Unraveling Cain's Jawbone: Narrative Detection and Analysis Using Louvain and Spectral Clustering Algorithms

MAXIMILIAN STOLLMAYER
ABDELBAST NASSIRI

July 2023

1 Introduction

Cain's Jawbone is an example of a literary puzzle book Published in 1930. The book presents a murder mystery where the reader is tasked with solving the crime by rearranging the 100 pages of the book in the correct order. The pages are printed on individual slips, and the reader must decipher the correct sequence of events to reveal the solution.

The primary goal of this project is to apply the Louvain and spectral clustering algorithms to detect and analyze the six narratives embedded within the puzzle. By leveraging these algorithms, our objective is to uncover the distinct narrative threads and story-lines contained within the puzzle's text.

2 Methodology

Our approach consists of the following steps:

2.1 Data Preparation

Before constructing the graph, we perform essential and standard routine for natural language processing to clean and standardize the text. The preprocessing steps include:
-Removing special characters, punctuation, and any non-alphanumeric symbols that may not contribute to the analysis.
-Tokenizing the text into individual words or phrases to facilitate further analysis. Removing common stop words, such as "the," "and," "is," etc., that are frequently used but often add little contextual value.

2.2   Graph Construction

The book was presented as fully connected graph, with each pages as nodes and measured the relationship between the pages using the cosine similarity between the vectors of the pages constructed using TF-IDF and represented it as the weights of the edges between the pages.

2.3   Louvain Algorithm

The Louvain algorithm is a community detectin algorithm. It was proven to be effective in detecting cohesive clusters within complex networks.
In our implementation, we used **louvain_community** function provided by the NetworkX [1] library in python. This function iteratively optimized the network partitioning by maximizing the modularity measure, which quantifies the quality of the clustering. The resulting partition obtained from the Louvain algorithm represented different clusters or communities within the graph. Each cluster was considered a potential narrative within Cain's Jawbone puzzle.

2.4   Spectral clustering algorithm

In addition to the Louvain algorithm, we employed spectral clustering to further analyze the narrative clusters obtained in the previous step. Spectral clustering is a powerful technique for partitioning data points based on spectral properties of the Laplacian matrix.
To apply the spectral clustering, we used the function **spectral_clustering()** [4] to compute the 6 smallest eigenvectors of the Laplacian matrix, and then cluster the points of row vectors of the 100x6 matrix M formed by the eigenvectors using k-mean clustering algorithm into clusters C1 to C6, and output the clusters A1 to A6 with $A_i = \{j, if\, y_j \in C_j\}$, where y_j is the jth row vector of the matrix M.

3   Comparison

After applying the two algorithms for 100 times since there is some randomness involved in both algorithms, we performed a comprehensive analysis of the identified clusters, evaluating their thematic coherence, and their relationships with the overall book, and the mean scores were 0.673 and 0.693 respectively indicating a slight win of the spectral Clustering algorithm in terms of the coherence of it's output clusters and also of the consistency of the output clusters.

4   Results

The application of the Louvain algorithm and Spectral Clustering allowed us to identify cohesive clusters of the nodes within the graph. These clusters represented distinct narrative threads and helped us to reduce the search size from 100! (which is $9.33\text{x}10^{157}$) and that is equal to the number of particles in the

universe with each particle being a universe of it's own to 1.94×10^{23} permutations.

References

[1] URL: https : / / networkx . org / documentation / stable / reference / algorithms / generated / networkx . algorithms . community . louvain . louvain_communities.html.

[2] Yair Weiss Andrew Y. Ng Michael I. Jordan. "On Spectral Clustering: Analysis and an algorithm". In: ().

[3] Ulrike von Luxburg. "A tutorial on spectral clustering". In: (2007).

[4] *Machine Learning in Python*. URL: https://scikit-learn.org/stable/ modules/generated/sklearn.cluster.SpectralClustering.html.