

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

Categories and Subject Descriptors

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

General Terms

Delphi theory

Keywords

ACM proceedings, L^AT_EX, text tagging

1. INTRODUCTION

高品質かつ高機能なソフトウェアの開発を続ける OSS コミュニティが注目を浴びようになっている。OSS は世界各地の開発者がネットワーク上でコミュニケーションをとりながら共同開発を行うという、分散型開発プロジェクトの一例である。OSS 開発のコミュニケーション構造の実態を分析することで、現在盛んに行われつつあるオフショア開発などの分散開発環境に対する有益な知見が得られると期待できる。

開発者同士のコミュニケーションの実態を分析した研究が多数報告されている。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 [?].

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。

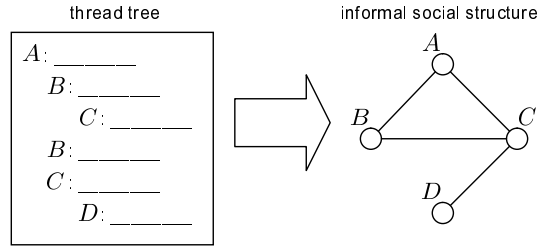


Figure 1: Constructing an informal social structure

open question そこで、本稿では 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、ある一期間における informal social structure の詳細についてと、時間の経過に伴う構造の変化という2つの観点から分析を行う。

This paper is organized as follows, ...

2. RELATED WORKS

- dev mail Ap vs Moz: Moccas Hervsleb ICSE 2007: 開発者昇格プロセス jensen SCACCI

- network howison bird

ユーザがどのように

3. 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形態を取るため、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。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を基に構築された 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。

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。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。役割に応じたサブコミュニティ内で議論を進めることで、ユーザはユーザにとって、開発者は開発者にとって必要な情報を得ることが容易となる。しかし、ユーザ同士の議論の中でソフトウェアの品質改良に繋がる有益な意見が表れていたとしても、その意見が開発者の目に入らず、結果としてログの中に埋もれてしまう可能性がある。そのため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

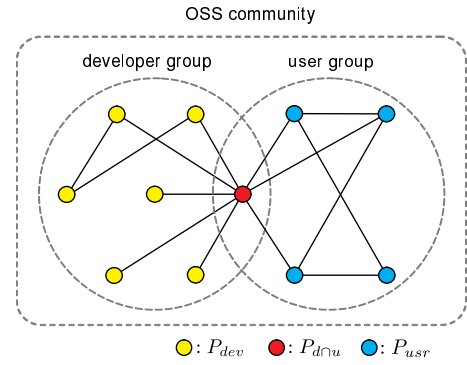


Figure 2: An example of oss community

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。

4. ANALYSIS METHODOLOGY

4.1 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。

4.1.1 static

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従来提案されている指標を用いた定量的な分析を行う。分析対象の informal social structure は、開発者コミュニティの参加者 ($P_{d\cap u}$, P_{dev}) が形成するネットワーク、ユーザコミュニティの参加者 ($P_{d\cap u}$, P_{usr}) が形成するネットワーク、そして、全ての参加者 ($P_{d\cap u}$, P_{dev} , P_{usr}) が形成するネットワークという3つの捉え方が可能である。これら3つのネットワークそれぞれについて、 $P_{d\cap u}$ を中心としてその構造的特徴を分析することで、2つのコミュニティをまたがる $P_{d\cap u}$ の役割をより詳細に分析することが可能となる。ネットワーク構造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 L とクラスタリング係数 C を用いる。各指標の詳細は次節で述べる。

4.1.2 dynamic

OSS コミュニティは誰でも自由に参加することが可能であり、参加者同士が自由に議論をしながら開発を進める。このため、OSS コミュニティの参加者が形成する informal social structure は、時系列と共に変化し成長/衰退する [it 生きた] ネットワークであるといえる。この生きたネットワークを分析するためには観点1での静的な分析のみでは不十分であり、時系列に沿った分析が必要である。そこで分析の観点2として、informal social structure の時間的な構造の変化についても分析を行う。さらに、ネットワーク全体の推移に限らず、個々の $P_{d\cap u}$ のコーディネーション能力の変化についても分析する。各期間における $P_{d\cap u}$ 全てを分析することは分析対象のネットワークの規模の大きさから困難であるため、ここでは観点1でコーディネーション能力の高い $P_{d\cap u}$ 5人のみを選択し、その推移を分析する。 $P_{d\cap u}$ コーディネーション能力を表す指標としては、「次数中心性はコーディネーション能力と相関がある」という*らの研究結果から次数中心性を用いることが考えられる。しかし、次数中心性は次数をネットワークの規模で正規化した値であり、同じネットワーク内（つまり同じ規模のネットワーク）のノー

ド同士を比較する際には本質的には次数と変わらない。本稿では、より直感的な次数をコーディネーション能力を表す指標として扱う。

4.2 Network Analysis Metrics

本節では分析に用いた 2 つの指標について説明する。

4.2.1 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辿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。本稿で対象とする informal socila structure は、ML などのオープンなメディア上で構成されるため情報の伝達網とは異なる。 L が

4.2.2 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。

ただし、図 2 に示すような、 $P_{d \cap u}$, P_{dev} , P_{usr} 全てを含むネットワークに対してのクラスタリング係数 C は意味を成さない。これは、本来別のネットワークに所属する P_{dev} と P_{usr} の間には必ずエッジが存在しないため、 $P_{d \cap u}$, P_{dev} , P_{usr} のエッジの三角系は表れないためである。分析の際には、開発者コミュニティの C 及びユーザコミュニティの C は算出するが、全参加者のネットワークの C は扱わない。

5. CASE STUDY

本章では実際の OSS コミュニティのコミュニケーション履歴データを用いたケーススタディについて説明する。

5.1 分析対象 OSS コミュニティ

本稿では、一般的に OSS として良く知られている 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。

Table 1: quantity perspective

		apache	netscape
perspective 1	time	2005/12	2002/08
	# participants	683	1,058
	# mails	4,068	3,509
perspective 2	start	2001/11	1999/09
	end	2007/09	2007/02
	# participants	12,710	23,386
	# mails	112,143	91,363

- Apache HTTP Server¹: 2008 年 2 月現在、世界一のシェアを誇る 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。

5.2 分析対象データ

分析対象のコミュニケーション履歴データは、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より収集した。各メディアは Apache は Mailing List, Netscape はニュースグループである。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、利用者が自身のメールアドレスをあらかじめ登録しておけば、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、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 と観点 2 の分析対象時期、及び各時期における履歴データの統計量を表 1 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を取得できた全ての期間を用いた。月をまたがった議論も多数存在するため、ある月 m と m の前後 1 ヶ月 ($m \pm 1$) の履歴データを、 m における informal social stcuture とした。

5.3 Result

5.3.1 Result of the Perspective1: static

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ネットワーク分析プログラムである pajek³を用い、図の配置アルゴリズムとしては kamada-kawai []

¹<http://httpd.apache.org/>

²<http://browser.netscape.com/>

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

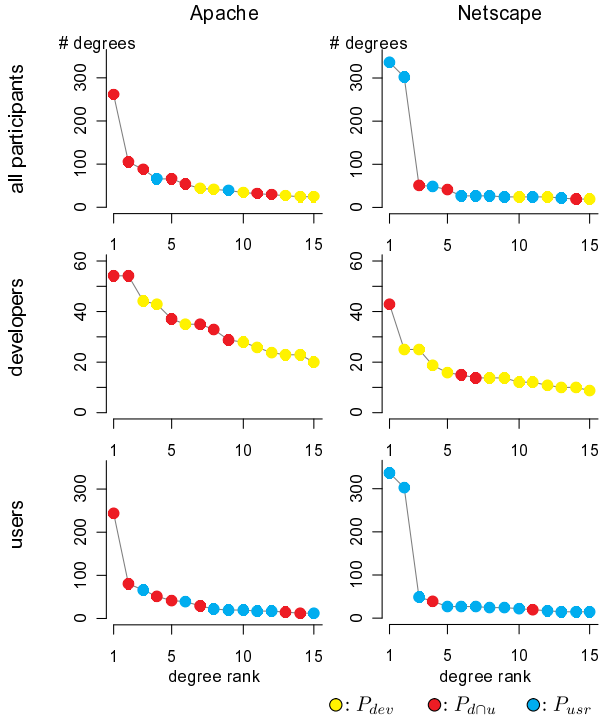


Figure 4: result all degree

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、左側はApache、右側はNetscapeである。図の上側は全ての参加者（ P_{dev} , P_{usr} , $P_{d \cap u}$ ）から形成されるinformal social structureであり、中央は開発者（ P_{dev} , $P_{d \cap u}$ ）、下側はユーザ（ P_{usr} , $P_{d \cap u}$ ）である。ノードの大きさは次数の多さを表し、ノードの色は参加者がいずれの役割に属するかを表しており、黄は P_{dev} 、赤は $P_{d \cap u}$ 、青は P_{usr} である。表2に各informal social structureより算出さ

Table 2: result all metrics

		Apache	Netscape
all	# nodes	661 (24)	1,022 (38)
	# edges	1230 (720)	1,600 (221)
	$\langle k \rangle$	1.86	1.57
	L	3.34	4.58
	C	0.10	0.08
dev	# nodes	131 (24)	319 (38)
	# edges	431 (302)	375 (121)
	$\langle k \rangle$	3.29	1.18
	L	2.76	6.64
	C	0.21	0.04
user	# nodes	554 (24)	741 (38)
	# edges	861 (540)	1246 (720)
	$\langle k \rangle$	1.55	1.68
	L	3.31	3.11
	C	0.08	0.10

れたネットワーク全体の特徴を表すメトリクスを示す。表中の括弧内の値は $P_{d \cap u}$ のみについての値を表す。各ネットワークの中心的な（次数の多い）人物は図3からも確認できるが、具体的な次数の値を確認するために、次数の降順で並び替えたときの各ノードの次数を図4に示す。グラフの縦軸は次数、横軸は各ノードの次数のランキングである。図3の配置と同様に、左側はApache、右側はNetscape、上から順に全参加者、開発者、ユーザである。ノードの色も図3と対応している。

まず、図3の上側、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 $P_{d \cap u}$ のエッジ数の違いである。表2より、全体の平均エッジ数 $\langle k \rangle$ は1.86と1.57であり大きな違いはない。しかし、図3より、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。

次に、開発者コミュニティの参加者（ P_{dev} と $P_{d \cap u}$ ）networkに着目する。表2より、Apacheの開発者コミュニティは人数がNetscapeの半分以下ではあるが、エッジ数が多く、平均エッジ数 $\langle k \rangle$ はNetscapeの倍近い値をとっている。この平均エッジ数 $\langle k \rangle$ の高さに起因して、Apache開発者コミュニティは平均パス長 L が全ネットワーク中最小、クラスタリング係数 C が全ネットワーク中最大であることも特徴である。逆に、Netscapeの開発者コミュニティは平均パス長 L が全ネットワーク中最大、クラスタリング係数 C が全ネットワーク中最小であった。また図3より、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、その末端にエッジの少ないノードが接続しており、巨視的にはスター型に近いネットワーク構造であることが分かる。一方でNetscapeの開発者ネットワークは、あるノードを中心とする10人ほどの小さなスターが細かいリンクで連結している構造である。図4より、次数の多い参加者トップ15人のうち $P_{d \cap u}$ の人数は、Apacheは6人、Netscapeは3人であった。

次にユーザコミュニティの参加者（ $P_{d \cap u}$ と P_{usr} ）networkに着目する。表2より、2つのOSSコミュニティを比較すると、ノード数とエッジ数についてはNetscapeのほうが多いものの、 $\langle k \rangle$, L , C ともに大きな差はなく、ネットワーク構造に際立った差は見られない。しかし図3から、Apacheは次数の多い $P_{d \cap u}$ を中心とした1つのクラスタが存在しており、その付近に多くの次数を持った $P_{d \cap u}$ が位置するという $P_{d \cap u}$ を中心とした構造となっている。一方でNetscapeは、 P_{usr} を中心とする巨大なクラスタが2つ存在しており、 P_{usr} 中心の構造となっていることが確認できる。このことは図4からも確認できる。また図4から、次数トップ15人のうち $P_{d \cap u}$ は2人だけであり、発言回数の多い $P_{d \cap u}$ が少ないこともNetscapeユーザnetworkの特徴である。結論として、ApacheとNetscapeのユーザnetworkは、構造的な違いよりも中心にいる参加者の役割に違いがあるといえる。

5.3.2 Result of the Perspective2

時間の経過に伴うinformal social structure構造の変化に着目した観点2の分析結果について説明する。全参加者ネットワークのメトリクスの推移を図5に示す。図の左側はApache、右はNetscapeであり、上側はノード数、エッジ数、下側は平均次数 $\langle k \rangle$ 、平均パス長 L である。縦の破線は観点1で分析した期間を表す。

図5より、Apacheネットワークのノード数は2002年半ばにピークを向かえ、その後は緩やかに減少し、およそ600人から700人の間で留まっている。平均パス長 L についても全ての期間で3から5の値と大きな変化はなく、Apacheネットワークは規模や構造に大きな変化のない安定したネットワークである。

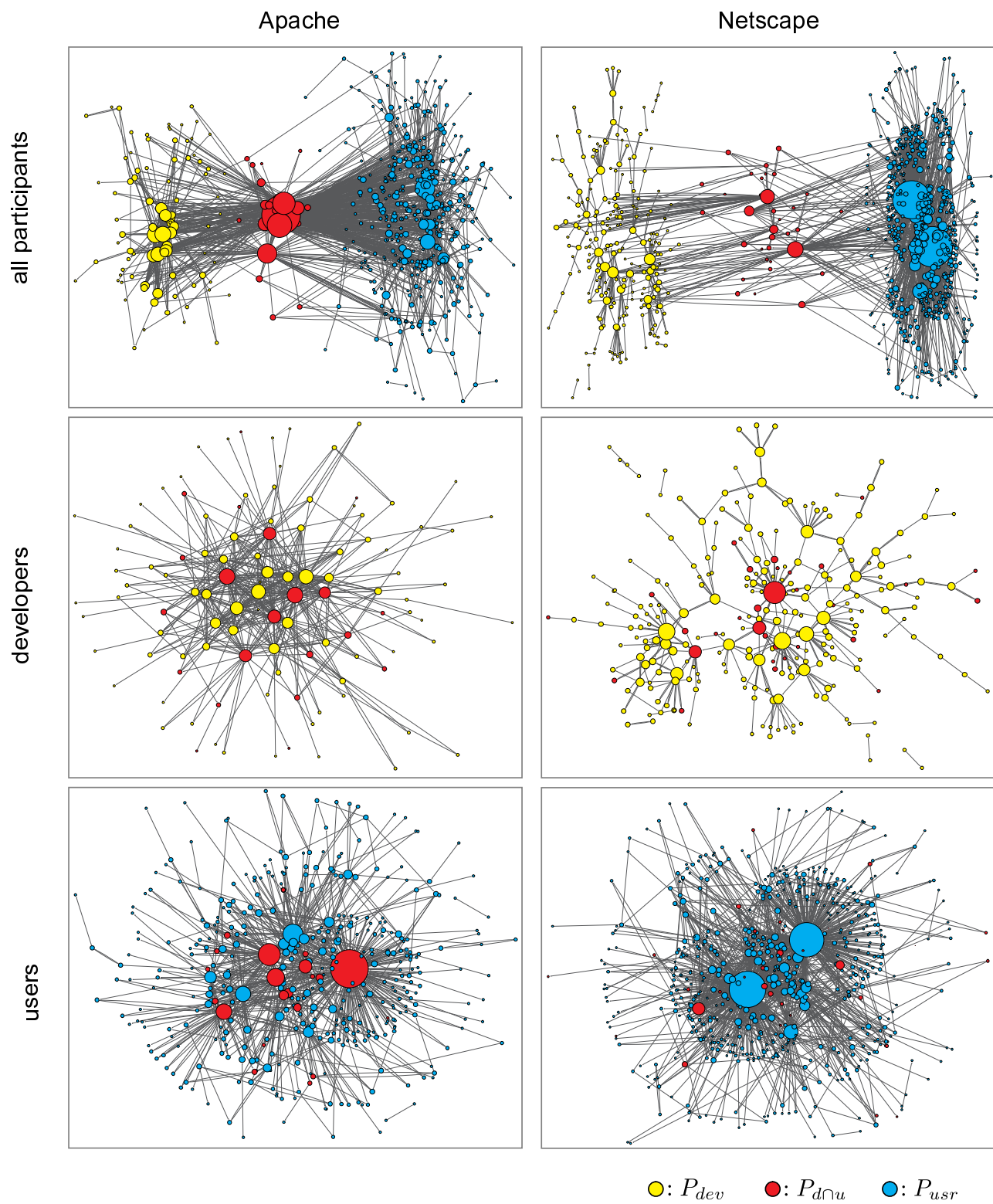


Figure 3: result all network

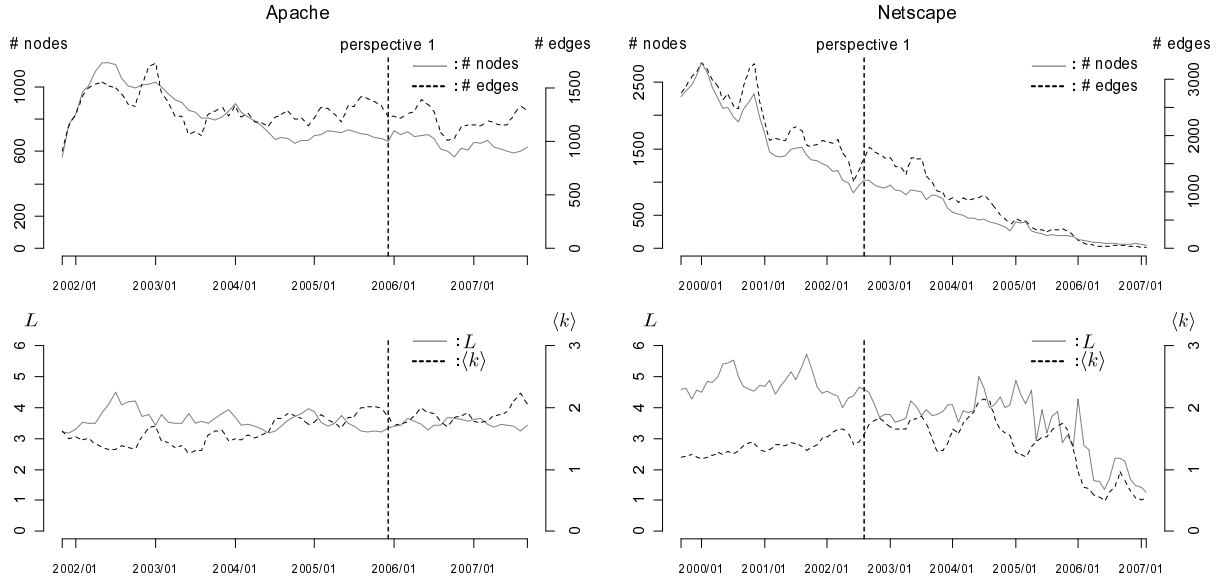


Figure 5: result graph metrics

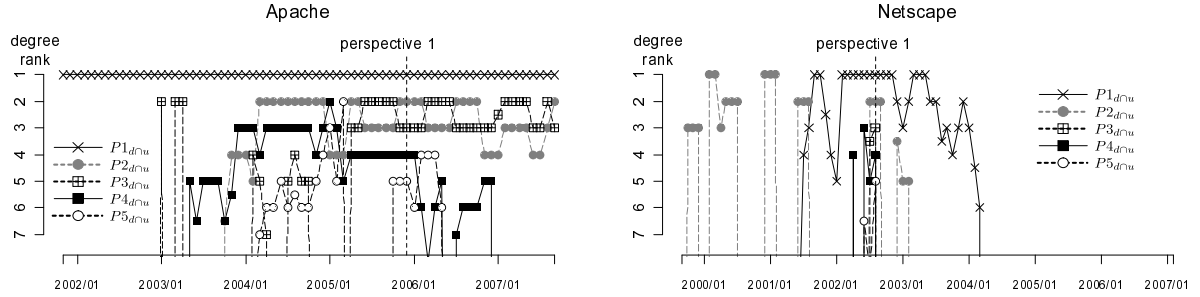


Figure 6: result degree ranking

それに対して、Netscape は参加人数、エッジ数共に単調に減少する傾向にあり、2000 年ごろのピーク時にはノード数が約 3500 であったのが、観点 1 の 2002 年 8 月には約 1000、2005 年には約 500 であった。2000 年から 2005 年にかけての平均パス長 L は緩やかに減少はしているが、ネットワーク規模の縮小度合いと比べて変化は小さい。

観点 1 では Apache はネットワーク構造に変化の少ない安定した時期、Netscape はネットワークの規模が縮小傾向にある時期を分析していたといえる。

次に、ネットワーク全体の構造の変化ではなく、次数の多い $P_{d \cap u}$ 5 人を着目した分析の結果を説明する。観点 1 で対象とした期間における次数が n 番目に多い $P_{d \cap u}$ を $P_{n d \cap u}$ と記す。図 6 に $P_{n d \cap u} (n = \{1, 2, \dots, 5\})$ それぞれの次数ランクの変化を示す。

Apache では分析期間とした約 6 年間の全ての期間において $P_{1 d \cap u}$ が常にトップであることが分かる。また、トップ 2, 3 についても 2004 年 1 月から分析終了時期の 2007 年 9 月まで、残り 4 人の $P_{n d \cap u}$ が占有している。図 5 よりネットワークの構造そのものが安定していることが確認できたが、両コミュニティ間の接続点となる $P_{d \cap u}$ 自体にも大きな変化がないということが分かる。

Netscape の $P_{1 d \cap u}$ は、観点 1 の時期の前後 1 年間は高い次数ランクを保っているが、それ以外の時期ではランク外となっている。また $P_{2 d \cap u}$ については、ランクの変動が激しく、多くの発言を行う時期とそうでない時期が数ヶ月単位で変動している。 $P_{3 d \cap u}$, $P_{4 d \cap u}$, $P_{5 d \cap u}$ は 5 ヶ月のみでランク外となっている。

6. DISCUSSION

6.1 Perspective 1

開発者コミュニティは L と C よりネットワーク構造そのものの違いがあり、ユーザコミュニティはネットワーク構造ではなく、中心とする人物が異なるということが分かった。

図 3 の配置や同様に、左側は Apache、右側は Netscape、上から順に全参加者、開発者、ユーザである。ノードの色も図 3 と対応している。

まず、図 3 の上側、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 $P_{d \cap u}$ のエッジ数の違いである。表 2 より、全体の平均エッジ数 $\langle k \rangle$ は 1.86 と 1.57 であり大きな違いはない。しかし、図 3 より、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。Netscape はブラウザであるというツールの特性上、ユーザ中心の informal social structure をなしている。

6.2 Perspective 2

ネットワークの規模，構造的特徴いずれも変化の少ない安定したネットワーク構造であるといえる．

6.3 Threads of Validity

7. CONCLUSION AND FUTURE WORK

8. ACKNOWLEDGEMENTS
