

A Comparison Study on the Coordination Patterns in Oss Community

Shinsuke Matsumoto
Graduate School of
Information Science, Nara
Institute of Science and
Technology
8916-5 Takayama, Ikoma
Nara, Japan
shinsuke-m@is.naist.jp

Yasutaka Kamei
Graduate School of
Information Science, Nara
Institute of Science and
Technology
8916-5 Takayama, Ikoma
Nara, Japan
yasuta-k@is.naist.jp

Masao Ohira
Graduate School of
Information Science, Nara
Institute of Science and
Technology
8916-5 Takayama, Ikoma
Nara, Japan
masao@is.naist.jp

Ken-ichi Matsumoto
Graduate School of
Information Science, Nara
Institute of Science and
Technology
8916-5 Takayama, Ikoma
Nara, Japan
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

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を対象とした研究が行われる一方で、OSS 開発の重要な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 [1]. 多くのユーザがソフトウェアを利用することにより、コードに含まれるバグが指摘され、結果として品質の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 [2].

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。分析では一般に良く知られた 2 つの OSS コミュニティを対象として、ある一時期における informal social structure 全体の構造的特徴についてと、時

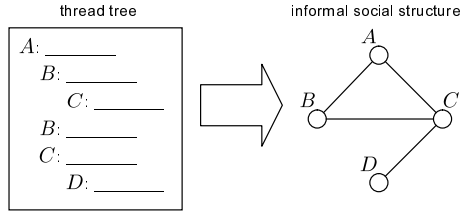


Figure 1: Constructing an informal social structure

系列に従った中心的な人物の**という2つの観点から分析を行う。

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。本論文では、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。informal social structure の例を図 1 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。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。

～なんかつなぎ～

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。

図 2 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。 P_{dev} は開発者コミュニティで議論を行った参加者を、 P_{usr} はユーザコミュニティで、 $P_{d \cap u}$ は 2 つのコミュニティで議論した経験のある人物を指す。このような両コミュニティに参加する $P_{d \cap u}$ は、開発者としての深い知識と経験を持ちながら、ユーザに対する FAQ や*を行うと考えられる。また逆に、ユーザからの新機能の要望や発見されたバグの報告などを開発者へ吸い上げる役割も担っているといえる。我々は、この $P_{d \cap u}$ 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が、参加者同士の円滑なコミュニケーションネットワークの形成に大きく関与していると考え、着目する。

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:

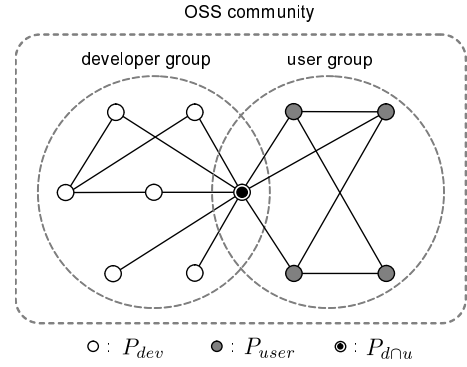


Figure 2: An example of oss community

ある一期間についての informal social structure を可視化しその構造の特長について、分析者による定性的な分析及び、従来提案されている指標を用いた定量的な分析を行う。ネットワーク全体の特徴を捉えるための指標としては、複雑ネットワークの分野で一般に用いられ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:

ネットワーク構造全体の傾向ではなく、informal social structure 中の中心的な人物に着目し、その次数の推移について分析を行う。ここでは次数中心性の最も高い人物を、そのネットワーク内の中心的な人物と見なす。ただし、分析対象のネットワークの規模が大きく、各期間における中心人物全てを描画することは不可能なので、ここでは観点 1 で次数の最も多い $P_{d \cap u} 5$ 人を選択し、その推移を分析する。次数中心性については次節で述べる。

「この観点で妥当なのか、根拠は」

3.2 Network Analysis Metrics

本節では分析に用いた 3 つの指標について説明する。

3.2.1 Average Path Length

ある 2 ノード間について、最短で到達可能なルートを通った場合の経由したノードの数のことを最短パスと呼ぶ。ノード v_i についての平均パス長 L_i とは、 v_i と接続している全ノードとの最短パス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パス長の平均を取った値を、そのネットワークに対する平均パス長 L と呼び、以下の式で定義される。

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

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

3.2.2 Clustering Coefficient

ノード 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値である。具体的にはあるノードと接続したノード同士が繋がっている割合のことであり、エッジの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．

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

この C が高いほど，高い凝集性を持ったネットワークであることを表す．

3.2.3 Degree Centrality

ノード v_i の次数中心性 $C_{deg}(v_i)$ は， v_i の持つ次数を，ネットワーク内でノード v_i が取りうる最大の次数によって正規化した値である． v_i の持つ次数を $deg(v_i)$ としたとき，次数中心性 $C_{deg}(v_i)$ は，以下の式で表される．

$$C_{deg}(v_i) = \frac{deg(v_i)}{n-1} \quad (3)$$

この値が高いほど，多くのノードと隣接しているノードである．つまり，多くの参加者と盛んに会話をを行っている中心的なノードであるといえる．

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 コミュニティの informal social structure の違いについて比較を行う．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 [1] 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用い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記載は見られなかった．

観点 1 における分析対象時期と総参加人数を表??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??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．図の描画及び指標値の計算には pajek³を用い，図の配置アルゴリズムとしては kamada-kawai [2] 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 5 は各 OSS コミュニティの informal social structure を可視化したネットワーク図であり，上側は Apache，下側は Netscape である．図の左側は全ての参加者 (P_{dev} , P_{usr} , P_{dru}) から形成される informal social structure であるが，視認性の向上のため，開発者 (P_{dev} , P_{dru}) とユーザ (P_{usr} , P_{dru}) を切り出した図をそれぞれ中央と右側に配置した．各図の白のノードは P_{dev} であり，赤のノードは P_{dru} ，青のノードは P_{usr} である．表??に各 informal social structure 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．

まず，図 5 の左側，全参加者からなるネットワーク構造の違いに着目するここで着目すべきは P_{dru} のエッジ数の違いである．表??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_{dru} は Netscape の P_{dru} と比べて多くのエッジを持っていることが図より確認できる．これは Apache の P_{dru} 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/>

Table 2:

		nodes	edges	<k>	L	C
all participants	Apache	661	1230	1.86	3.34	0.10
	Netscape	1022	1600	1.57	4.58	0.08
developers	Apache	131	431	3.29	2.76	0.21
	Netscape	319	375	1.18	6.64	0.04
users	Apache	554	861	1.55	3.31	0.08
	Netscape	741	1246	1.68	3.11	0.10

¹<http://httpd.apache.org/>

²<http://browser.netscape.com/>

Table 1: 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

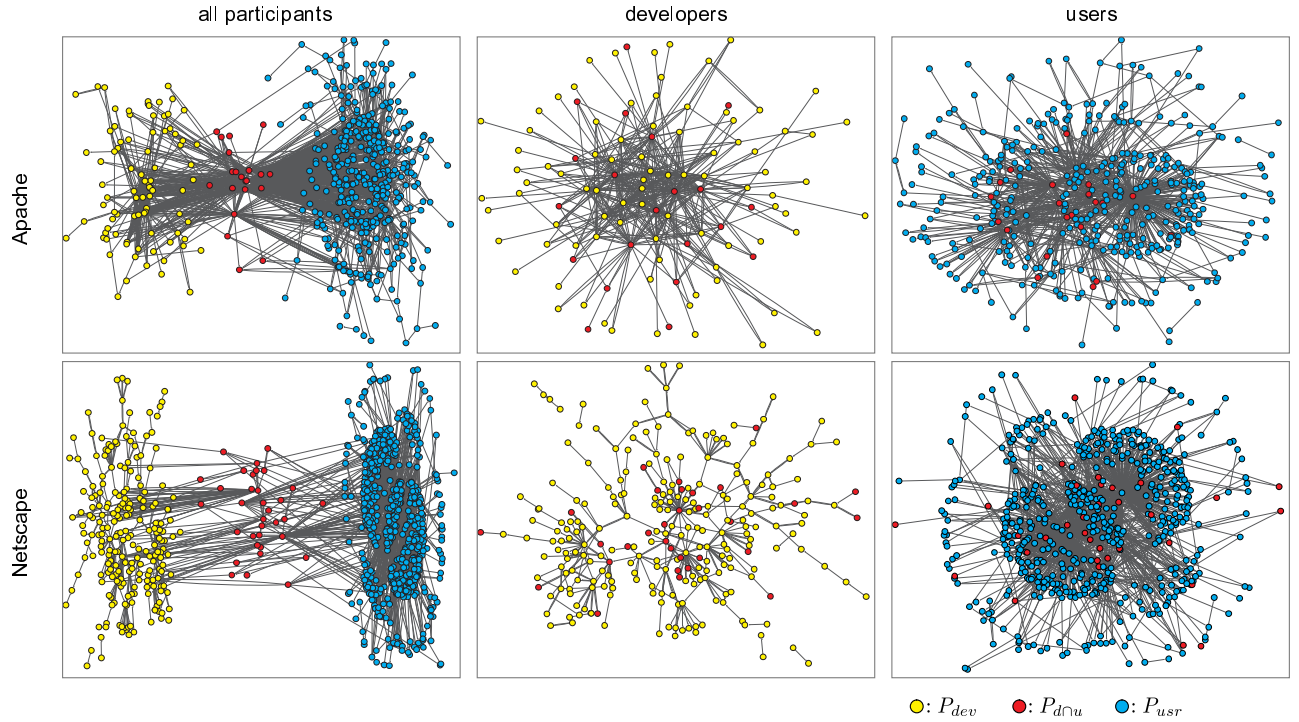


Figure 3: result all network

次に、開発者コミュニティの参加者 (P_{dev} と $P_{dev \cap u}$) network に着目する。Apache の開発者コミュニティは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傾向にあり、巨視的にはスター型に近いネットワーク構造であるといえる。また参加者の数に対して平均エッジ数 $\langle k \rangle$ 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。一方で、Netscape は平均パス長 L が 6 ステップと長く、クラスタリング係数 C が小さいことから、情報や知識の伝達に多くの開発者を迎える必要があると考えられる。また、各ノードのエッジ数に大きな隔たりはなく、コミュニティをまとめる中心的な人物が存在しないといえる。

中心的な人物の分布を確認するために、次数オーダーで並べた各参加者の次数のグラフを図 4 に示す。縦軸は次数であり、横軸は各ノードの次数のランキングである。

Apache は次数の多い上位 10 人のうち 6 人と半分以上が $P_{dev \cap u}$ であり、多くの $P_{dev \cap u}$ が中心となって発言を行っていることがわかる。一方で Netscape は次数が最も多い参加者が $P_{dev \cap u}$ にはあるが、上位 10 人中の $P_{dev \cap u}$ は 3 人のみと少なく、また全体的に値も低い傾向にある。 $P_{dev \cap u}$ は開発者とユーザの両コミュニティに参加する人物であり、協調作業に有益な情報を抱えている参加者である可能性が高い。そのため多くの発言を行うノードは P_{dev} であるよりも $P_{dev \cap u}$ であるほうがコミュニティの協調作業の円滑化に繋がると我々は考える。よって開発者コミュニティの構造としては Apache のほうが優れているといえる。

次にユーザコミュニティに着目する。両 OSS コミュニティの L , C ともに大きな差はないが、Apache はある $P_{dev \cap u}$ を中心とした 1 つのクラスタが存在する一方で、Netscape には 2 人の P_{usr} を中心とするクラスタが 2 つ存在していることが図から確認できる。この次数の多い Apache の $P_{dev \cap u}$ と次数の多い Netscape の 2 人の P_{usr} は図 5 から確認できる。これらの人物はユーザからの質問に対して積極的に答えるような人物であると考えられるが、議論の中で得られたユーザのリクエストなどを知識として得、開発者コミュニティへ伝えるという点では $P_{dev \cap u}$ が中心にいる Apache コミュニティの構造のほうが優れていると考えられる。また、次数の多いトップ 10 人のうち $P_{dev \cap u}$ の人数は Apache は 4 人、Netscape は 1 人と、開発者コミュニティと同様に Apache のほうが $P_{dev \cap u}$ が多かった。

4.3.2 Result of the Perspective2

次数の多い $P_{dev \cap u}$ 5 人を選択し、動画・メッセージ数の推移 k ・参加者数の推移・オーダー・value

5. DISCUSSION

6. CONCLUSION AND FUTURE WORK

7. ACKNOWLEDGEMENTS

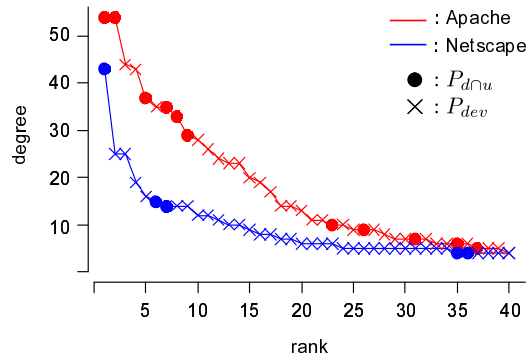


Figure 4: result degree dev

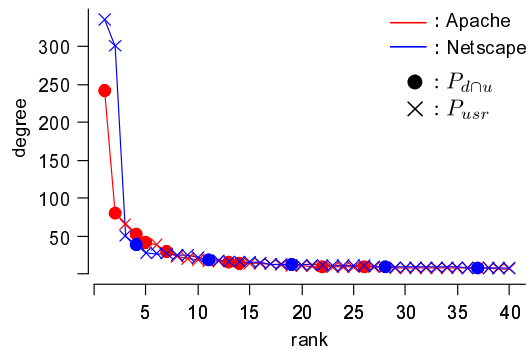


Figure 5: result degree dev