An Investigation of the Community Structures within Social Networks of the Gospels

Maki Miyake (mmiyake@lang.osaka-u.ac.jp)

Osaka University

1. Introduction

Graph representation and the techniques of network analysis are useful for detecting and investigating the intricate patterns of connectively within large-scale corpora. A number of studies have applied the graph theory approach to investigating linguistic resources (Dorow, Widdows, Ling, Eckmann, Danilo, and Moses, 2005; Steyvers and Tanenbaum 2005; Gfeller, Chappelier, and De Los Rios, 2005). Recently, I have successfully applied a graph clustering technique to data processing that utilizes a clustering-coefficient threshold in creating sub-networks for the Gospels of the New Testament (Miyake, 2008). Creating semantic networks of the Gospels is an effective way of capturing the conceptual structures that underlie similar words. In this study, I construct social networks from the four books of the Gospels. Specifically, I propose a soft graph clustering technique as a method of detecting the community structures within the social networks of the Gospels. The principle objectives of this study are to investigate the interaction between the characters in the stories and the transitions (the series of action and events taking place) during the Early Christian period.

2. Applied Methods

In order to reveal the appropriate communities within the social network, this study proposes combining a basic statistical measure of clustering coefficients and the hard clustering of Markov Clustering (MCL). Following Watts and Strogatz (1998), a clustering coefficient represents the inter-connective strength between neighboring nodes in a graph and assumes values between 0 and 1. In the context of social network analysis, the clustering coefficient can be utilized as a measure of role ambiguity or community hubs that have numerous links. MCL is an effective method for detecting the patterns and clusters within large and sparsely connected data structures. Based on random walks for a graph and using the two simple algebraic operations of expansion and inflation, MCL simulates the flow on a stochastic transition matrix in converging towards an equilibrium state. Within the MCL process, a graph is partitioned into hard clusters. However, one particular problem with applying MCL to linguistic resources is that the hard clustering approach is not appropriate for words that have multiple meanings. In order to overcome this problem, I apply the hard clustering method without the hub nodes which are identified by taking the clustering coefficient as a threshold. After the MCL process, the resultant crisp clusters are expanded with neighboring nodes to produce overlapping clusters that include in the hub nodes.

3. Building the Social Networks of Gospels

In recent years, several notable studies have sought to gain insights into the interactions between the people and places in the Bible through social networks. One impressive study by Chris Harrison (2007) has produced a number of graph visualizations for the King James Bible. Similarly, The English Standard Version Bible Blog (ESV Blog, 2007) has reported on mapping the social networks of the New Testament using Many Eyes, which is IBM's new online data visualization site. The corpus used in the present study is the Greek version of the Gospels (Nestle-Aland, 1979) and the data consists of names and places and a set of nouns such as son, father, mother that are important in representing the characteristics and the roles of people in the stories. In creating four social networks from the books of Mark, Matthew, Luke and John, co-occurrence data is computed for pairs of words that appear in the same verse (sentence) and for which the clustering coefficient value is more than two. Table 1 shows the number of co-occurrence words represented as nodes to be clustered for each Gospel network. The four average clustering coefficients for the total nodes are all quite high values of 0.5 or more, indicating strong connectedness between nodes, which is a characteristic of small-world networks.

4. Graph Clustering Results

Table 2 presents the numbers of expanded nodes after the soft clustering process, MCL cluster sizes, and the cluster

sizes that are included within the "Jesus" cluster. The numbers of "Jesus" clusters are indicative of this character's central importance, and each cluster represents a series of relationships and events around Jesus, such as the temptation and the passion stories. Comparing between clusters across the four social networks illustrates the different roles associated with characters in the Gospels, such as Jesus as the son of God in Mark and Jesus as Savior in Luke. The clustering results for the social networks are implemented as graph network structures developed by the Prefuse interactive data visualization toolkit, as shown in Figure 1. The co-occurrence nodes are connected when they appear in the same verse and the grouping cluster is identified by the same color.

Table 1: Network features

Mt: Matthew, Mk: Mark, Lk: Luke, Joh: John

	Mt	Mk	Lk	Joh
Number of nodes for MCL (in total)	258 (329)	179 (218)	257 (352)	146 (193)
Average clustering coefficient	0.58	0.55	0.57	0.52

Table 2. MCL results

	Mt	Mk	Lk	Joh
Number of nodes expanding with neighboring nodes	276	189	285	160
MCL cluster sizes	44	35	41	22
Clusters included in "Jesus"	10	13	16	19

Figure 1. visualization of clustering

5. Conclusions

In conclusion, I have proposed a soft clustering method combining the clustering coefficient and Markov Clustering in order to determine multiple clusters, especially for nodes that are connected to numerous other nodes. Examining social networks opens up an effective approach to exploring the interactions between characters and the features that underlie word groups within the Gospels. Although some issues remain to be solved, such as the treatment ambiguous names and places, this social network approach may provide significant clues for context analysis in furthering biblical studies.

References

S. van Dongen. (2000). *Graph Clustering by Flow Simulation*, PhD thesis, University of Utrecht.

B. Dorow, D. Widdows, K. Ling, J. Eckmann, D. Sergi, and E. Moses. (2005). Using Curvature and Markov Clustering in Graphs for Lexical Acquisition and Word Sense Discrimination, *Proceeding of 2nd Workshop organized by MEANING Project (MEANING-2005)*.

M. Steyvers, and J. B. Tenenbaum. (2005). The Large-Scale Structure of Semantic Networks: Statistical Analyses and a Model of Semantic Growth, *Cognitive Science*, 29 (1): pp.41-78.

Gfeller, D., Chappelier, J.-C., and De Los Rios, P.. (2005). Synonym Dictionary Improvement through Markov Clustering and Clustering Stability, *International Symposium on Applied Stochastic Models and Data Analysis*, pp. 106-113.

Maki Miyake (2008). Investigating word co-occurrence selection with extracted sub-networks of the Gospels Employing Clustering Coefficients, *Digital Humanities 2008*, pp.258-260.

ESV Blog. (2007). Mapping New Testament Social Networks, <http://www.esv.org/blog/2007/01/mapping.nt.social.networks>

Chris Harrison. (2007). Visualizing the Bible, <http://www.chrisharrison.net/projects/bibleviz/index.html>