

福音書ソーシャルネットワークにおける コミュニティ構造の考察

三宅 真紀

大阪大学 言語文化研究科

本研究では、ソーシャルネットワーク分析を新約聖書の4福音書に適用して、登場人物や地名の共起情報から作成した福音書ソーシャルネットワークから、物語に登場する人物の相互関係や福音書間のコミュニティ構造の特徴について考察する。コミュニティ構造の抽出手法としては、ハードクラスタリング手法であるマルコフクラスタリングとクラスタリング係数を組み合わせて、1単語が複数のクラスターに属することを可能とするソフトグラフクラスタリング手法を実現する。ネットワーク構造の体系を把握することを目的として、得られたクラスタリング結果は、ネットワークの状態をインタラクティブに視覚化できる web アプリケーションとして実装する。

Thoughts on the Community Structures within Social Networks of the Gospels

Maki Miyake

Graduate School of Language and Culture
Osaka University

This paper reports on the application of some network analyses to the four social networks of the Gospels of the New Testament constructed based on co-occurrences between individuals and locations. In particular, the paper discusses the interactions between story characters and features of the community structures. In order to detect community structures within the social networks, this study applies a soft graph clustering technique combining the hard clustering of Markov Clustering and the clustering coefficient. The clustering results for the social networks are implemented as graph network structures through the development of a web-based application for interactive data visualization.

1. まえがき

実世界のネットワーク解析法としてのグラフ理論に基づくネットワーク分析は、大規模なデータ体系を直感的に把握する有効性から、辞書をはじめとする言語使用にも適用した応用研究が報告されている[1][2]。新約聖書の福音書を対象にしたネットワーク分析では、ネットワーク基本特性であるクラスタリング係数を活用したデータ処理による部分グラフのクラスタリング結果の精度に関する報告があり[3]、単語の関係性を理解するために、ネットワークによる文書構造の体系的な把握は有効である。

本研究では、4福音書における固有名詞から成るソーシャルネットワークを利用したネットワーク解析から、物語の登場人物の特徴や相互関係を考察する。コミュニティ構造を抽出する方法には、1単語が複数のクラスターに含まれるようにソフトグラフクラスタリング手法を適用する。構造の体系を把握するためのグラフ描画は、ネットワークの状態をインタラクティブに視覚化できる web アプリケーションとして実装する。

2. MCL ソフトグラフクラスタリング

本研究では、ネットワーク基本特性量であるクラスタリング係数とノード間のオーバーラップを許さないハードクラスタリング手法である Van Dongen による MCL (Markov Clustering) [5]を組み合わせて、MCL ソフトクラスタリングを実現する。この手法は Dorow ら[2]による Curvature クラスタリング手法から着想を得たものであり、多数のノードに繋がるハブ的ノードを制御しながらソーシャルネットワークのコミュニティを適切に抽出する。

クラスタリング係数は、Watts と Strogatz [4]が提唱した「知り合いの知り合いが知り合いである確率」を定量的に表した指標である。ノード n に隣接するノードを $N(n)$ と表すとき、 n のクラスタリング係数 $C(n)$ は次のように定義される。

$$C(n) = \frac{\text{隣接ノード間のエッジ数}}{N(n) \times (N(n) - 1) / 2}$$

ここで、クラスタリング係数の値は、0 から 1 の範囲で表される。ある 1 つのノードに他のノードがぶらさがるスターグラフの場合には、そのノードのクラスタリング係数は 0 であり、逆に、あるノードに隣接するノード間が全て隣接している完全グラフの場合は、その係数は 1 である。ソーシャルネットワーク分析では、クラスタリング係数は複数のコミュニティにまたがる曖昧性や多くのノードに繋がっているハブ的役割のノードの抽出するための指標として使用されている。

ハードクラスタリングには、Van Dongen による MCL (Markov Clustering) [5] を適用する。MCL は、ランダムウォークに基づいたシンプルなアルゴリズムであり、グラフの隣接行列から得られた遷移行列を初期値とするマルコフ過程の反復操作により、全体をクリスプな部分グラフに分割したハードなクラスタを抽出する。このようなハードクラスタリング手法は、多義語や多面的役割を持ち合わせている単語に関しては、その中の 1 つのクラスタにしかならずに属することができないため、偏りのあるクラスタが形成されるという問題点を含んでいる。

本研究では、多数のノードに繋がるハブ的役割を持つノードに対して、複数のクラスタに格納するために、MCL ハードクラスタの結果を拡張させてソフトクラスタを生成する。まず、ネットワークのデータ前処理として、クラスタリング係数を閾値としてハブノードを省き、MCL 計算用のネットワークを作成する。低クラスタリング係数を閾値とした場合、ハブ的役割を持ったノードと同時に、低次数のぶらさがりノードが除去されることになる。次に、MCL クラスタリング結果から生成されたハードクラスタから、各クラスタ要素のノードが隣接するノードから 2 つ以上のノードに共通するノードを抽出し、同クラスタとして取り入れる。このようにして、データ処理段階で省かれていたハブ的ノードが、複数のクラスタに格納され、ソフトクラスタを形成することが可能になる。本研究では、各文書ネットワークから形成されるクラスタ間の類似性を考察することを考慮して、低次数のぶらさがりノードをネットワークから予め除去しているが、次数-クラスタリング係数の 2 つのパラメータを閾値として、ハブノードだけを省く処理も可能である。

図 1 に示すような、12 個のノード数から構成される、平均クラスタリング係数が 0.6 とクラスタ性の高い性質を持つ無向グラフを例として、ソフトクラスタリングの手順を示す。まず、ネットワークの隣接行列から得られる MCL の結果は、 $\{1, 2, 3, 5, 6, 7, 10\}$ と $\{4, 8, 9, 11, 12\}$ の 2 つのクラスタに分割される。しかしながら、最大次数であるノード番号 5 に関しては、 $\{4, 8, \dots\}$ のグループにも繋がっており、そのクラスタリング係数は、0.3 と小さい値であり、ノード 5 に繋がるノード間の繋がりが度合い

は小さい。ノード 5 のハブ的性質が MCL の結果に影響して、クラスタ性の低い大クラスタを生み出す結果となっている。

次に、低クラスタリング係数のノード 5 を除いたネットワークを使用して、MCL を行った結果は図 2 のような、 $\{1, 6, 7, 10\}$, $\{2, 3, 5\}$, $\{4, 8, 9, 11, 12\}$ の 3 グループに分割され、ノード 5 を含んでクラスタリングを行った計算と比べて、小さいクラスタが抽出された。

最後に、各クラスタ内で、2 ノード以上と隣接しているノードに関して、クラスタ要素として追加するプロセスを行い、最初に除去したハブノードをクラスタに復活させる。図 3 に、隣接ノードを追加した結果を示す。前段階のハードクラスタは、最終的には、全てのクラスタにノード 5 が含まれた、 $\{1, 5, 6, 7, 10\}$, $\{2, 3, 5\}$, $\{4, 5, 8, 9, 11, 12\}$ の 3 グループに分けられる。

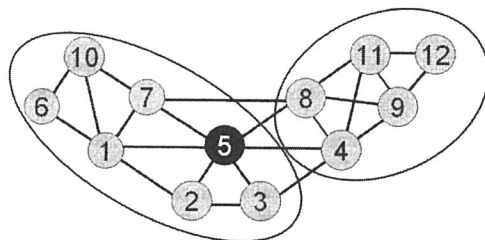


図 1. MCL クラスタリング例

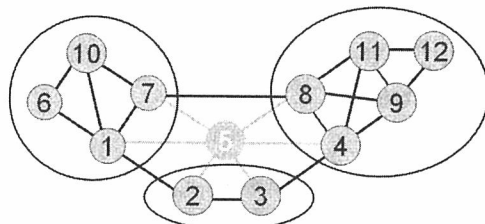


図 2. ハブノードを除去

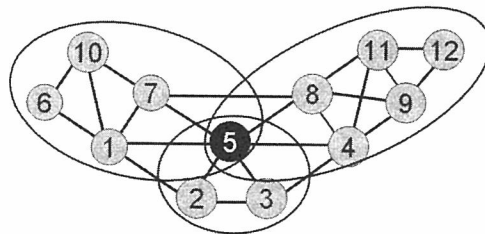


図 3. 隣接ノードを追加

3. ソーシャルネットワークの作成

近年、聖書に描かれた人物の関係性や活動場所を体系的に捉える手段として、聖書からソーシャルネットワーク作成し、視覚的に関係性を捉える研究が盛んになりつつあり、聖書のソーシャルネットワーク研究が重要視されてきている。その代表研究としては、英語訳聖書(King James Version)を利用したChris Harrisonによる聖書ソーシャルネットワークのグラフ視覚化[6]や、IBMのオンラインデータ視覚化サイト”Many Eyes”に実装した新約聖書ソーシャルネットワーク[7]が挙げられる。

本研究では、古典ギリシヤ語の Nestle-Aland[8]をテキストとして使用し、マルコ(Mk)、マタイ(Mt)、ルカ(Lk)、ヨハネ(Joh)の4福音書を対象にして、4種類のソーシャルネットワークを作成する。固有名詞に限定した単語をデータとするソーシャルネットワークに関していえば、原本、翻訳による結果に大きな差異はないと予想されるが、原本である古典ギリシヤ語の格変化を伴う特徴を活かして、「誰が誰に向かって話した」というような有向グラフによる発展研究も可能になるので、先行研究の英語訳版ではなく、古典ギリシヤ語版を使用する。そのため、形態素の未処理の単語をデータとして用いる。使用単語は、固有名詞(人物・場所)と一部の普通名詞に限定し、同じ節内に出現した単語を共起データとしてカウントする。普通名詞に関しては、物語における登場人物が果たす役割や特徴を把握するための重要であると考え、息子、父、教会等の36種類の名詞を選定した。

表1. 福音書ソーシャルネットワーク

	Mt	Mk	Lk	Joh
ノード数	329	218	352	193
平均最短パス	3.7	3.4	3.9	3.3
平均次数 (全ノード数に対する割合)	5.1 (2%)	5.9 (3%)	5.3 (2%)	6.96 (5%)
平均クラスタリング係数	0.58	0.55	0.57	0.52

各福音書ソーシャルネットワークの基本特徴を把握するために、各ネットワークにおける最短パス、次数、クラスタリング係数の平均値をそれぞれ表1に示す。各ネットワークの基本特性量は、同様な傾

向を示しており、平均最短パスに関しては、ミルグラムが検証した「6次の隔たり(Six Degrees of Separations)」[9]よりも短い3~4次長である。また、クラスタリング係数の平均値は0.5以上と非常に高い値であり、ノード間の結びつきの高さを表しており、これら2つの特性量の結果から、スモール・ワールド性質を持っていることが分かる。

平均次数に関しては、各ネットワークとの約5~7のノードと隣接していることを示しており、繋がり密度を全ノード数に対して占める割合として示すと、2~5%と非常にスパースなネットワークであることが分かる。

図4に、ノード数が329個のマルコ福音書から作成されたソーシャルネットワークの次数-クラスタリング係数を分布したものを表す。グラフ図の右下にプロットされた最多次数・低クラスタリング係数のノードは「イエス(ιησους)」であり、物語の中心人物として多数のノードと繋がりを持つハブ的ノードであることが分かる。次に、高次数・高クラスタリング係数を持つノードとして、「ヨセ(ιωσητος)」ノードを例に取る。この単語は、他の文書では登場しない人物であり、マルコ文書の中でも3回しか出現しない(6:3, 15:40, 15:47)、低頻度の単語ではあるが、イエスの兄弟として、「マリア」、「ヤコブ」といった特定の人物と常に出現されるため、高クラスタリング係数を作り出していることが読み取れる。また、低次数の特徴を持つノードは、その繋がり少なからず高クラスタリング係数をとる傾向にあるが、例外的に低クラスタリング係数を持つノードも観察された。例えば、イエスに敵対する役割として現れる「ファリサイ人(φαρισαιου)」は、マルコ文書中には8回(2:18, 2:24, 3:6, 7:1, 7:3, 7:5, 8:11, 10:2)出現され、様々な登場人物と関わりがあるため、クラスタリング次数の値が0.14と非常に小さい。

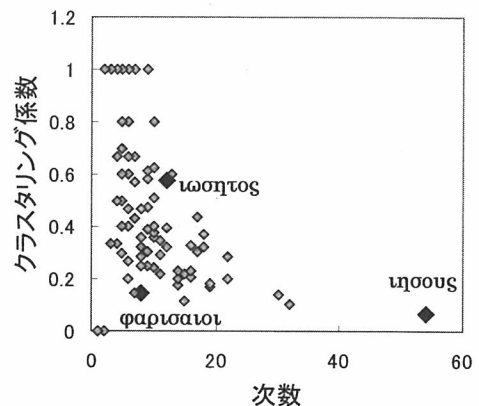


図4. 次数-クラスタリング係数分布 (Mk)

4. グラフクラスタリング結果

前節で作成した福音書ソーシャルネットワークは、「イエス」ノードをはじめとする高次数-高クラスタリング係数で表されるハブノードの影響を大きく受けるため、そのままグラフクラスタリング計算を行うと大クラスタに集約されてしまい、適切なコミュニティの抽出が困難である。本研究では、クラスタリング係数を閾値として設定し、値が 0.2 以上のノードを抽出して、ネットワークを再構築し、2 節で説明した MCL ソフトクラスタリングを適用する。

表 2 に、(1) 閾値をもとに再構築したノード数、(2) MCL の結果からさらに隣接ノードに拡張して得られたノード数、(3) MCL ソフトクラスタリングの結果のクラスタ数、(4) 「イエス」ノードを含むクラスタ数を各ネットワークについて示す。複数のクラスタにまたがる単語としては「イエス」が最も多く、この「イエス」を含んでいるクラスタの数を併記する。「イエス」クラスタの占める割合の大きさは、登場人物「イエス」が物語の中心的存在であることを表している。

図 5 は、表 2 の各文書における (2) と (3) のデータをもとにして、初期ネットワークノードから MCL クラスタリングから抽出されたクラスタ数の推移をプロットしたものである。

表 2. MCL ソフトクラスタリング結果

	Mt	Mk	Lk	Joh
MCL 計算用ノード数	258	179	257	146
ソフトクラスタリング後のノード数	276	189	285	160
MCL クラスタ数	44	35	41	22
「イエス」クラスタ数	10	13	16	19

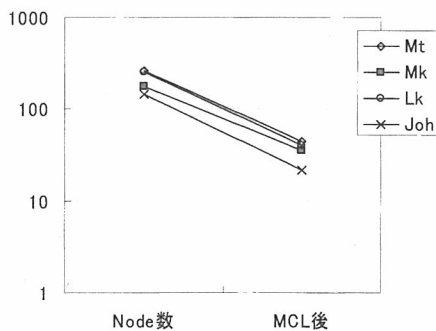


図 5. クラスタ数の推移

4 福音書間のクラスタの要素を比較するにあたり、Gfeller ら[15]が導入した類似度指数を適用する。2つのクラスタ (C_i, C_j) 間の類似度 S に関して、次式で示される Jaccard 係数を用いて計算した。

$$s_{ij} = \frac{C_i \cap C_j}{C_i \cup C_j}$$

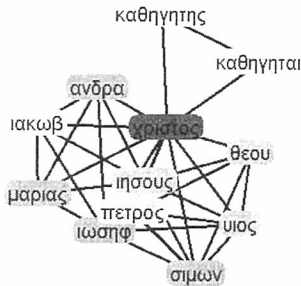
ここで、2つのクラスタが完全に一致する場合は、その類似度は 1 であり、共通する要素が一つもない場合は 0 である。2つのネットワーク A, B のクラスタ間の類似度に関しては、ネットワーク A の個々のクラスタに対して、ネットワーク B の全クラスタの類似度の最高値を取得し、それらの平均値をクラスタ間の類似度とする。

表 3 に、各ネットワーク間におけるクラスタの平均類似度を算出した結果を示す。クラスタ間の類似度値から、マタイとルカのクラスタが共通している要素で構成されており、マルコ、マタイ、ルカのいわゆる「共観福音書」間の類似度は約 0.2 であり、ヨハネに対する値と比較すると高い値を示している。これは、登場人物の構成からも共観福音書がヨハネ福音書と比較して、共通したコミュニティを形成していることが分かり、一方でヨハネ福音書は独自の構成をもつことが分かる。

表 3. クラスタ間の類似度

	Mt	Mk	Lk	Joh
Mt	1	0.198	0.215	0.128
Mk	-	1	0.195	0.125
Lk	-	-	1	0.109
Joh	-	-	-	1

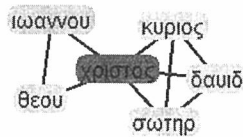
表 4 に、共通部分が多いマタイ・マルコ・ルカの 3 福音書において、クラスタ間の類似度指数の値が高い結果の一例を示す。各文書間のクラスタの Jaccard 類似度係数は、それぞれ、 $S(\text{Mt}, \text{Mk})=0.6$, $S(\text{Mt}, \text{Lk})=0.4$, $S(\text{Mk}-\text{Lk})=0.5$ と高い値を示している。クラスタの要素から、「カエサルのはカエサルに…」で始まるイエスの有名な言葉を容易に思い浮かべることができる(表 5)。この節は、イエスの思想を表し、反対派がイエスを十字架につけた要因と考えられる重要な箇所である。このように、福音書のクラスタ間を比較することによって、登場人物間のコミュニティの考察に留まることなく、福音書の中で重要なテーマを追うことも可能である。



(a) マタイ



(b) マルコ



(c) ルカ



(d) ヨハネ
図 7. 部分グラフ

謝辞

本研究は、科学研究費補助金若手研究（B）19700238 の支援を受けて行ったものである。言語データへのグラフクラスタリング手法の応用研究に関して、東京工業大学・赤間啓之准教授ならびに鄭在玲さんから有益な助言を頂いたことに心より感謝いたします。

参考文献

- [1] Steyvers, M., Tenenbaum, J., The Large Scale Structure of Semantic Networks: Statistical Analyses and a Model of Semantic Growth, *Cognitive Science*, 29 (1) pp.41-78, 2005.
- [2] Dorow, B. et al., Using Curvature and Markov Clustering in Graphs for Lexical Acquisition and Word Sense Discrimination, *MEANING*, 2005.
- [3] 三宅真紀、グラフクラスタリングに基づく共視福音書意味ネットワークの実装、*人文科学とコンピュータシンポジウム論文集 じんもんこん-2006*、pp.161-165、2006.
- [4] Watts, D. and Strogatz, S., Collective dynamics of 'small-world' networks, *Nature*, 393:440-442, 1998.
- [5] Van Dongen, S., *Graph Clustering by Flow Simulation*. PhD thesis, University of Utrecht, 2000.
- [6] ESV Blog. (2007). Mapping New Testament Social Networks, <<http://www.esv.org/blog/2007/01/mapping.nt.social.networks>>
- [7] Chris Harrison. (2007). Visualizing the Bible, <<http://www.chrisharrison.net/projects/bibleviz/index.html>>
- [8] Nestle-Aland, *Novum Testamentum Graece 26th edition*, German Bible Society Stuttgart, 1979.
- [9] Milgram, Stanley: *Psychology Today*, pp.60-67, 1967.
- [10] Gfeller, D., et al., Synonym Dictionary Improvement through Markov Clustering and Clustering Stability, *ASMDA*, 106-113, 2005.
- [11] 新約聖書翻訳委員会（佐藤研訳）、「新約聖書」、岩波書店、2004.