

A Comparison Study on the Coordination Patterns in Oss Community

Shinsuke Matsumoto, Yasutaka Kamei, Masao Ohira and Ken-ichi Matsumoto
Graduate School of Information Science, Nara Institute of Science and Technology
8916-5 Takayama, Ikoma, Nara, Japan
{shinsuke-m, yasuta-k, masao, matumoto}@is.naist.jp

ABSTRACT

abst abst abst abst abst abst abst abst abst abst abst abst
abst abst abst abst abst abst abst abst abst abst abst abst
abst abst abst abst abst abst abst abst abst abst abst abst
abst abst abst abst abst abst abst abst abst abst abst abst
abst abst abst abst abst abst abst abst abst abst abst abst
abst abst abst abst abst abst abst abst abst abst abst abst
abst abst abst abst abst abst abst abst abst abst abst abst
abst abst abst abst abst abst abst abst abst abst abst abst
abst abst abst abst abst abst abst abst abst abst abst abst
abst abst abst abst abst abst abst abst abst abst abst abst
abst abst abst abst abst abst abst abst abst abst abst abst
abst abst abst abst abst abst abst abst abst abst abst abst
abst abst abst abst abst abst abst abst abst abst abst abst
abst abst abst abst abst abst abst abst abst abst abst abst
abst abst abst abst abst abst abst abst abst abst abst abst
abst abst abst abst

Categories and Subject Descriptors

H.4 [Information Systems Applications]: Miscellaneous; D.2.8 [Software Engineering]: Metrics—complexity measures, performance measures

General Terms

Delphi theory

Keywords

ACM proceedings, L^AT_EX, text tagging

1. INTRODUCTION

厳格な指揮系統を持たない開発環境を取りながら、高品質かつ高機能なソフトウェアの開発を続ける OSS コミュニティを対象とした研究が現在盛んに行われている。Howison らは OSS コミュニティ内の開発者同士が形成するコミュニケーション構造を informal social structure と名づけ、分析を行った。According to them, 多くの OSS コミュニティでは開発者コミュニティの中で中心的な役割を担う開発者が長い期間、存在していることが明らかとなった。Bird et al. examined the correlation between centralities and Apache developer contributions from developer mailing lists and the changed history of source code[?].

The analysis results indicated that developers communicating with many other developers contributed more to source code changes.

また、OSS 開発者のコミュニケーションについての研究される一方で、OSS の重要な要素の役割の一つとしてユーザの存在についても指摘されている。Raymond らは開発されたソフトウェアを利用するユーザを OSS コミュニティにおける共同開発者と位置づけ、その重要性を指摘した。多くのユーザがソフトウェアを利用することで、コードに含まれるバグが指摘され、また高速改良にも繋がるためである。また、Ye らは OSS コミュニティのユーザがコミュニティ参加者同士での学習を通じてコミュニティへの帰属意識を持ち、次第に重要な役割へと変化するという参加形態（正統的周辺参加）が開発者のモチベーションに繋がるとしている [?].

このように開発者とユーザの協調作業の重要性が指摘されている一方で、開発者とユーザが実際にどのような informal social structure を構築しているかについては明らかにされていない。そこで、本論文では OSS コミュニティにおける開発者とユーザの構築する informal social structure について、2 つのコミュニティ（開発者、ユーザ）に所属し、仲介役を担っていると考えられる人物に着目し分析を行う。

This paper is organized as follows, ...

2. INFORMAL SOCIAL STRUCTURE IN OSS COMMUNITY

一般的に、OSS コミュニティはオンライン上の分散開発環境を取るため、参加者同士のコミュニケーションには Mailing List (ML) やニュースグループ、掲示板などのオンラインメディアが用いられる。本稿では、これらのオンラインメディアへの投稿されたメッセージの送信者をノード、メッセージに対する返信をエッジとしたネットワーク構造を OSS コミュニティにおける informal social structure と見なす。例えば、コミュニティ参加者 A の投稿メッセージに対して、B 及び C が返信をした場合、ノード A に対して B と C からエッジが張られる。informal social structure の例を図??に示す。図の左側はオンラインメディアのスレッド構造の例であり、図の右側がこのスレッド構造から構築された informal social structure を表す。

～つなぎ～

OSS 参加者には開発者やテスト、ユーザなど様々な役割が存在しており、各役割に応じたサブコミュニティが形成されている。例えば、新機能追加のための技術的な議論や、検出された不具合の関する議論は開発者コミュニティ内で行われ、ソフトウェアの利用方法に対する質問とその回答はユーザコミュニティ内

で行われる。このため、ユーザ間で議論された、新機能の要望などの開発者にとって有益な情報がユーザコミュニティの中で埋もれてしまう可能性がある。我々は、このような単一コミュニティ内で議論された情報を他コミュニティへ媒介するような参加者を～～～と考え着目する。

図??に2つのサブコミュニティを持つOSSコミュニティの例を示す。

以降、開発者コミュニティで議論を行った参加者を P_{dev} 、ユーザコミュニティで議論を行った参加者を P_{user} 、両コミュニティで議論を行った参加者を $P_{d\cap u}$ と呼ぶ。

3. ANALYSIS METHODOLOGY?

3.1 Perspective of Analysis

OSSコミュニティの参加者が形成する informal social structure を分析するために、以下の2つの観点から分析を行う。

- snapshot:
ある一期間についての informal social structure 全体を可視化し、その構造の特長について従来提案されている指標を用いて分析する。
ー ネットワークの特徴を捉えるために用いる指標としては、一般にネットワーク分析で用いられている平均パス長とクラスタリング係数を用いる。各指標の詳細は次節で述べる。
- transition:
各期間における informal social structure の推移について分析を行う。ただし、分析対象のネットワークの規模が大きく参加者全体の中心性の推移を扱うことは不可能なので、ここでは観点1で特徴的な $P_{d\cap u}$ を選択し、分析の対象とする。

3.2 Network Centrality

本節では分析に用いた3つの指標について説明する。

ネットワーク全体の特徴を表す指標である平均パス長とクラスタリング係数を用いる。

Characteristic Path Length 平均パス長 L とはネットワークに存在する全てのノード間の最短パス長を平均した値のことであり以下のように表される。

$$L = \quad (1)$$

頂点 v_i についてのクラスタリング係数 C_i とはネットワーク内のノードがどの程度密集しているかの度合いを表す。グラフ全体のクラスタリング係数 C は、グラフ全体の n 個の頂点の C_i を平均した値であり、以下の式で定義される。

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

2つの値を～

networkの中心的な人物を表す指標として3つの中心性が提案されている。近接中心性は～近さを表す指標であり、あるノードからあるノードまで到達するためのステップを基に算出される。しかし、MLやニュースグループなどの誰もが情報を取得することのできるオープンなメディアでは、情報伝達のステッ

プがそもそも存在しないため、～、表す指標としては適切ではない。また、次数中心性は～で適切ではない。そこで～を分析した～らの研究[]と同様に、 $P_{d\cap u}$ を評価するための指標として～を表す媒介中心性を用いる。

ノード v_i の媒介中心性 $C_{betweenness}(v_i)$ は、他の2つノード v_j から v_k への最短経路にノード v_i が含まれる割合である。媒介中心性は、式(3)のように表される。

$$C_{betweenness}(v_i) = \frac{\sum_{j < k}^n \sum_{j < k}^n p_{jk}(v_i)}{\sum_{j < k}^n \sum_{j < k}^n p_{jk}} \quad (3)$$

ここで、 n はネットワーク内のノード数を、 p_{jk} は v_j から v_k への最短経路、 $\sum_{j < k}^n \sum_{j < k}^n p_{jk}$ は v_j から v_k への最短経路の総数を表し、 $\sum_{j < k}^n \sum_{j < k}^n p_{jk}(v_i)$ は v_i を含む v_j から v_k への最短経路の総数を表す。

$C_{betweenness}(v_i)$ は値域 $[0,1]$ を取り、値が大きいほど媒介中心性が高いことを表す。

本稿では、媒介中心性が高い $P_{d\cap u}$ を、2つのコミュニティ間での情報の伝達や調整行動を行っている人物と見なし、分析する。

4. CASE STUDY

本章では～ケーススタディについて説明する。

4.1 分析対象 OSS コミュニティ

本論文では一般的に良く知られている Apache HTTP Server と Netscape Browser の2つのOSS開発コミュニティを対象として分析を行う。各OSSコミュニティの統計量を表3に示す。また、各OSSコミュニティの開発ソフトウェアについては以下のとおりである。

- Apache HTTP Server¹: 2008年2月現在、世界一のシェアを誇るWebサーバ(HTTP Server)ソフトウェアである。3つのメジャーバージョンの並行開発をはじめ、欠陥の修正パッチを継続的に多数リリースしており、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。
- Netscape Browser²: Webブラウザソフトウェアであり、1996年ごろの最盛期にはブラウザ市場の8割を占めていた。1998年にはInternet Explorerのシェア拡大に対抗するためRaymondらの研究[]を受けてオープンソース化を実施したが、シェアを取り返すことができずコミュニティの規模は縮小化していった。2008年2月に、ソフトウェアの開発及び、ユーザサポートは停止された。

4.2 分析対象データ

分析対象としたコミュニケーション履歴データはApacheはMailing List、Netscapeはニュースグループであり、それぞれ各OSSコミュニティが主として議論の場に使っていたメディアである。これら分析対象メディアの差については、informal social structureの分析に対しては大きな違いは表れないと我々は以下の様に考える。MLとニュースグループの違いは情報の受信方法にある。MLの場合、利用者自身のメールアドレスを登録しておけば、メールサーバより自動的に投稿メッセージが送信されてくるが、ニュースグループの場合、利用者が能動的

¹<http://httpd.apache.org/>

²<http://browser.netscape.com/>

Table 1: 観点 1：各 OSS コミュニティの統計量

	総参加人数		
	対象期間	開発者	ユーザ
Apache	2005/12	-	-
Netscape	2002/08	-	-

に投稿メッセージを選択し受信する必要がある。しかし、我々の分析対象とする informal social structure はメディアに対するメッセージの送信という行為から形成されるため、受信方法の違いという分析メディアの違いはないと考えられる。

また、分析対象とした期間ではニュースグループというメディアそのものの衰退時期と重なってはいるが、Netscape コミュニティが別のオンラインメディアに移行したというような

観点 1 における分析対象時期と総参加人数を表 1 に、観点 2 における分析対象期間、総参加人数及び、総メッセージ数を表 2 に示す。観点 1 での分析対象とした時期は、それぞれメジャーバージョン (Apache: Ver.2.2.0, Netscape: Ver.7.0) がリリースされた時期である。観点 2 の分析対象期間については、Web 上に公開されている履歴データの取得できた期間すべてを用いた。

観点 2 での time sliding method 前後 1ヶ月を扱う

4.3 Result

4.3.1 Result of the Perspective1: snapshot

各 OSS コミュニティのメジャーバージョンがリリースされた時期についての informal social structure を分析した観点 1 の結果について説明する。描画及び指標値計算には pajek³を用い、図の配置アルゴリズムとしては kamada-kawai⁴を用いた。ただし、分析対象のネットワークはノード数とリンク数が非常に大きいため、すべてを図示した場合、視覚的に構造を把握することは厳しくなる。ここでは、視認性を上げるためにエッジを持たないノード (新規に発言を行ったが返信が得られなかった参加者) を図から除いてある点に注意されたい。

図 3 は P_{dev} , P_{user} 及び $P_{d \cap u}$ を含む全ての参加者から形成される informal social structure を可視化したものである。図の左側、白のノードは P_{dev} であり、中央の赤のノードは $P_{d \cap u}$ 、右側、青のノードは P_{user} である。また、表 3 に図 3 より算出されたネットワーク全体の特徴を表すメトリクスを示す。

ここで着目すべきは $P_{d \cap u}$ のエッジ数の違いである。表 3 より、全体の平均エッジ数 k に大きな違いはないが、Apache の $P_{d \cap u}$ は Netscape の $P_{d \cap u}$ と比べて多くのエッジを持っていることが確認できる。これは Apache の $P_{d \cap u}$ は、2つのコミュニティで頻繁に発言することにより、開発者とユーザ間の協調作業を円滑に進めていると考えられる。

³<http://vlado.fmf.uni-lj.si/pub/networks/pajek/>

		nodes	edges	<k>	L	C
all participant	Apache	661	1230	1.86	3.34	0.10
	Netscape	1022	1600	1.57	4.58	0.08
developer $P(d) \cap P(d \cap u)$	Apache	131	431	3.29	2.76	0.21
	Netscape	319	375	1.18	6.64	0.04
user $P(u) \cap P(d \cap u)$	Apache	554	861	1.55	3.31	0.08
	Netscape	741	1246	1.68	3.11	0.10

Table 3:

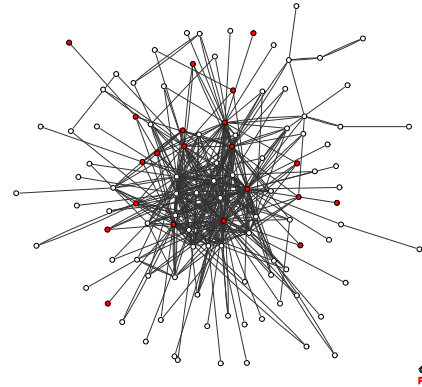


Figure 1: result apache dev

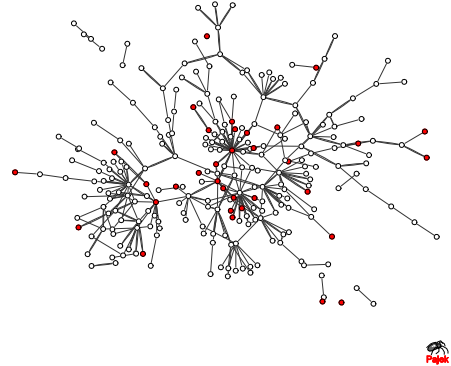


Figure 2: result netscape dev

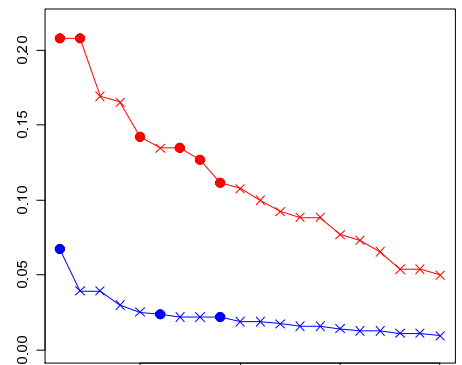


Figure 3: result netscape dev

Table 2: 観点 2：各 OSS コミュニティの統計量

	分析対象期間			総参加人数		総メッセージ数	
	開始年月	終了年月	期間 (月)	開発者	ユーザ	開発者	ユーザ
Apache	2001/11	2006/09	59	1,570	9,347	32,352	67,114
Netscape	1999/09	2007/02	92	8,271	15,115	29,417	61,946

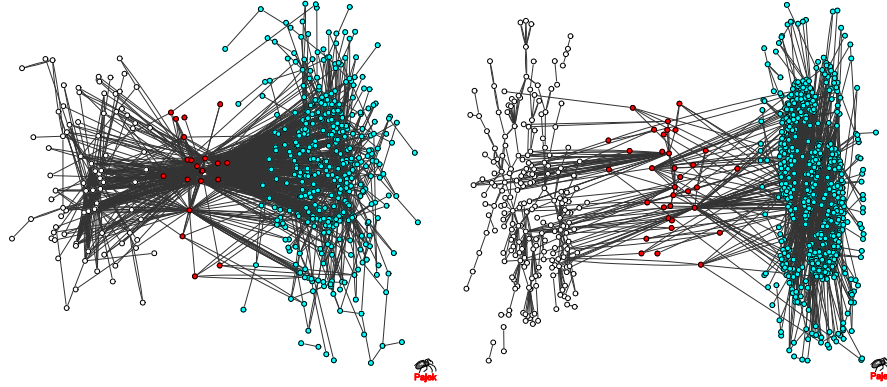


Figure 4: result netscape dev

次に、開発者コミュニティの参加者 (P_{dev} と $P_{d \cap u}$) network とユーザコミュニティの参加者 (P_{user} と $P_{d \cap u}$) network をそれぞれ切り出し、個別に分析を行う。

開発者コミュニティのネットワーク構造を切り出した図を Apache については図 3 に、Netscape については図 3 に示す。各図のメトリクスを表 3 に示す。

Apache は図の中央に位置するエッジの多いノード同士の密度が非常に高く、その周辺にエッジの少ないノードが付随しているというスター型に近いネットワーク構造を形成している。一方で、と少ないノードに大きな偏りがあり、図の中央に位置する発言の多い参加者同士のエッジが特に多いことが分かる。Ns は広く分布している。このため～

参加者の数に対して平均エッジ数 k が大きく、少数の参加者が活発に議論を進めていると考えられる。

Netscape は Apache と比べて平均パス長 L が大きい傾向にある。

参加者の数に対して次数が大きいという点はこの図からも確認できるが、着目すべきは $P_{d \cap u}$ である。

これは Netscape の開発者コミュニティでは、 P_{dev} 同士の距離が長くコミュニティ全体としてまとまっていない情報とか知識の伝達に多くの開発者を辿る必要があり、うまくいっていないと考えられる。また、図より Ns は中心的な人物がいないことも確認できる

縦軸は各ノードの次数であり、横軸は次数の多いほうから並べてある Ap はトップ 2 人が $P_{d \cap u}$ で中心的な位置にいる。30 人中 8 人が $P_{d \cap u}$ である一方 Ns は値低くて、30 人中 8 人しかいないこれは開発者の中で多くの発言をする人物がいないことを指す

次にユーザコミュニティに着目する。リーダーが一人

4.3.2 Result of the Perspective2

動画・メッセージ数の推移・参加者数の推移・オーダー・value

5. DISCUSSION

6. CONCLUSION AND FUTURE WORK

7. ACKNOWLEDGEMENTS