

オープンメディアを活用した知識コミュニティのデザインに関する一考察

Implications for Designing a Knowledge Community using Open Media

大平 雅雄(おおひら まさお・Masao Ohira)¹、松本 真佑(まつもと しんすけ・Shinsuke Matsumoto)²、
伊原 彰紀(いはら あきのり・Akinori Ihara)³、松本 健一(まつもと けんいち・Ken-ichi Matsumoto)⁴

¹奈良先端科学技術大学院大学情報科学研究科 助教

²奈良先端科学技術大学院大学情報科学研究科 博士後期課程

³奈良先端科学技術大学院大学情報科学研究科 博士前期課程

⁴奈良先端科学技術大学院大学情報科学研究科 教授

{ ¹masao, ²akinori-i, ³shinsuke-m, ⁴matumoto } @ is.naist.jp

[Abstract]

In this paper we try to provide useful implications for designing a knowledge community using open media, based on a comparative analysis of large-scale knowledge networks in three knowledge communities. As analysis targets, we selected three different kinds of knowledge communities (Yahoo!Q&A community, Apache HTTP Server community, and Enron community). Yahoo!Q&A community is a knowledge community for knowledge-sharing through interactions among participants on open media, while Apache community is a traditional online knowledge community for developing open source software (knowledge-creation). Enron community is a knowledge community in a corporation without open media. We analyzed large-scale knowledge networks in the three knowledge communities by using social network analysis. As the results of our analysis, we have found that characteristics of each knowledge community reflected on structural features in the knowledge networks.

[キーワード]

知識コミュニティ、知識ネットワーク、社会ソーシャルネットワーク分析、時系列分析

1. はじめに

Web2.0 と呼ばれる言葉が象徴するように、ウェブ上での情報の流通形態が近年大きく変化してきている。Blog、SNS、Wikiなどの新しい情報（知識）発信・共有技術は、CGM (Consumer Generated Media)とも呼ばれ、ウェブを活用して個人が情報（コンテンツ）を手軽に発信するための有力な手段となっている。情報技術により情報能力を高めた個人が爆発的に増加する中で、カリスマブロガーと呼ばれるような、豊富な知識を有し人々に大きな影響を与え得る個人も既に数多く登場してきている。また、Wikipedia¹は不特定多数の個人による質の高い知識共有活動の結果、インターネット上の百科事典として確固たる地位を築いている。このように、CGMをはじめとするオープンメディアを通じた知識共有および知識創造活動が近年広く盛んに行われるようになり、必ずしも質の高い知識が専門家のみによって生み出される状況ではなくなりつつある。本論文では、知識の共有や創造活動を行うコミュニティを「知識コミュニティ」と定義し、知識コミュニティにおける参加者間のインタラクションの分析を行う。特に、それぞれの知識コミュニティの特徴が、参加者のインタラクションによって形成される「知識ネットワーク」とどのような関係があるのかを明らかにしたい。

本論文の構成は以下の通りである。続く2章では、本論文で分析の対象とする知識コミュニティの特徴をまとめる。3章では、知識コミュニティに参加する人々との間のインタラクションをネットワーク構造として捉え、知識の共有と創造がどのような構造によって支えられているのかを観察するために用いた社会ネットワーク分析の方法について述べ、4章において分析結果を示す。5章では、分析によって得られた結果に基づいて今後の知識コミュニティのデザインに関して考察し、最後にまとめと今後の課題について述べ、本論文を結ぶ。

2. 知識コミュニティの特徴

本章では、本論文で扱う知識コミュニティの定義と特徴について述べる。また、本論文で分析の対象とする3つの知識コミュニティについて述べる。

2.1. 知識コミュニティ

本論文では、知識コミュニティを「共通の目的をもって知的協調活動に取り組む人間の集団」と定義する。この定義上、知識コミュニティは研究グループなど従来から存在する集団も当然含むが、本論文では特に、オンラインメディアを活用して知的協調作業に取り組む知識コミュニティを対象とする。先に述べたように、近年オンラインでの質の高い知識共有や知識創造活動が盛んに行われており、現代社会の一部として今後ますます重要な役割を担うと予想される。オンラインの知識コミュニティはすでに多様な形態のものが登場してきている。このような状況の中で、オンラインの知識コミュニティと実社会との関わり方や新たな知識コミュニティを構築する方法を模索するためにも、現在最も成功していると考えられる知識コミュニティについて深く理解することは社会的に意義のあることと考えている。

本論文では、3つの知識コミュニティを取り上げる。1つはQ&A型の知識共有を目的とするYahoo!知恵袋コミュニティである。Q&A型知識共有コミュニティは、オンラインの知識コミュニティの中でも非常に多くの参加者(ユーザ)を集めているコミュニティの1つである。次に、近代型のYahoo!知恵袋コミュニティの対比として、インターネット黎明期から存在している知識コミュニティであるオープンソースコミュニティに着目する。本論文では、古い歴史を持ちオープンソースコミュニティの代表例として考えることのできるApacheコミュニティを取り上げる。さらに、オープンメディアを用いた前述の2つの知識コミュニティの対比として、企業における知識コミュニティの分析を行う。分析対象としては、企業内の知識活動を調べることのできるEnronコミュニティ(Enron Email Dataset¹¹)を取り上げる。表-1は、3つの知識コミュニティの特徴をまとめたものである。以降では、それぞれの知識コミュニティについて述べる。

2.2. Yahoo!知恵袋コミュニティ

Yahoo!知恵袋コミュニティは、2004年4月からヤフー株式会社がサービスを提供している質問・回答型の知識共有コミュニティの1つである。日常の疑問や悩みを質問として投稿する質問者と、その質問に答える回答者からなるコミュニティである。ユーザのコミュニティ参加に対する動機付けの仕組みとして、質問や回答を行うと獲得できる「知恵コイン」や、優れた回答として選ばれた「ベストアンサー」に基づくランキング方式が採用されている。本論文では、国立情報学研究所がヤフー株式会社との契約に基づき研究者に対して提供している2004年4月から2005年10月の間に約24万人のユーザによって投稿された質問と回答の本文約16億文字、質問数311万件、回答数約1347万件のデータを分析の対象とする。Yahoo!知恵袋コミュニティにおけるユーザの参加理由や回答の質などは三浦らの研究チームによってすでに詳細に検討されている[11][12][13][14]。本論文では、Yahoo!知恵袋コミュニティ内の知識ネットワークを社会ネットワーク分析を用いて調べるものであり先行研究とはアプローチが大きく異なるが、三浦らの研究は本論文の分析結果を考察する際の根拠としても参考にできる。

2.3. Apache コミュニティ

Apache (HTTP server) コミュニティは、現在 (2008年8月時点)、世界一の市場シェアを誇るWebサーバ (HTTP

表-1 本論文で分析対象とする知識コミュニティの特徴

	Yahoo!知恵袋 (近代型知識コミュニティ)	Apache (従来型知識コミュニティ)	Enron (企業型知識コミュニティ)
知的協調活動の目的	知識の共有	知識の創造 (ソフトウェア開発)	情報の共有や 組織活動の調整
コミュニケーション メディア	Q&A型の掲示板	メーリングリスト	電子メール
メディアの 開放性	オープン	オープン	クローズド
参加者の特徴	不特定多数の 匿名ユーザ	不特定多数の 開発者	Enron社の 従業員

Server) ソフトウェアを開発しているオープンソースコミュニティである。欠陥の修正パッチを継続的に多数リリースしているため、非常に高品質なソフトウェアとして広く認知されている。世界中に分散しているボランティアの開発者がメーリングリストを主なコミュニケーション手段として用いてソフトウェアの共同開発を行っている。本論文では、開発者メーリングリストを対象とし、2001年11月から2007年9月までの間に約12,700人の開発者によってやり取りされたメッセージ約11万件を分析する。Apache コミュニティの開発者は対面でのコミュニケーションをほとんど行わないが質の高いソフトウェアを実現しているため、Apache コミュニティの成功要因を企業における分散開発のヒントとして活用するために多くの研究者が Apache コミュニティを対象とした研究を行っている[6][9][11][15]。

2.4. Enron コミュニティ

Enron 社はかつて、エネルギー関連事業を中心として成長し、全米第7位の年間売上高(2000年度)や2001年には社員数約21,000人を有するなど、米国での有力企業の1つであった。しかしながら、巨額の不正会計処理が明らかとなり2001年12年に破綻した。これをきっかけに、企業の会計不信が広まり世界経済にも大きな影響を与えた。本論文では、連邦エネルギー統制委員会(Federal Energy Regulatory Commission: FERC)が調査期間中に公開したEnron社の管理職約150人分のemailデータに基づき、カーネギーメロン大学のCALO(A Cognitive Assistant that Learns and Organizes)プロジェクトがいくつかの処理を施した後に研究者に対して公開しているデータ(Enron Email Dataset)を分析する。データセットには、約24,000人の社員間でやり取りされた約51万件のemailデータが含まれる。Enron Email Datasetは公開されている企業のメールデータとしては最大のものである。本論文で用いる社会ネットワーク分析アプローチが適用可能であることはHossainらの研究[4]により示されている。

3. 知識コミュニティにおける参加者のインタラクションの分析

本論文では、3つのコミュニティそれぞれにおいて行われる知識共有・創造活動が、コミュニティ参加者の間でどのようなインタラクション(すなわちメッセージのやり取り)によって支えられているのかを理解するための分析を行う。そのために、参加者間のインタラクションをネットワーク構造(知識ネットワーク)として捉え、その構造的特徴を調べるために社会ネットワーク分析[16][17]を用いる。

ネットワークの構造的特徴を知るための指標がこれまで数多く提案されているが、本論文では、代表的かつ基本の特徴量として用いられている「平均経路長」「クラスタリング係数」「次数分布」の3つの指標を用いる。以下ではまず、本論文でのネットワーク分析に関する諸定義を行う。

3.1. ネットワーク分析

3.1.1. ネットワークの定義

本論文では、各点(ノード)とそれぞれを繋ぐ線(エッジ)の集合をネットワークとする。例えば、コミュニティ参加者Aが参加者Bに対してメッセージを送信した場合、参加者A(ノードA)から参加者B(ノードB)にエッジを結ぶ。参加者(ノード)間のメッセージのやり取りをエッジとして表現することによってネットワークを構築しネットワーク分析を適用する。以下、ネットワーク全般の用語(次数、経路、距離)について図-1を用いて説明する。

- 次数: あるノードの持つエッジの数である。例えば、ノード v_1 の次数は2である。
- 経路: あるノードから他のノードへ到達するために通るノードの順路である。例えば、 v_1 から v_6 への経路としては、 $v_1 \rightarrow v_5 \rightarrow v_6$ や $v_1 \rightarrow v_3 \rightarrow v_4 \rightarrow v_5 \rightarrow v_6$ などがある。あるノードから他のノードへ到達できる(経路が存在する)経路を到達可能な経路、到達できない(経路が存在しない)場合を到達不可能な経路と呼ぶ。例えば、 v_1 から v_6 へは到達可能であり、 v_1 から v_7 へは到達不可能である。
- 距離: あるノードから他のノードへの経路の長さである。例えば、経路 $v_1 \rightarrow v_5 \rightarrow v_6$ の距離は2であり、経路 $v_1 \rightarrow v_3 \rightarrow v_4 \rightarrow v_5 \rightarrow v_6$ の距離は4である。 v_i から v_j への経路のうち距離が最も短い経路を最短経路、その距離を最短経路長と呼ぶ。例えば、 v_1 から v_6 の最短経路は $v_1 \rightarrow v_5 \rightarrow v_6$ で、最短経路長は2である。

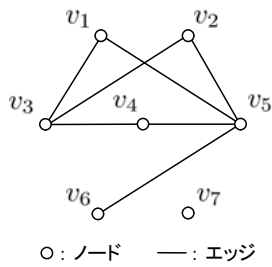


図-1 ネットワークの例

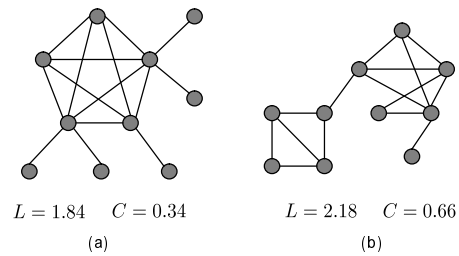


図-2 ネットワーク指標の例

3.1.2. 平均経路長

任意の2ノード間を最短で到達可能なルートを通った場合に経由したノードの数のことを最短経路長と呼ぶ。ノード v_1 についての平均経路長 L_i とは、 v_i と接続している全ノードとの最短経路長の平均のことを指す。さらにネットワークに存在する全てのノードの最短経路長の平均を取った値を、そのネットワークに対する平均経路長 L と呼び、以下の式で定義される。

$$L = \frac{1}{n} \sum_{i=1}^n L_i \quad (1)$$

一般に、この L が小さいネットワークほど、各ノード間の距離が近く情報伝達効率の高いネットワークであるといえる。

3.1.3. クラスタリング係数

ノード v_i についてのクラスタリング係数 C_i は、ノード v_i と接続した2つのノード同士が繋がっている割合であり、3つのノードがエッジで接続された三角系を形成している確率のことである。ネットワーク全体のクラスタリング係数 C は、グラフ全体の n 個のノードの C_i を平均した値であり、 L と同様、以下の式で定義される。

$$C = \frac{1}{n} \sum_{i=1}^n C_i \quad (2)$$

クラスタリング係数 C が大きいネットワークほど、ネットワーク内のノード同士が密に繋がりが合っているため「自分の友人同士が互いに友人である」確率が高くなる。

図-2は、ある2つのネットワークそれぞれに対する平均経路長とクラスタリング係数の例である。図-2(a)は5人の参加者を中心に議論が一つにまとまっているが、図-2(b)では議論が2つの場で発生しており、 L の値は(a)のほうが低い。論文では、 L の値を情報の伝達効率ではなく、参加者の議論が集中している（多人数の関与により議論が分散していない）度合いと見なす。クラスタリング係数 C については、コミュニティの「まとまり度合い（参加者同士がコミュニケーションとお互いの存在を認知している割合）」として解釈する。ただし、本論文で扱う知識ネットワークは、メッセージの送受信関係に基づいて構築し分析を行うため、単純化のために図-1、図-2のような無向グラフではなく、有向グラフを分析対象としている点については注意されたい。

3.1.4. 度数分布

度数分布とは、ネットワークにおけるノードのエッジ数（ノードの度数）の分布のことである。多くの大規模なネットワークは、ランダムグラフが予期するポアソン分布には従わずべき乗則に従い、スケールフリー性^[1]を持つことが明らかになってきている。

3.2. 分析方法

本論文では、Yahoo!知恵袋の質問・回答データ、Apacheのメーリングリストデータ、Enron社のemailデータを用いてネットワーク構造を構築し分析する。以下では、各コミュニティにおけるノードとエッジの説明とノードとエッジの情報の入手方法を述べる。次に、ネットワーク構造の分析に用いる特徴量の時間的な変化を捉える

ために用いた時系列分析の手法、並びに一定期間ごとに構築する各コミュニティのネットワークの構築方法について述べる。そして最後に分析に用いる各特徴量の導出方法について述べる。

3.2.1. 各知識コミュニティにおけるノードとエッジ

Apache コミュニティ、Enron コミュニティではメッセージのやり取りを行う人物らを各コミュニティの参加者と見なすため、各コミュニティの参加者をノードで表す。また、送信者、受信者間で行うメッセージのやり取りをエッジで表す。Apache コミュニティ、Enron コミュニティにおけるノードとエッジの情報は各メッセージのヘッダ情報から入手する。一方、Yahoo!知恵袋コミュニティではメッセージのやり取りを行っているわけではない。しかし、質問に対して回答を行うことで知識共有を実現していると考える。よって、メッセージの送信者を回答者、受信者を質問者とみなし、質問者、回答者をそれぞれノード、質問に対する回答をエッジで表す。

3.2.2. ネットワーク構造の構築

本論文では送信者（又は回答者）のノードから受信者（又は質問者）のノードに向けてエッジを繋ぐことでネットワークの構築を行う。しかし、知識コミュニティへの参加と離脱や参加者間のインタラクションの頻度に依存して、知識ネットワーク内はノードの出現と消失やエッジの増減が時間の経過とともに動的に変化している。すべての期間のデータを用いて1つのネットワークを構築し、その特徴を分析することは適切ではない。そのため、ある一定期間ごとにネットワークを構築し、各指標を算出する必要がある。本論文では、ある一定期間 P ごとに知識ネットワークを構築する際に、知識ネットワークを求める期間 P の開始時期を一定期間（例えば、 $P/2$ ）ずつスライドさせる Sliding Time Method を用いる[8]。

本論文では、オープンソースコミュニティにおける開発者の社会的構造を分析した[5]と同様、知識ネットワークを求める期間 P を3ヶ月、スライドさせる期間を1ヶ月とした。本論文では社会ネットワーク分析を行うために使用される大規模ネットワーク可視化ツール「Pajek」ivを用いて各コミュニティのネットワーク構造を可視化した。(図-5、図-8、図-11)

3.2.3. 定量データの導出

各コミュニティにおけるデータから Sliding Time Method に則って各期間の参加者数を導出し、ノード数の推移を分析する。また、平均経路長、クラスタリング係数、次数分布のそれぞれの特徴量は、Pajek を使用して導出した。

4. 分析結果

4.1. 基本統計量

分析データセットの基本統計量を表-2に示す。Yahoo!知恵袋コミュニティは、分析対象期間が他の2つに比べ最も短い17ヶ月であるが、参加者総数・メッセージ総数はともに最も多いデータセットとなっている。一方、Apache コミュニティは、分析対象期間が3つのコミュニティの中で最も長い71ヶ月（約6年間）あるにもかかわらず、参加者総数・メッセージ総数はともに最も少ないデータセットとなっている。Apache コミュニティは他の2つのコミュニティと比べると参加者総数は少ないが、1つのプロダクト（Apache HTTP sever）を開発している人間（開発者）の総数であることを考慮すると、大規模なオープンソース開発コミュニティであるといえる。

それぞれのコミュニティで行われている知識共有や知識創造活動の種類が大きく異なるため、コミュニティの参加人数やメッセージ総数などからコミュニティ間の比較を単純に行うことはできない。しかし、本論文で分析対象としているデータセットは、それぞれのコミュニティにおいて観察される「社会現象」をある程度の一般性を持って議論するための材料としては十分な規模があると思われる。

表-2 各コミュニティの基本統計量

	分析対象期間			参加者 総数 (人)	メッセージ 総数 (件)
	開始年月	終了年月日	期間 (月)		
Yahoo!知恵袋	2004/04	2005/08	17	242,333	16,593,794
Apache	2001/11	2007/09	71	12,710	112,143
Enron	1998/10	2002/07	46	24,773	517,431

4.2. Yahoo!知恵袋コミュニティ

Yahoo!知恵袋コミュニティのノード（参加者）数、平均経路長、クラスタリング係数の時系列変化を図-3に、次数分布（分析期間：2004年6月、2004年12月、2005年6月）を図-4に、知識ネットワークを図-5に示す。

Yahoo!知恵袋コミュニティのノード数（図-3(a)）はサービス開始（2004年4月）から単調増加しており、コミュニティが順調に拡大している様子が見て取れる（用いたデータセットはコミュニティの発展途上のものである可能性が高い）。一方、平均経路長（図-3(b)）はサービス開始当初からほぼ一定している。ノード数の単調増加に対して平均経路長が一定しているのは、ノード数が増加しても議論が大きく分散することがないためと思われる。ノード数の単調増加とは対照的にクラスタリング係数（図-3(c)）は徐々に減少し、分析期間後期辺りから一定してきている。Yahoo!知恵袋の「質問カテゴリ」ごとに参加者が分散した結果、知識ネットワークが数多くのクラスタを持つ構造になった（お互いの存在を認識していない参加者が存在する割合が高くなった）ためと考えられる。

次数分布（図-4）からは、いわゆるスケールフリー構造¹を見て取ることができる。分析期間前期、中期、後期へと進むにつれて、4,000以上の次数を持つノードの分布確率 $P(k)$ が高くなっており、コミュニティの中で活発に知識共有に関与する参加者が増加していることがわかる。ネットワーク図（図-5）からは、可視化に用いたデータを大幅に削減させていることもあるが、時間経過によるネットワークの構造の変化は視覚的には観察できない。したがって、質問カテゴリごとで参加人数の分布には大きな違いがあると思われるが、質問カテゴリごとの参加人数の分布確率はほぼ一定で推移しているのではないかと予想できる。

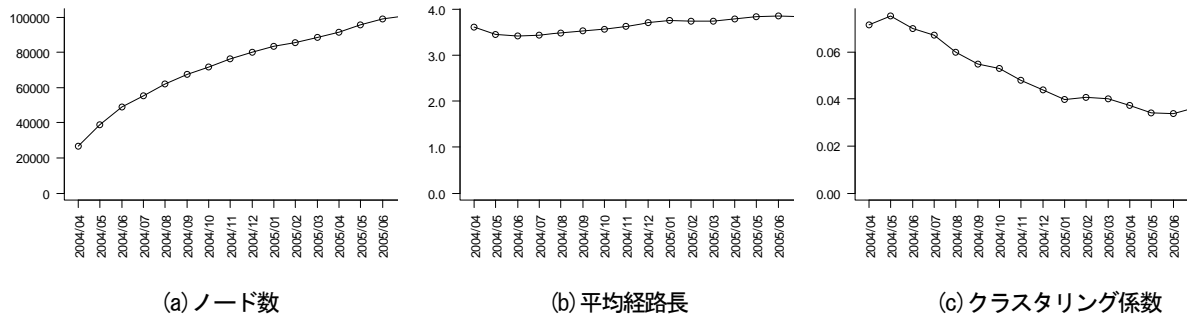


図-3 Yahoo!知恵袋コミュニティにおける各指標値の時系列変化

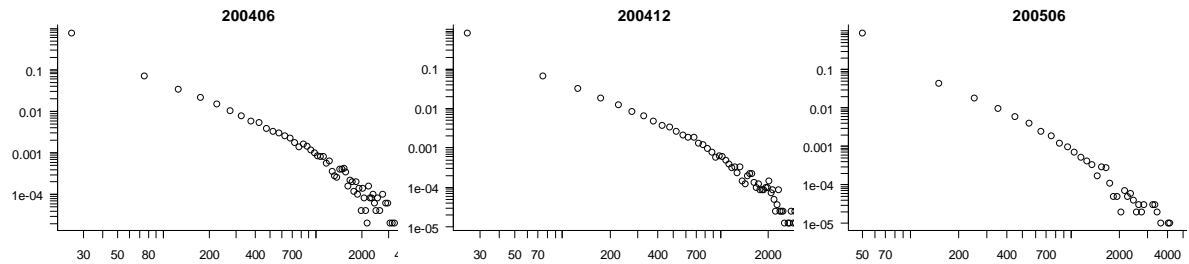


図-4 Yahoo!知恵袋コミュニティ前期・中期・後期における次数分布

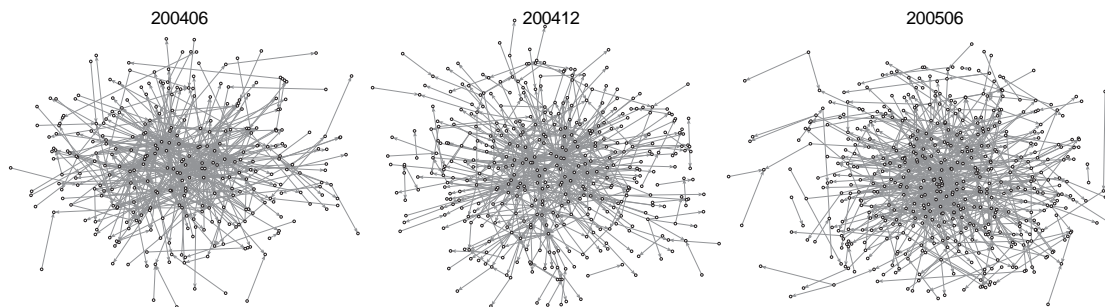


図-5 Yahoo!知恵袋コミュニティ前期・中期・後期における知識ネットワーク

4.3. Apache コミュニティ

Apacheコミュニティのノード(参加者)数、平均経路長、クラスタリング係数の時系列変化を図-6に、次数分布(分析期間:2002年10月、2004年10月、2006年10月)を図-7に、知識ネットワークⁱⁱⁱを図-8に示す。

Apacheコミュニティのノード数(図-6(a))は分析期間前半の2002年5月にピークに達した後、徐々に減少している。プロダクト(Apache HTTP Sever)の機能や品質が成熟するにつれて大きな技術的革新が少なくなってきたことから、新しい技術の習得に興味を持つことのできる開発者[3][7]が徐々に他のオープンソースコミュニティへ流出している可能性が高い。分析期間が長い(71カ月)であることを考えると分析期間中期および後期のApacheコミュニティは成熟段階にあるものと考えることができる。Yahoo!知恵袋コミュニティとは異なり、Apacheコミュニティの平均経路長(図-6(b))は分析時期に値が異なっている。分析期間前期では、ノード数の増加とともに平均経路長の値が大きくなっており、議論が分散していることが窺える。一方、分析期間中期や後期では値が増減している。新しいバージョンのソフトウェアのリリース前後では新機能やバグ報告について様々な議論が活発化するため値が高くなり、ソフトウェアのリリースがない時期には議論すべき対象が相対的に少なくなるため集中した議論が行われているものと推測できる。クラスタリング係数(図-6(c))も平均経路長の結果と同様の傾向が観察できる。

次数分布(図-7)からは、一部のノードが多数の次数を持ち分析期間後期にその傾向が最も顕著に観察できる。ネットワーク図(図-8)からもその傾向を読み取ることができ、分析期間後期には特定のノードを中心としたクラスターが構成されている。Apacheコミュニティの知識ネットワークには中心的なリーダーが数人存在し、ソフトウェア共同作業のコーディネーションを行っているという従来の知見[9][10]とも一致する結果となっている。

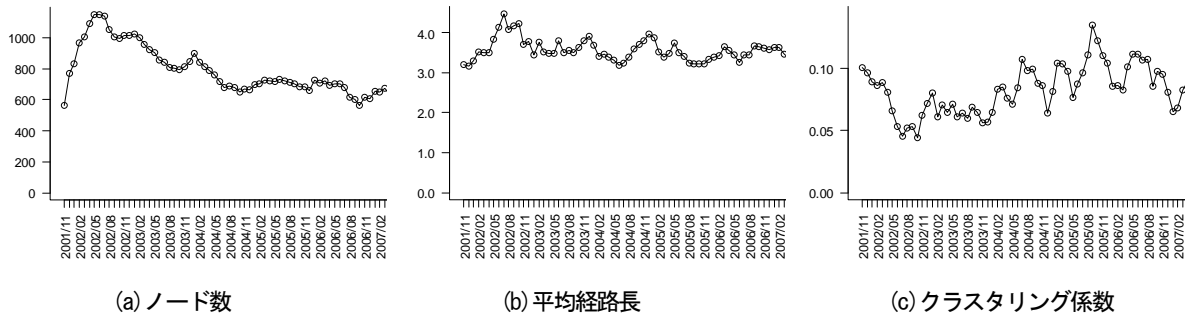


図-6 Apache コミュニティにおける各指標値の時系列変化

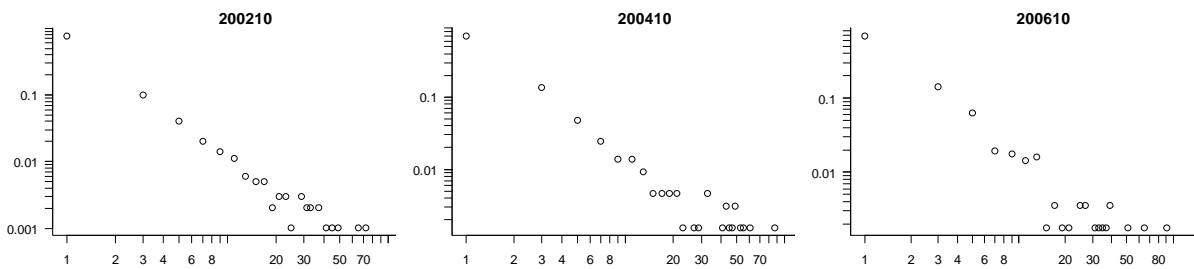


図-7 Apache コミュニティ前期・中期・後期における次数分布

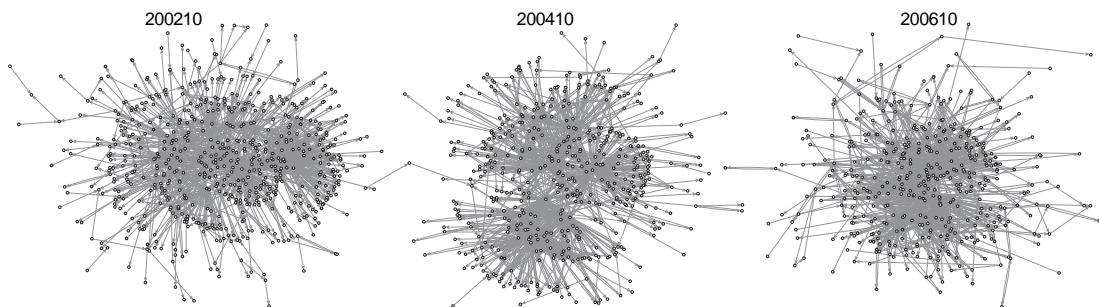


図-8 Apache コミュニティ前期・中期・後期における知識ネットワーク

4.4. Enron コミュニティ

Enronコミュニティのノード(参加者)数、平均経路長、クラスタリング係数の時系列変化を図-9に、度数分布(分析期間:1999年5月、2000年9月、2002年1月)を図-10に、知識ネットワーク⁴⁴⁾を図-11に示す。

Enron コミュニティのノード数(図-9(a))は分析期間前半は中期・後期に比べると非常に少ない。2001年5月にピークに達した後、急激に減少している。2000年度までは米国のトップレベルの企業であったEnron社が、不正会計処理が明らかとなった後2001年12月に破たんした様子がデータにも表れているものと考えられる。平均経路長(図-9(b))は、(電子メールを用いたコミュニケーションが主流でなかったと思われる)分析期間前期にはノード数の増加とともに値が増加しているが、中期以降では値が安定している。クラスタリング係数(図-9(c))も平均経路長の結果と同様の傾向が観察できるが、ここで注目すべきはクラスタリング係数の値の大きさである。全分析期間のクラスタリング係数の平均値は、Enronコミュニティでは0.21となっており、Yahoo!知恵袋コミュニティ(0.05)やApacheコミュニティ(0.09)よりおよそ2倍から4倍の開きがある。Yahoo!知恵袋やApacheコミュニティで用いられている不特定多数が参加するオープンメディアにおける知識ネットワークと比較すると、Enronコミュニティにおける知識ネットワークは、企業活動を営む組織としての「まとまり」が存在することがわかる。

度数分布(図-10)からは、分析期間前半はノード数自体が極端に少ないため当てはまらないが、分析期間中期・後期にはスケールフリー構造が存在することを見て取ることができる。このネットワークにも中心的人物が存在し、企業活動の方向性を制御している人物が存在していたと思われる。特に、ネットワーク図(図-11)からその傾向を顕著に読み取ることができ、分析期間後期には特定のノードを中心とした2つの大きなクラスタが構成されていることを確認できる。

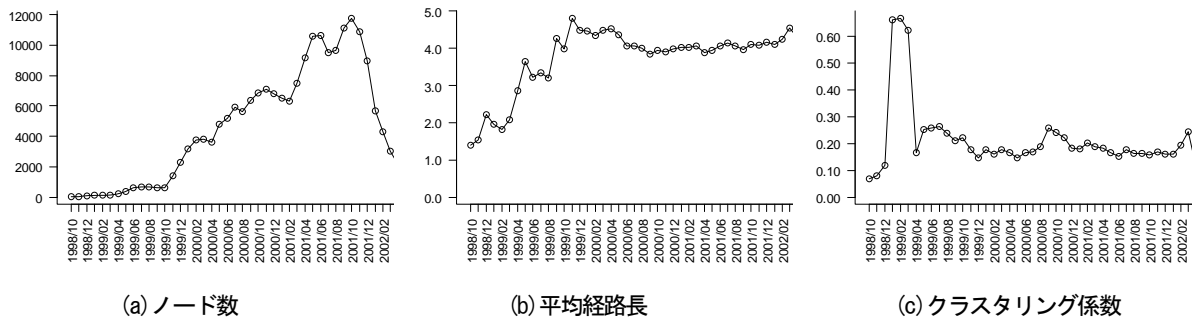


図-9 Enron コミュニティにおける各指標値の時系列変化

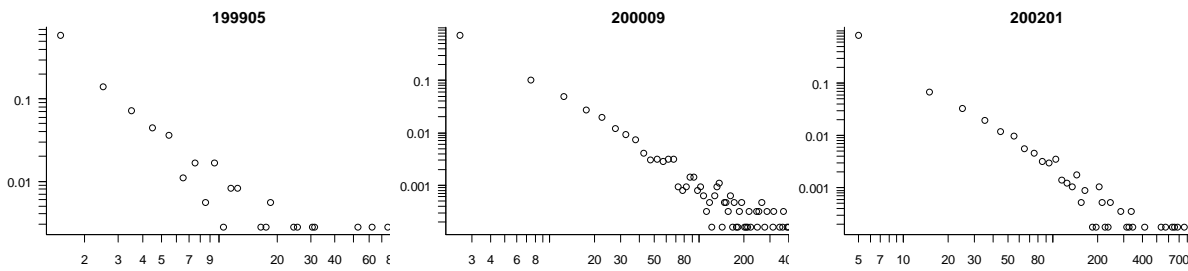


図-10 Enron コミュニティ前期・中期・後期における度数分布

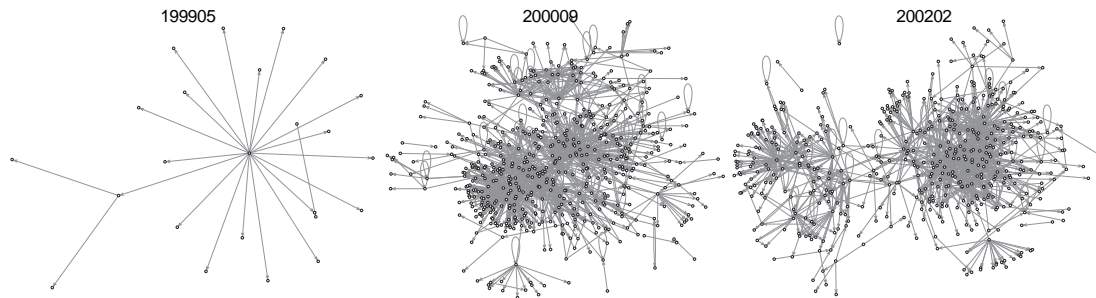


図-11 Enron コミュニティ前期・中期・後期における知識ネットワーク

5. オープンメディアを活用した知識コミュニティのデザインに関する考察

本論文では、種類の異なる3つの知識コミュニティにおける知識ネットワークを、社会ネットワーク分析を用いて横断的に分析した。分析に用いたデータセットは、Yahoo!知恵袋コミュニティの場合は成長段階、Apache コミュニティの場合は成長から成熟段階、Enron コミュニティは成長から終焉段階までを含んだものと思われるため単純な比較はできないが、2章表-1で対比させたような知識コミュニティの特徴の違いが反映された知識ネットワークを確認することができた。得られた分析結果より、オープンメディアを活用した知識コミュニティのデザインについて考察する。

Yahoo!知恵袋コミュニティは発展途上にあるため、今後Apache コミュニティのような成熟段階が訪れることは必然である。人間であるユーザは有限であり、人間が一日に発言できる数も有限であるためである。前例を見ない速度でユーザ数を増加させてきたソーシャルネットワーキングサイト(SNS)のmixiもすでに成長に陰りが見え始めている。ノード数の頭打ちを避けることはできないため、知識コミュニティとしての最終形態がどのようにあるべきかについては、運営者あるいはコミュニティ参加者の共通意識として考えておく必要があると思われる。

現状ではポイントやランキングなどのインセンティブが参加者数の増加や質問および回答数の増加が有効に作用していると考えられるが、コミュニティが飽和状態になった際に同じ仕組みが有効に働くかどうかは現在のところ不明である。目的指向型のより質の高い知識データベースとしてコミュニティを機能させたい場合には、ApacheやEnronコミュニティで見えてきたように、クラスタリング係数が現在よりも高くなるように知識ネットワークを構築する必要がある。しかしながら、Yahoo!知恵袋コミュニティはスケールフリー構造が観察されており、クラスタリング係数を高めることによりさらに大きな「ハブ」となる参加者が出現することが予想される。「ハブ」となって知識コミュニティに中心的に貢献していた参加者が「世話疲れ」などが原因となりコミュニティを脱退すると、コミュニティが一気に崩壊する可能性があることもオープンソースコミュニティの分析により報告されている[9]。コミュニティとしての「まとまり」作りと、「まとまり」を形作るために精力的に貢献する参加者の負荷とのバランスをどのように取っていくかが、知識コミュニティのデザインにおいて今後重要な課題の1つであると言えよう。

6. まとめ

本論文では、「近代型 vs. 従来型」「オープンメディア vs. クローズドメディア」という二つの対比軸を使って、3つの異なる知識コミュニティを横断的に俯瞰した。分析に用いたデータセットはそれぞれのコミュニティの異なる段階(成長期 vs. 円熟期)を捉えているため、厳密なコミュニティ間の比較が行えない点が本論文の制約であり今後の課題である。また、近代型の知識コミュニティにはQ&A型知識コミュニティ以外に種類の異なるものも多く存在しており、その他の知識コミュニティを分析していく必要がある。

謝辞

本論文で用いたYahoo!知恵袋データセットはYahoo!株式会社が国立情報学研究所に提供したものである。本研究の一部は、文部科学省「次世代IT基盤構築のための研究開発」の委託に基づいて行われた。また、本研究の一部は、文部科学省科学研究補助費(若手B:課題番号20700028)および公益信託マイクロソフト知的財産研究助成基金による助成を受けた。

[参考文献]

- [1] Albert-Laszlo Barabasi, *Linked: How Everything Is Connected to Everything Else and What It Means*, Penguin, 2003.
- [2] Gerard Beenen, Kimberly Ling, Xiaoqing Wang, Klarissa Chang, Dan Frankowski, Paul Resnick, Robert E. Kraut, *Using Social Psychology to Motivate Contributions to Online Communities*, Proceedings of the conference on Computer Supported Cooperative Work (CSCW'04), pp.212-221, 2004.
- [3] Paul A. David, Andrew Waterman, Seema Arora, *FLOSS-US: The Free/Libre/Open Source Software Survey for 2003*, available from <<http://www.stanford.edu/group/floss-us/>> (accessed 2007-06-30).
- [4] Liaquat Hossain, André Wu, Kenneth K S Chung, *Actor Centrality Correlates to Project Based Coordination*, Proceedings of the 2006 Conference on Computer Supported Cooperative Work (CSCW'06), pp.363-372, 2006.

- [5] James Howison, Keisuke Inoue, and Kevin Crowston, Social Dynamics of Free and Open Source Team Communications, Proceedings of the 2nd International Conference on Open Source Systems (OSS'06), pp. 319-330, 2006.
- [6] Chris Jensen, Walt Scacchi, Role Migration and Advancement Processes in OSSD Projects: A Comparative Case Study, Proceedings of the 29th International Conference on Software Engineering (ICSE'07), pp. 364-374, 2007.
- [7] International Institute of Infonomics (University of Maastricht) and Berlecon Research GmbH: Free/Libre and Open Source Software: Survey and Study, available from <<http://www.infonomics.nl/FLOSS/report/>> (accessed 2007-06-30).
- [8] Takashi Kakimoto, Yasutaka Kamei, Masao Ohira, Ken-ichi Matsumoto, Social Network Analysis on Communications for Knowledge Collaboration in OSS Communities, Proceedings of the International Workshop on Supporting Knowledge Collaboration in Software Development (KCS'D'06), pp. 35-41, 2006.
- [9] Karim R. Lakhani, Eric von Hippel, How Open Source Software Works: "Free" User-to-User Assistance, Research Policy, Vol. 32, No. 6, pp. 923-943, 2003.
- [10] Shinsuke Matsumoto, Yasutaka Kamei, Masao Ohira, Ken-ichi Matsumoto, A Comparison Study on the Coordination Between Developers and Users in FOSS Communities, Proceedings of Socio-Technical Congruence (STC 2008), CD-ROM-NO. 8, 1-9, 2008.
- [11] 三浦麻子、川浦康至、地福節子、大瀧直子、岡本真、知識共有コミュニティを創り出す人たち、人工知能学会全国大会 (第20回)、3D3-1、2006年
- [12] 三浦麻子、川浦康至、地福節子、大瀧直子、岡本真、知識共有コミュニティを創り出す人たち(2): 「質問タイプ」から見た参加行動、日本社会心理学会 47 回大会論文集、pp. 496-497、2006年
- [13] 三浦麻子、川浦康至、地福節子、大瀧直子、岡本真、知識共有コミュニティを創り出す人たち(3): 「回答者」データから見るコミュニティ内の「知識」、日本社会心理学会 47 回大会論文集、pp. 498-499、2006年
- [14] 三浦麻子、川浦康至、地福節子、大瀧直子、岡本真、知識共有コミュニティを創り出す人たち(4): 形態素解析による「知識」の分析、人工知能学会全国大会 (第21回)、3D3-1、2007年 Audris Mockus, Roy T.
- [15] Fielding, James D. Herbsleb, Two Case Studies of Open Source Software Development: Apache and Mozilla, ACM Transactions on Software Engineering and Methodology (TOSEM), Vol. 11, No. 3, pp. 309-346, 2002.
- [16] John P Scott, Social Network Analysis: A Handbook, SAGE Publications, 2000.
- [17] Stanley Wasserman and Katherine Faust, Social Network Analysis: Methods and Applications, Cambridge University Press, 1994. M

ⁱ Wikipedia: <http://www.wikipedia.org/> 市販の百科事典と比較しても遜色のないレベルの情報量と信頼性を誇っている。

ⁱⁱ Enron Email Dataset: <http://www.cs.cmu.edu/~enron/>

ⁱⁱⁱ さまざまなネットワーク (インターネット、伝染病の感染経路、口コミによる情報伝搬経路など) において近年観察されているネットワーク構造。ネットワーク中のごく少数のノードが膨大な量の次数をもち「ハブ」として機能し、大多数のノードは少数の次数しかもたないようなネットワークの構造的特徴を指す。

^{iv} Pajek: Program for Large Network Analysis: <http://vlado.fmf.uni-lj.si/pub/networks/pajek/>

^v 分析の基本単位期間は3ヶ月であるが、3ヶ月分のデータを用いて可視化するとノード・エッジ数ともに膨大な量になるため関係の把握が困難になる。このため、本論文で示すネットワーク図はデータ量を大幅に削除している。知識ネットワークの大まかな特徴を視覚的に理解するために利用する。

^{vi} さまざまなネットワーク (インターネット、伝染病の感染経路、口コミによる情報伝搬経路など) において近年観察されているネットワーク構造。ネットワーク中のごく少数のノードが膨大な量の次数をもち「ハブ」として機能し、大多数のノードは少数の次数しかもたないようなネットワークの構造的特徴を指す。

^{vii} 分析の基本単位期間は3ヶ月であるが、3ヶ月分のデータを用いて可視化するとノード・エッジ数ともに膨大な量になるため関係の把握が困難になる。このため、本論文で示すネットワーク図はデータ量を大幅に削除している。知識ネットワークの大まかな特徴を視覚的に理解するために利用する。

^{viii} 分析の基本単位期間は3ヶ月であるが、3ヶ月分のデータを用いて可視化するとノード・エッジ数ともに膨大な量になるため関係の把握が困難になる。このため、本論文で示すネットワーク図はデータ量を大幅に削除している。知識ネットワークの大まかな特徴を視覚的に理解するために利用する。