

The silver bullet for OSS communities

Yasutaka Kamei¹, Shinsuke Matsumoto¹, Hirotaka Maeshima¹,
Yoji Onishi¹, Masao Ohira¹, and Ken-ichi Matsumoto¹

Graduate School of Information Science,
Nara Institute of Science and Technology
8916-5 Takayama, Ikoma, Nara 630-0192, Japan
+81-743-72-5312

{yasuta-k,shinsuke-m,hirotaka-m,yoji-o,masao,matumoto}@is.naist.jp

Abstract. Your text goes here. Separate text sections with the standard L^AT_EX sectioning commands.

1 Introduction

Open source software(OSS) have been being used in governments and education institutions, since it is free, and high quality and capacity. Generally, participants(e.g. developers and users) in the community develop OSS through communication with each other on the WWW[?]. There are cases that OSS development is stopped for some factor and end-users are not given continual support, such as bug fixes and functional additions. Therefore, system administrators in a company need to decide carefully whether they should introduce OSS instead of a commercial software system. As OSS becomes widely used in infrastructure, it needs to know whether OSS community continuous activity .

Various studies have reported in order to attempt to reveal the process of development and a success factor in OSS community[?][?]. 例えば, Howison らは OSS コミュニティの成功要因を理解するために, OSS コミュニティにおける開発者グループのインフォーマルな社会的構造 (the informal social structure) を明らかにしようとした [?]. Howison らの分析対象は, 開発者間のコミュニケーションパターンである. 一方で, Raymond は, 成果物 (ソフトウェア) に対するユーザからの反応も, OSS コミュニティの成功要因としている [?]. 新機能追加の要求や, 欠陥の報告といったユーザからの反応は, OSS コミュニティ外からの評価であり, 社会的なニーズに対応しているという実感を開発者に与える [?][?]. つまり, ユーザは成果物の享受者であると同時に, 開発者へのモチベーションの提供者でもある.

我々も, レイモンドと同様, ユーザからの反応が OSS コミュニティの成功要因の1つとして重要であるという立場を取る. そのため, 我々は, 開発者グループだけではなくユーザグループを含めた社会的構造を調べるべきであると考え. ただし, 我々の知る限り, 開発者グループとユーザグループのインフォーマルな社会的構造を明らかにしようとした研究はない.

In this study, we investigate the informal social structure of OSS community by analyzing a communication data between the developers group and the

users group. As an approach to examine the communication data, we use social network analysis(SNA) that views socail relationships between the individual actors[?][?]. In particular, we focus on the betweenness centrality that is a measure to know the state of informal socail structure. By using the betweenness centrality, for example, we can know the community in which each developer has strong connections with others developer, or several participants connects developers group to users group, and so on.

This paper presents a case study to investigate the informal social structure using a communication data changing over time in Apache project that is a large OSS community. We selected Apache project from OSS communities since Apache is the most widely deployed Web server in the Netcraft survey[?] and is very successful in Mockus et al[?].

(分析の手順を入れる.)

In this case study, we calculated 3 centralities and graphed the socail network from the communication data of developers group and users group.

In what follows, Section 2 introducts the related work. Section 3 explains (仮) collaborative development in OSS community, and Section 4 describes the details of analysis method. Section 5 provides our case study, and Section 6 discuss the result of analysis. Section 7 summarizes the paper and presents some future topics of research.

2 Related Work

コーディネータに関する研究 (伊原論文から) エンロン, ほいそん, XXX
Apache に関する研究. Apache Vs Mozilla, XXX, XXX

3 OSS コミュニティにおけるコミュニケーション構造

OSS コミュニティはオンライン上の分散開発環境であるため, Mailing List (ML) や掲示板などのオンラインメディアを通じた非対面のコミュニケーションを基本とする. また, OSS コミュニティの参加者はオンラインメディアを利用し, 役職や年功による序列にほとんど縛られることなく自由に対等な立場で情報交換や議論を行うことができる. ただし, 参加者の役割 (開発者やユーザなど) によって議論したい話題の種類 (新機能についての話題やソフトウェアの使い方についての話題など) が大きく異なる. そのため, 多くの OSS コミュニティでは, 複数の ML や掲示板を用意することでグループ (開発者グループやユーザグループ) が形成されている.

上記の特徴を有する OSS コミュニティ参加者らのコミュニケーション構造は Figure 1 のように表現することができる. 各参加者はそれぞれの役割に応じてグループ内で他の参加者とコミュニケーションを行う. 今後, 開発者グループのみで議論した参加者を P_d , ユーザグループのみで議論した参加者を P_u , 開発者グループとユーザグループで議論した参加者を $P_{d \cap u}$ と呼ぶ.

一般的に、OSS コミュニティが活発に活動し続けるためには、全く指揮系統のない状態ではなく、ある程度の調整を務める人物が重要であるとされている [?], [?], [?]. Figure 1 に示すようなコミュニケーション構造を有する OSS コミュニティでは、開発者グループの議論は開発者グループの中で、ユーザグループの議論はユーザグループの中で閉じられている。そのため、OSS コミュニティ全体の調整には、開発者だけでなく、ユーザだけでもない、開発者とユーザの間に立って調整を務める人物（コーディネータ）の存在が重要であると考えられる。

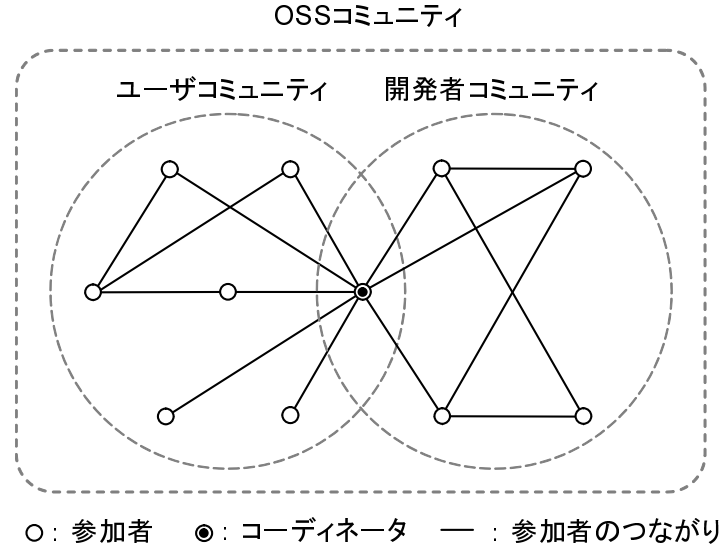


Fig. 1. Organizational structure of an OSS community

4 Analysis Method

4.1 Analysis Procedure

本稿では、開発者グループとユーザグループで議論した参加者 $P_{d \cap u}$ 、特に、両グループを媒介している度合いの大きい $P_{d \cap u}$ が、開発者とユーザの調整を務める人物、コーディネータではないかという考えに沿って分析を行う。媒介している度合いが大きいほど、両グループを強く結びつける $P_{d \cap u}$ であり、そういった $P_{d \cap u}$ ほど、両グループ間の調整を務める可能性が高いためである。我々は、 $P_{d \cap u}$ が両グループを媒介している度合いとして、Freeman の中心性の 1 つ、媒介中心性を用いる。媒介中心性については、Section 4.2 で詳しく説明する。具体的な分析手順を以下に示す。

1. インフォーマルな社会的構造 (the informal social structure) のグラフ化
インフォーマルな社会的構造の大まかな全体像をつかむために、インフォーマルな社会的構造をグラフ化し、俯瞰する。
2. 媒介中心性の算出
両グループを媒介している度合いの大きい $P_{d \cap u}$ を特定するため、それぞれの媒介中心性を求める。
3. メッセージの読解
両グループを媒介している度合いの大きい $P_{d \cap u}$ が、コーディネータであるかを確認するために、媒介中心性の大きい上位 5 名のメッセージログを読む。5 名という人数の設定は、howison らの研究 [?] を参考にした。

4.2 Betweenness centrality

In this paper, we use betweenness centrality of social network analysis proposed by Freeman [?]. We now describe the centrality below.

Betweenness centrality $C_{betweenness}(v_i)$ of a node v_i shows a ratio what is including shortest path between v_j and v_k . Betweenness centrality is shown as Formula (1).

$$C_{betweenness}(v_i) = \frac{\sum_{j < k}^n \sum_{p_{jk}(v_i)}^n}{\sum_{j < k}^n \sum_{p_{jk}}^n} \quad (1)$$

where n shows the number of the network, p_{jk} shows the shortest path from v_j to v_k , $\sum_{j < k}^n \sum_{p_{jk}}^n$ shows the number of path from v_j to v_k , and $\sum_{j < k}^n \sum_{p_{jk}(v_i)}^n$ shows the number of path from v_j to v_k including v_i .

$C_{betweenness}(v_i)$ takes a value from 0 to 1, and a node which takes higher value has a high betweenness centrality. A node which takes higher betweenness centrality plays a role as a coordinator.

In the OSS community, we can assume a node which takes high betweenness centrality as a participants who connected between other participants. Thus, if a participants who takes high betweenness centrality decrease suddenly, communications in the OSS community become decrease.

4.3 調整行動の観点

媒介中心性の大きい $P_{d \cap u}$, 上位 5 名のメッセージを読み, $P_{d \cap u}$ が調整行動を行っているかを確認する. 本稿では, 以下の行動を調整行動とする.

誘導 開発者→ユーザ:

開発者グループで議論されているがユーザグループで議論されるべきトピックに対して, ユーザグループで議論するように誘導している行動を, 調整行動であると判断する. 例えば, 機能追加の要望について議論が続く場合, ユーザグループで議論するべきだと誘導している行動である. 機能の追加については, 利用者の立場から議論するべきで, 開発者グループよりユーザグループで議論されるほうが望ましい.

誘導 ユーザ→開発者:

ユーザグループで議論されているが開発者グループで議論されるべきトピックに対して、開発者グループで議論するように誘導する行動を、調整行動であると判断する。例えば、ある機能で発生する症状がバグではないのかというトピックが発生し議論が続く場合、開発者グループで議論するべきだと誘導している行動である。

伝達 開発者→ユーザ:

開発者グループでのみ知りえる情報を、ユーザグループに紹介する行動を調整行動であると判断する。例えば、次期バージョンがリリースされる時期などの情報を、ユーザグループに告知する行動である。

伝達 ユーザ→開発者:

ユーザグループでのみ知りえる情報を、開発者グループに紹介する行動を調整行動であると判断する。例えば、ユーザグループ内で高評価を得ているバージョンを、開発者グループに伝える行動である。

サポート 開発者→ユーザ:

ソフトウェアの導入やエラーコメントの解釈といった質問に対する $P_{d \cap u}$ のサポートを、調整行動であると判断する。サポートはソフトウェアの関する知識を持つ開発者のみが可能な行動であるため、ユーザグループで発生した質問とそれに対する $P_{d \cap u}$ のサポートのみを考える。

その他:

上記の条件には該当しないが、 $P_{d \cap u}$ が調整していると考えられる行動も、調整行動であると判断する。

上記の観点のいずれかに合致するかは、5人の実験者によって判断された。5人の実験者は、各メッセージを閲覧し、合致する観点到1点投票する。ある観点に対して、3人以上の実験者が合意が得られたメッセージを、調整行動を行っているメッセージと判断した。合意が得られないメッセージは、調整行動を行っていないメッセージと判断した。

5 Case Study

In this section, we present our case study that we analyzed OSS community in perspective of Section ??.

5.1 Apache Community

The Apache HTTP server (the OSS software under consideration in this case study) is a web server software, and has big market share. According to the Netcraft survey, over half of the 150 million or so web sites queried in the Netcraft data collection deployed the OSS software at this writing. In addition, it is widely recognized as high quality and functionality software because of frequent development and maintenance. For this reason, Apache HTTP server is very successful.

5.2 Apache Data Source

We used the following archival source of data.

Developer mailing list: Anyone with an interest in working on Apache development can join developer mailing list, which archived monthly. It contains many different sorts of messages, including technical discussions, proposed changes, and automatic notification messages about changes in the code and problem reports since the period starting March, 1995.

User mailing list: Anyone with an interest in Apache software(???) can join this which archived monthly. It contains many different sorts of messages, including problem reports, *** and *** since the period starting November, 2001.

Developer mailing list で公開されているメール数は XXX 件で, User mailing list で公開されているメール数は XXX 件であった. そのため, 本論文では, Version2.2.0 がリリースされた前後 3 か月分のメーリングリストを分析した.

5.3 Result of analyses

インフォーマルな社会的構造のグラフ化 Apache コミュニティのインフォーマルな社会的構造の大まかな全体像をつかむために, インフォーマルな社会的構造をグラフ化し, 俯瞰した. A communication network for the period released of Apache Ver 2.2.0 is shown in Figure 3. Developers group is on the left, users group is on the right and coordinators group is on the center.

XXX のグラフは, 開発者グループとユーザグループを媒介している $P_{d \cap u}$ が存在していることを示している. なお, P_d の人数は XXX 名で, P_u の人数は XXX 名で, $P_{d \cap u}$ の人数は XXX 名であった.

調整行動の確認 $P_{d \cap u}$ それぞれの媒介中心性を求め, 媒介中心性の大きい上位 5 名のメールを閲覧した. リリース時期における媒介中心性の大きい上位 5 名の各統計量は, Table 1 の通りである.

Table 1. Degree of top 5 coordinators. (リリース時期)

	$P1_{d \cap u}$	$P2_{d \cap u}$	$P3_{d \cap u}$	$P4_{d \cap u}$	$P5_{d \cap u}$
媒介中心性 (Median : 0.0005 Mean : 0.0102)	0.0951	0.0636	0.0439	0.0264	0.0155
メール数 (Median : 12.0000 Mean : 52.4444)	321	196	261	127	119
調整活動メール数	XXX	XXX	XXX	XXX	XXX
開発者との次数	16	30	34	19	28
ユーザとの次数	89	33	29	21	1

Table 1 に示すとおり, 媒介中心性の大きい上位 5 名は調整行動を取っていた. 書面の都合上, 一部を抜粋し, 以下に示す.

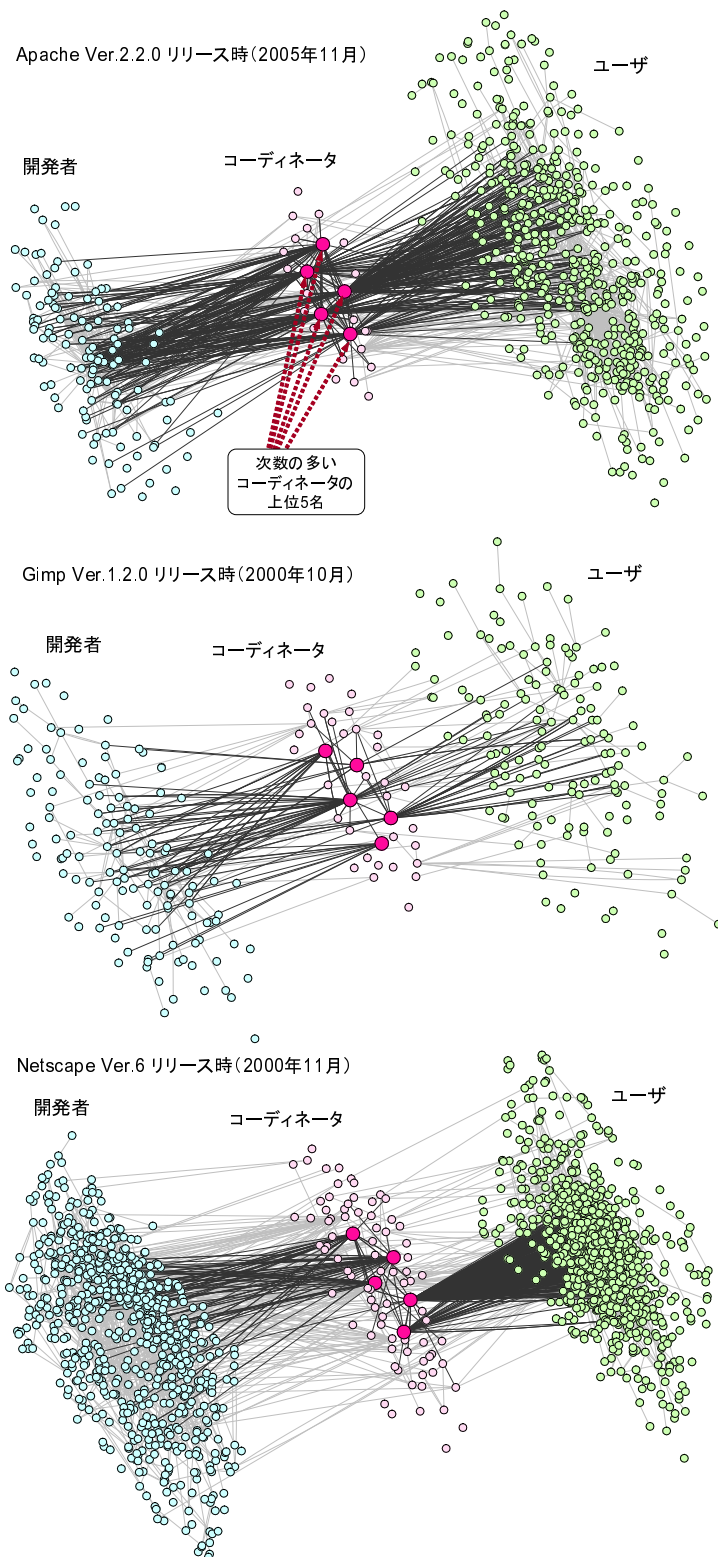


Fig. 2. Communication network of during one period of time.

$P1_{d \cap u}$ による調整行動:

誘導 ユーザ→開発者: あるユーザ P_u が, Apache のモジュールを Windows 上で実装したが Linux 上で動作させたいというメールをユーザグループに送った. そのメールに対して, $P1_{d \cap u}$ は開発者 ML で質問するべきだと, ユーザ P_u をユーザ ML から開発者 ML に誘導していた.

伝達 開発者→ユーザ: あるユーザ P_u が, 今後の開発状況が知りたいので質問のメールをユーザグループに送った. そのメールに対して, $P1_{d \cap u}$ は開発者はそもそもタイムラインを見ずに, 開発者同士の同意によってリリースしていることをユーザ P_u に伝えていた. また, $P1_{d \cap u}$ は, 次期バージョンがリリースされる時期を伝えた.

サポート: あるユーザ P_u が, 正規表現や API の使い方がわからないという質問のメールをユーザグループに送った. そのメールに対して, $P1_{d \cap u}$ は, そのたびに, 質問に答えたり, FAQ のページを紹介していた.

 $P2_{d \cap u}$ による調整行動:

誘導 開発者→ユーザ: ある開発者 P_d が, CGI の高速化の計画に関するメールを開発者グループに送った. そのメールに対して, $P2_{d \cap u}$ は, 要望に関するメールはユーザグループで議論するほうが適切だと, 開発者 P_d を開発者 ML からユーザ ML に誘導していた.

 $P3_{d \cap u}$ による調整行動:

伝達 ユーザ→開発者: ある開発者 P_d が, 一部のバージョンの人気のないというメールを開発者グループに送った. そのメールに対して, $P3_{d \cap u}$ は, 少なくとも個人的なフィードバックがあるのはもちろんのこと, ユーザグループでも指示を得ていることを開発者 P_d に伝えていた.

その他: ある開発者 P_d が, バージョンのリリース発表をユーザグループに送った. しかし, それ以外のコミュニティにも送っていないことに対して, $P3_{d \cap u}$ がその他にも告知するべきだと注意していた.

 $P4_{d \cap u}$ による調整行動:

誘導 ユーザ→開発者:

 $P5_{d \cap u}$ による調整行動:

誘導 ユーザ→開発者:

6 Discussion

XXX のグラフを, 媒介中心性の高い 5 人にのみ対象をしばってグラフ化した. Nodes and edges of top 5 coordinators of degree are highlighted.

グラフと, テーブルの次数を確認すると, 5 人のうち, 5 番手のやつを除いた 4 人は, 開発者グループとユーザグループをバランスよく媒介していることがわかる.

なんで見たのか.

Changes of statistic values(the number of participants and messages) and (→これだと, 3つの中心性をまとめた平均値?) each mean of centralities(degree, betweenness and closeness) for the period of analyses are shown in Figure 4. The vertical axis shows the time and horizontal axis shows statistic values or

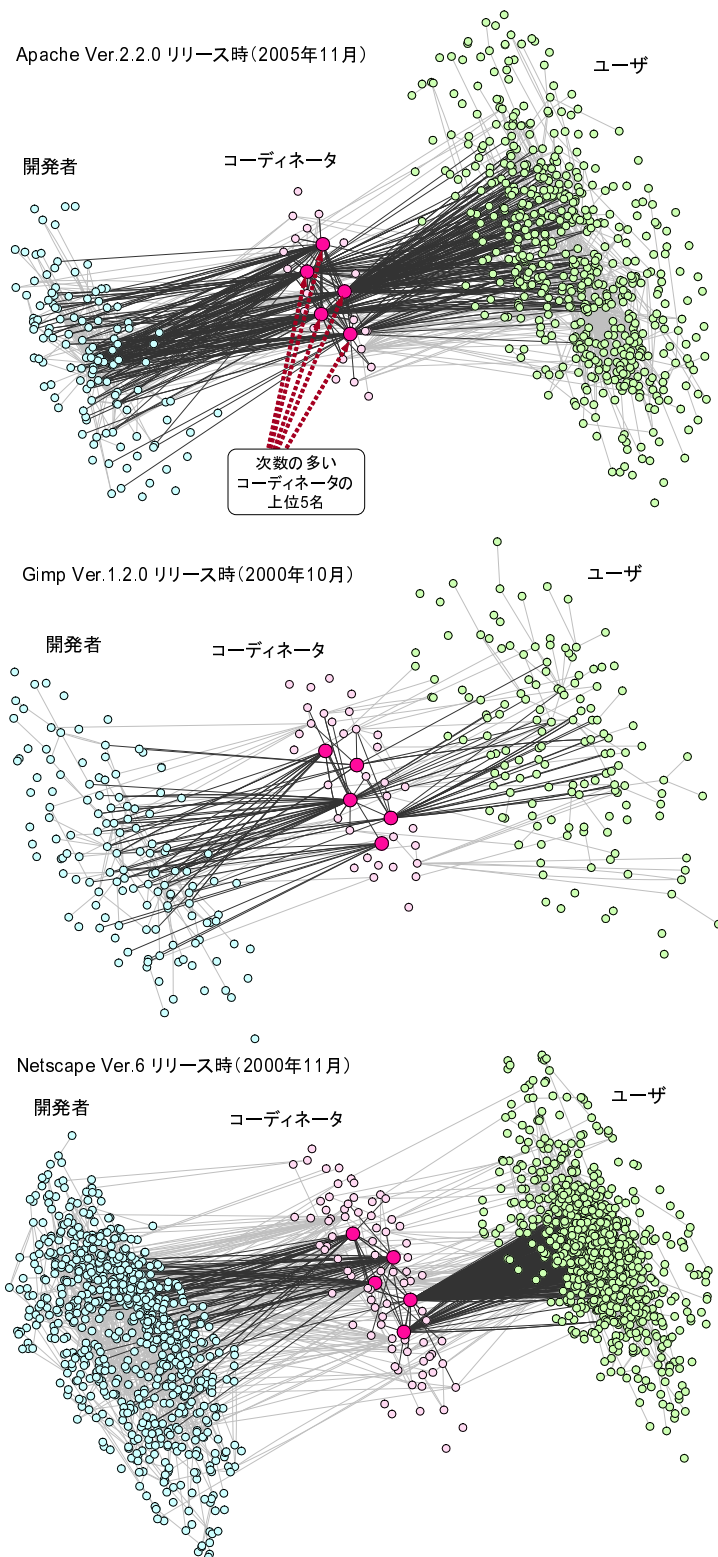


Fig. 3. Communication network of during one period of time.

3 centralities. The vertical dashed line shows the released period as a major version.

3ヶ月窓の話を若干入れる。

今回分析した期間 (XXXX-XXXX) においては, あるアンドマンが常に順位が高い状態であった。これは, Howison らが述べている, リーダーは入れ替わらないという結果と反しない [?].

成功コミュニティである Apache の媒介中心性の上位 5 人中 XXX は, 調整行動を行っていた。媒介中心性の高い人をモニタリングするなどの利用用途が考えられる。

なお, 本論文では, Apache コミュニティのデータセットのみを用いたケーススタディの結果であることに注意する必要がある。結果の信頼性を向上させるために, 他の OSS コミュニティのデータセットを用いることは, 今後の課題となる。また, メール加工アルゴリズムによる制約はある。

7 Conclusions

本論文では, OSS コミュニティのイノベーションのプロセスを明らかにするために, 2つの参加者コミュニティが形成するコミュニケーションネットワークに対して中心性分析を行った。3つの著名な OSS コミュニティケーススタディにより得られた知見は以下の通りである。

- OSS コミュニティが活発に活動するためにはコーディネータの存在が重要である。
- 成功コミュニティでは2つの参加者コミュニティをバランスよく媒介するコーディネータが存在する。
- 成功コミュニティは時間の経過に伴いコミュニケーション構造が過密化する。

また, コーディネータの重要性を調べる方法としては, 3つの中心性の指標のみでは不十分であることが分かった。本論文ではコーディネータの次数を確認することでその重要性を示したが, コーディネータの媒介能力を示すことのできるモデル及び指標の構築が今後の課題である。

References

1. Kajan E (2002) Information technology encyclopedia and acronyms. Springer, Berlin Heidelberg New York
2. Broy M (2002) Software engineering – From auxiliary to key technologies. In: Broy M, Denert E (eds) Software Pioneers. Springer, Berlin Heidelberg New York
3. Che M, Grellmann W, Seidler S (1997) Appl Polym Sci 64:1079–1090
4. Ross DW (1977) Lysosomes and storage diseases. MA Thesis, Columbia University, New York

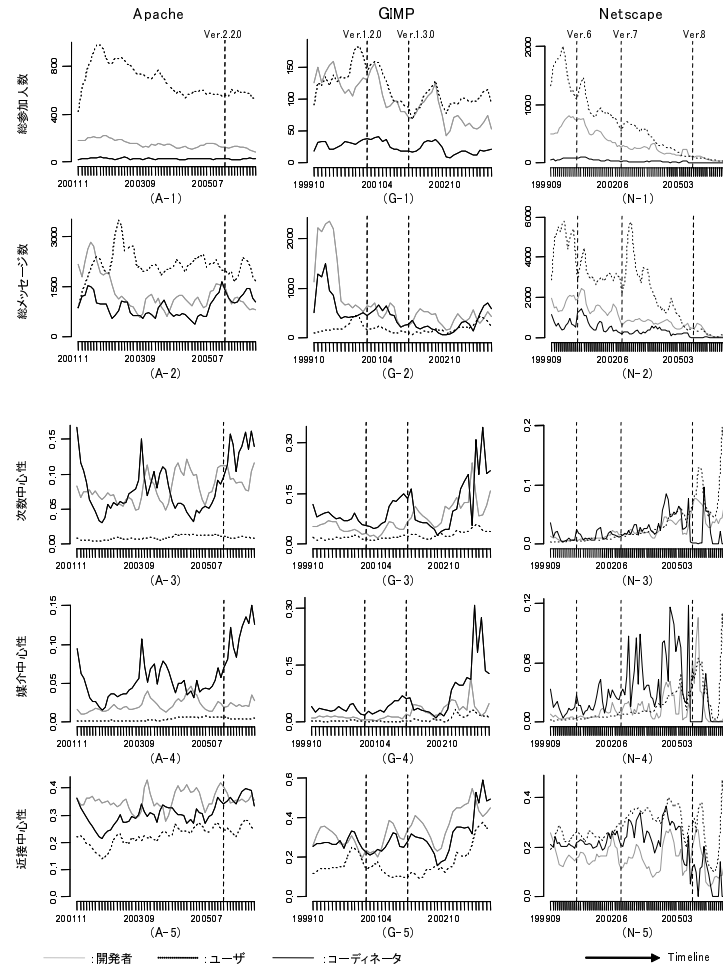


Fig. 4. Changes of average of network centralities of Apache community.