerank	Betweeness	Closeness
D&C	Malachi	D&C	Ruth	3 John
Alma	Mormon	2 Nephi	Malachi	Philemon
2 Nephi	Proverbs	Alma	D&C	2 John
Isaiah	Hebrews	1 Nephi	Proverbs	Joseph Smith Matthew
1 Nephi	Leviticus	3 Nephi	Matthew	Articles of Faith

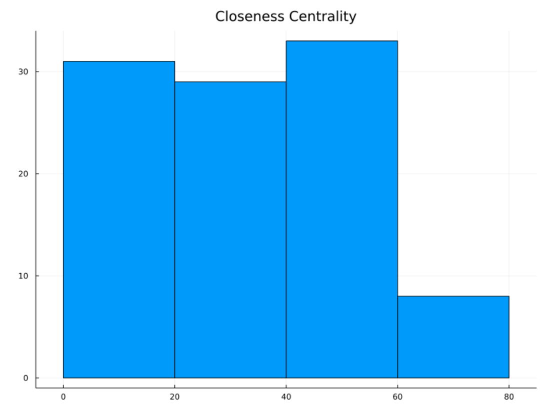
Isaiah is commonly referred to by other prophets and 2 Nephi contains several passages that are comparable to Isaiah, which is why these two books appear so often. Malachi serves as a “bridge” between all books, cited by several books in each text in spite of its short length and small number of references. The reason for which 2/3 John and Philemon score so high for closeness centrality is unknown.

The degree distribution, similar to most real-world networks, appears to follow a rough power law with exponent $\alpha = 2.1524$, within the expected range for a real-world network. Most abnormal is the distribution of the eigenvector centrality, which appears almost normal (the high number of books with eigenvector centrality in $[0, 0.03]$ consists mostly of the out-component). This is likely due to the large number of edges flowing into (and out of) Doctrine and Covenants.

Distribution of Centralities







3 Community Structure

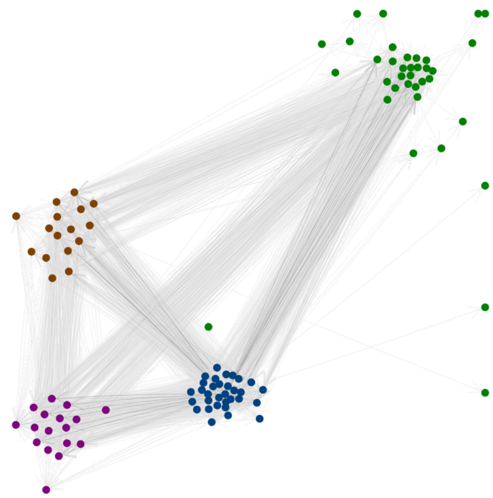
The network is highly connected, having a global clustering coefficient of $C = 0.1646$. There is no agreed-upon convention for reciprocity of directed weighted graphs, but in this case I was able to compute two measures that are very similar. If we consider reciprocity of an unweighted graph to be the expected number of edges reciprocated to a random edge, then we may consider weights in the network to record multiple edges and compute the same value through probability theory. The computation yields an expected value of 1.051 back-edges for an arbitrarily chosen edge (i, j). Alternatively, for an edge (i, j) with weight w_{ij}, the expected value of w_{ji} is $1.051w_{ij}$.

If we consider the reciprocity as the probability that $w_{ji} > 0$ given $w_{ij} > 0$, then the computation yields $r = 0.862$, which is incredibly high for any network.

I used two community-detection algorithms to examine the network. Figure 1 positions the nodes of the network according to their core number. The innermost shell corresponds to nodes with core number 62, which composes half of the graph. The second shell contains vertices with core number between 50 and 62, and combined with the first shell represents about 3/4 of the graph. The third shell contains vertices with core number between 5 and 50. The first three shells are the strongly connected component, and the final shell the out-component. This community structure is comparable to the expected structure of a directed network (the "bow-tie"), but with the in- and out-components being replaced with cores.

The graph is dissassortative with respect to degree, having a modularity coefficient of $Q = -0.2103$. I had expected the network to show a strong assortativity when the books were grouped according to their text (Old Testament, New Testament, etc.), but even this clustering had a score of $Q = 0.0657$, only weakly assortative. The result of a modularity maximization is shown below

Communities based on modularity maximization



Interestingly enough, the clusters created by modularity maximization appeared to correspond roughly to the function or role of the books. The largest cluster, colored green, contains all of the Joseph Smith Translations referenced and several books of the minor prophets. The second-largest cluster, which appears rust-colored above, contains books of historical significance. Each cluster appears to be centered around a book of major doctrine or prophecy such as Isaiah, Revelations, or 2 Nephi. However, even this clustering was only weakly assortative, with $Q = 0.128$.

4 Modeling the Network

We didn't discuss many models designed to create multi-edges, so the only feasible model for emulating the structure of this network was the configuration model. I briefly considered using a modified version of Price's model with the additional constraint that whenever a node is added, each node in the network is allowed to attach edges, but this does not seem to capture the properties of this network very well.

5 Conclusion

The centralities and degree distribution have properties similar to real-world networks. The clustering coefficient is incredibly high, even for a real-world network, but this may be attributed to the density of the network. The network does follow the small-world property, but this is most likely due, again, to the density of the edges.

The sheer density of the network makes it difficult to believe that it is similar to other real-world networks. The network is certainly not a citation network, as it is not acyclic. However, although the community structures of the graph are somewhat weak, the other properties of the network make it seem like a real-world network; my hypothesis is that modifying the level at which references are recorded would change the community structure to match other real-world networks.