

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

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

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るようになってきている。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、OSS 開発のみならず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 informal social structure について分析を行った [3]。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、長期に渡って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 [1]。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 [4]。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 [5]。

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、“How structure of an informal social structure formed by users and developers in an OSS community.” という点については明らかにされていない。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。具体的には、「informal social structure の中心的な人物はだれか?」「開発者、ユーザの両方のサブコミュニティに属する人物はどのような役割を担うか?」という点について着目する。

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 WORK

[コピペ含]

OSS コミュニティの開発形態を分析した事例が、数多く報告されている [?, ?, ?]。Mockus らは、これまで OSS コミュニティでの開発について指摘されていた事柄の正しさについて、CVS データ、バグ報告データから調査し、その結果を報告した。分

析では、開発者全体の人数、主な開発者の人数、バグの含有密度、バグの修正期間などに着目している。Jensen らは、OSS 開発者の role migration や project career advancement process について、3 つの OSS 開発プロジェクトを比較し、その違いについて報告を行っている。これらに対して我々は、開発者とユーザの形成する informal social structure について分析を行い、その調査結果を報告する。

また我々の分析と同様に、OSS 参加者の形成する informal social structure を一つのネットワーク構造と見なし、その構造について分析した研究も行われている。Hossain らは、Enron 社内で用いられていた email の履歴データから、テキストマイニングと中心性に基づくネットワーク分析を行い、社員 (employee) 間のコーディネーション能力について分析した。結果、コーディネーション能力の高い社員は、高い中心性を持つことが明らかとなった。Howison らは、バグ報告スレッドの送受信関係をネットワーク構造として、バグ報告者と中心性との関係について分析を行った。5 つのプロジェクトを扱ったケーススタディにより、巨大なプロジェクトほど中心的な役割を担う開発者が長期に渡って存在していることが明らかとなった。本稿ではこれらの分析方法と同様に、オンラインコミュニケーション履歴データから参加者同士のコミュニケーションネットワーク構造を抽出し分析の対象とする。

これまでに我々と同様、OSS コミュニティを中心性の観点から分析した研究はいくつか報告されている [?][3][1]。Yamauchi らは、開発者の中心性と調整行動との関係をメーリングリストデータから調査した結果を報告している [?]。ケーススタディでは、FreeBSD Newconfig と GNU GCC プロジェクト、2 つの事例を紹介し、開発者間の調整行動を行っている人物が存在することを明らかにした。また、Howison らは、バグ報告者の人数と中心性との関係を、バグ報告データから調査した結果を報告している [3]。ケーススタディでは、SF.net に登録されている OSS コミュニティ、5 つの事例を紹介し、コミュニティの大半で、相当長い期間コミュニティの中心である特定の人物が存在することを明らかにした。特に、Bird らは、我々と同様 Apache コミュニティを対象として、ソースコードの更新履歴から求めた開発者の活動の度合いと開発者の中心性との相関を調査した結果を報告している [1]。ケーススタディでは、多くの開発者と交流している開発者ほど、活動の程度が高いことを明らかにした。

- 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

Since, in general, participants in FOSS communities are geographically-distributed, they rely heavily on online media as a way to communicate each other. So the history of communications among participants on online media such as mailing lists news groups and bulletin board systems helps us extract the informal social structure of participants in FOSS communities. In this paper, we define the informal social structure by sender-receiver relationships on online media. Figure 1 is an example of the informal social structure in FOSS community. The left side in Figure 1 shows a thread tree in online media and the right side is a network structure constructed from the thread tree, regarding

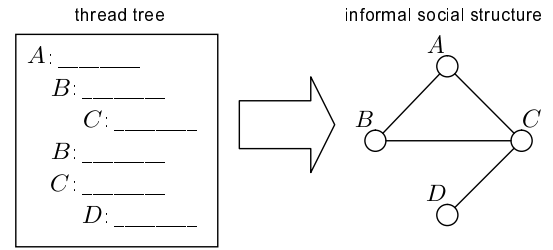


Figure 1: Constructing an informal social structure

nodes and edges as message senders and replies respectively. For example, if a participant B and C replies to a message sent by participant A , edges are linked from node B and C to node A . In this way, we can represent and analyze the informal social structure in FOSS communities as network structures. #ひとまず冠詞のチェックは後回し。

In FOSS community, there are some roles of participants (e.g. developer, tester and users), and subgroups according to the roles of participants. Through using online media, technical matters (for instance, topics related to new features and/or defects of FOSS products) are discussed in a developer group, Q&As about usage of software are discussed in user group, and so forth. Figure 2 illustrates an example of the informal social structure formed by two subgroups in OSS community. P_{dev} (yellow node) is a participant who had sent a message to the developer group (the left circle in the figure), P_{usr} (blue) is a participant who had sent a message to the user group (the right circle in the figure), and $P_{d \cap u}$ is a participant who had sent to both the developer group and user group. By discussing topics corresponding to the role of each subgroup, participants of subgroups can efficiently obtain relevant information for themselves. However, for instance, even if P_{usr} in the user group discussed useful ideas which might improve the quality of software products, P_{dev} in the developer group could not notice the ideas. As the result, the ideas would be buried in communication logs of the user group. Therefore, $P_{d \cap u}$ who participates in the both groups has an important role of transferring feedbacks from users such as feature requests and/or defect reports to developers. In parallel, $P_{d \cap u}$ also have a role of user support because have s/he are knowledgeable about developed software as a developer. We consider that $P_{d \cap u}$ who transfers information discussed in one group to other group largely involves in facilitating communication and collaboration among participants in FOSS community, that is, $P_{d \cap u}$ have strong influence on the informal social structure. In this paper, we focus on the role and network structure of $P_{d \cap u}$ based on the reasons described above.

4. ANALYSIS METHODOLOGY

4.1 Perspective of Analysis

For analyzing the informal social structure formed by oss participants, we describes two analysis perspectives.

4.1.1 static

In this perspective, we perform static analysis about an informal social structure in one period of time based on visualizing and using previously proposed metrics. Specifically,

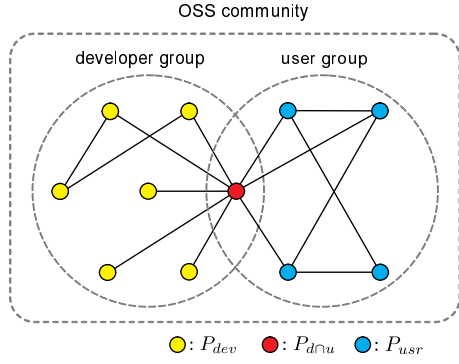


Figure 2: An example of FOSS community

we analyze whole network structure and individual coordination ability. The informal social structure is assumed to be constructed by 3 networks.

- developers network formed by $P_{d\cap u}$ and P_{dev} . (developer group in Figure 2.)
- users network formed by $P_{d\cap u}$ and P_{usr} . (user group in Figure 2.)
- all participants network formed by $P_{d\cap u}$, P_{dev} and P_{usr} . (OSS community in Figure 2.)

Understanding the role of $P_{d\cap u}$ which bridges between developers and users in details becomes possible by analyzing structure of each network with focus on $P_{d\cap u}$.

We use average path length L and clustering coefficient C which are generally used in complex networks for showing the feature of network structure, L and C are described in Section 4.2. It is possible that Freeman's centrality[?] can show metric of individual coordination ability from result from "there is significant correlation between coordination and Freeman's centrality" by Hossain et al [2]. Since, Hossain et al conclude betweenness has the highest correlation with coordination ability, Howison et al conclude betweenness is?? Also, when comparing two nodes in same network (this means same scale of network), degree centrality is essentially same as degree (the number of edges), because degree centrality is a value of degree normalized by a size of network (the number of nodes), In this paper, we use degree that is intuitive understanding as metric of the coordination ability for analyze 3 networks..

4.1.2 dynamic

In FOSS community, anyone can freely participate and they can freely discuss ideas and exchange information on equal footing. For this reason, nformal social structure formed by participant in FOSS community can be said to be living network that grows/decays with time. In order to analyze this living network, only static analysis in the perspective 1 is inadequate and analysis in time series is required. Therefore as an perspective 2, we perform dynamic analysis about an informal social structure on progress of time. In the perspective 2, we analyze change of whole network structure

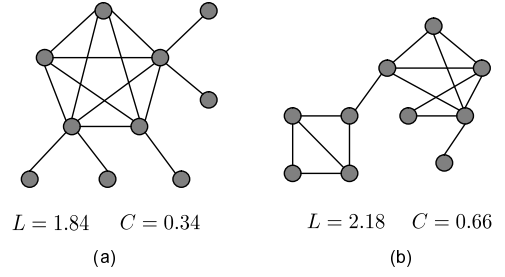


Figure 3: An example of analysis metrics.

and each $P_{d\cap u}$ coordination ability as perspective 1. However, it is difficult to analyze all $P_{d\cap u}$ in all period of time because of network scale, We select top 5 $P_{d\cap u}$ of the degree in perspective 1, and analyze the transition of them.

4.2 Network Analysis Metrics

In this section, we describe average path length L and clustering coefficient C which are used in analysis

4.2.1 Average Path Length

Average path length L is defined as the average number of steps along the shortest paths for all reachable pairs. When average path length of node v_i is L_i and number of nodes is n , L is defined as follows.

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

Generally an information transfer network with smaller L has near distance of each nodes, and the network has high information transmission efficiency. However, since an informal social structure considered in this paper consists of on open media such as mailing list, it differs from the information transfer network. Figure 3 is an example of L and C about two networks. In the Figure 3(a) there is a intensive discussion between five participants although in Figure 3(b) there is two discussion and value of L is low. In this paper, we consider L shows not transmission efficiently but how aggregate discussion are held.

4.2.2 Clustering Coefficient

Clustering coefficient C_i of node v_i is proportion of links between two nodes which are linked to v_i , this means C_i is proportion of triangle between three nodes. Clustering coefficient C of whole network is an average number of C_i , it defined as follows.

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

In the Figure 3(a) there is five participants that has only one edge (they cannot make triangle with other participants), C_i of participants who linked with them are low. Although

a network is separated into two clusters in the case of Figure 3(b), the ratio of connected nodes have connected each other is high. In this paper, we consider C shows how many participants are acquaintance with each other. Note that, although we calculate C to developer network and user network, we does not calculate C to all participants network. Because C is not suitable for applying to a network that contains $P_{d \cap u}$, P_{dev} and P_{usr} as shown in Figure 2, due to P_{dev} and P_{usr} essentially belongs to other network. Since an edge between P_{dev} and P_{usr} does not appear, a triangle of $P_{d \cap u}$ and P_{dev} and P_{usr} does not apper in any case.

5. CASE STUDY

In this section,

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。

- 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 [4] 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 Table 1 に示す。表中の participant 及び mail はユニークな人物及びメールの総数を表す。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,

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

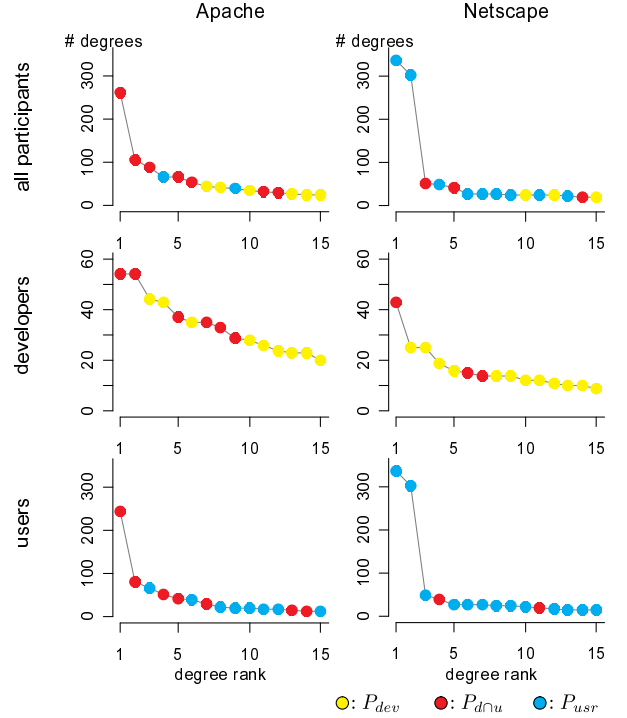


Figure 5: result all degree

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]]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。

Figure 4 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

¹<http://httpd.apache.org/>

²<http://browser.netscape.com/>

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

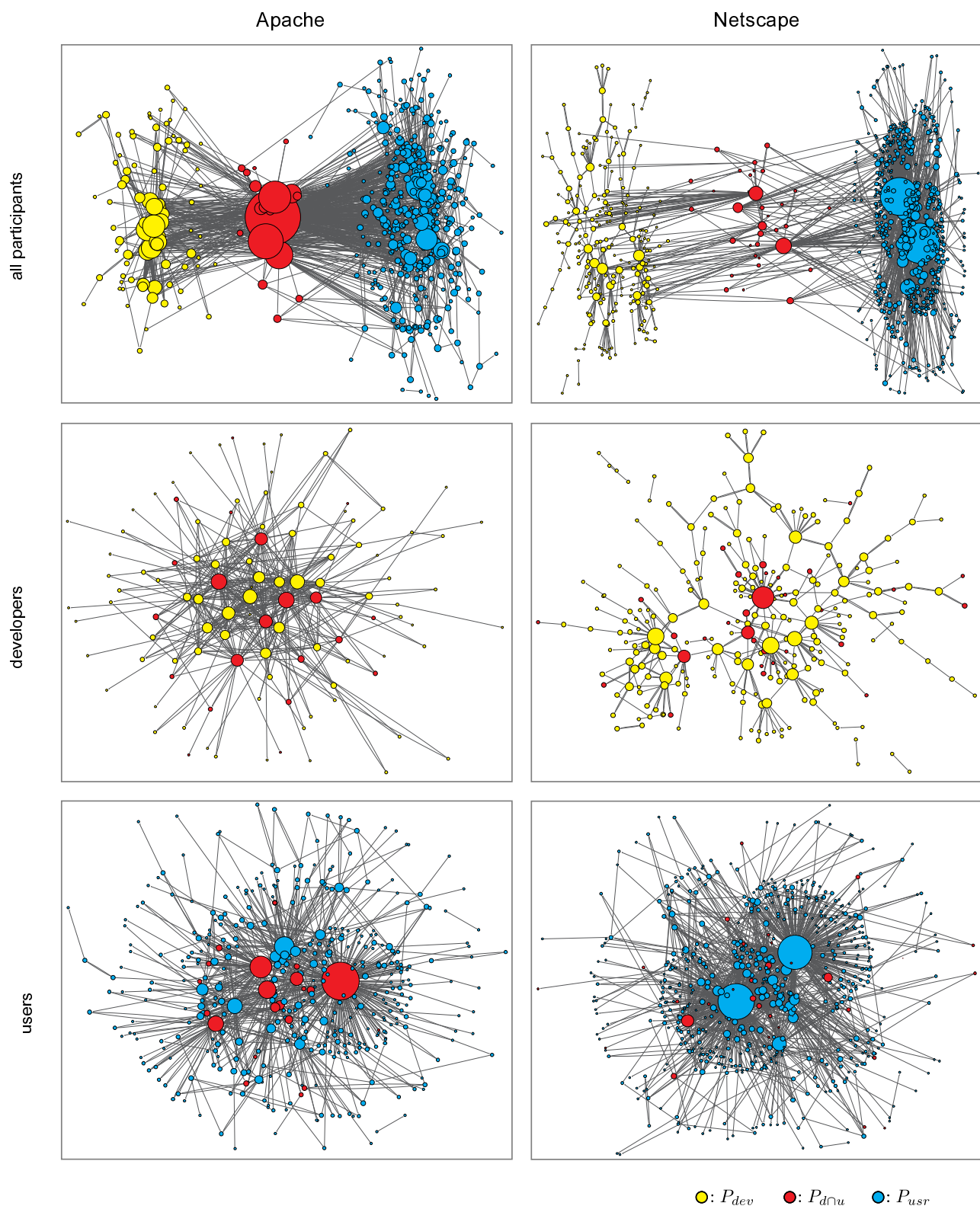


Figure 4: result all network

Table 2: result all metrics

		Apache ($P_{d\cap u}$)	Netscape ($P_{d\cap u}$)
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 (154)
	$\langle k \rangle$	1.55	1.68
	L	3.31	3.11
	C	0.08	0.10

Netscape である。図の上側は全ての参加者 ($P_{d\cap u}$, P_{dev} , P_{usr}) から形成される informal social structure であり、中央は開発者 ($P_{d\cap u}$, P_{dev})、下側はユーザ ($P_{d\cap u}$, P_{usr}) である。ノードの大きさは次数の多さを、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、赤は $P_{d\cap u}$ 、黄は P_{dev} 、青は P_{usr} である。Table 2 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。表中の括弧内の値は $P_{d\cap u}$ のみについての値を表す。各ネットワークのコーディネーション能力の高い (次数の多い) 人物は Figure 4 から確認できるが、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 Figure 5 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。Figure 4 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。各点の色についても Figure 4 のノードの色と対応している。

まず、Figure 4 の上側、Apache と Netscape の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。Table 2 より、 $P_{d\cap u}$ のノード数は Apache は 24、Netscape は 34 と Netscape のほうが数が多いものの、 $P_{d\cap u}$ の総エッジ数は Apache は 720、Netscape は 221 であり、Apache の $P_{d\cap u}$ は少ない人数で多くの発言を行っていることが確認できる。これは、 $P_{d\cap u}$ の次数の差については Figure 4 から視認することができる。

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。Table 2 より、Apache の開発者コミュニティはノード数が 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、クラスターリング係数 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 は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、クラスターリング係数 C が全ネットワーク中最小であった。また Figure 4 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。Figure 5 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。Table 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 Figure 4 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 Figure 5 から確認できる。また Figure 5 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 Figure 6 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。

Figure 6 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。

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。Figure 7 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。Figure 6 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。

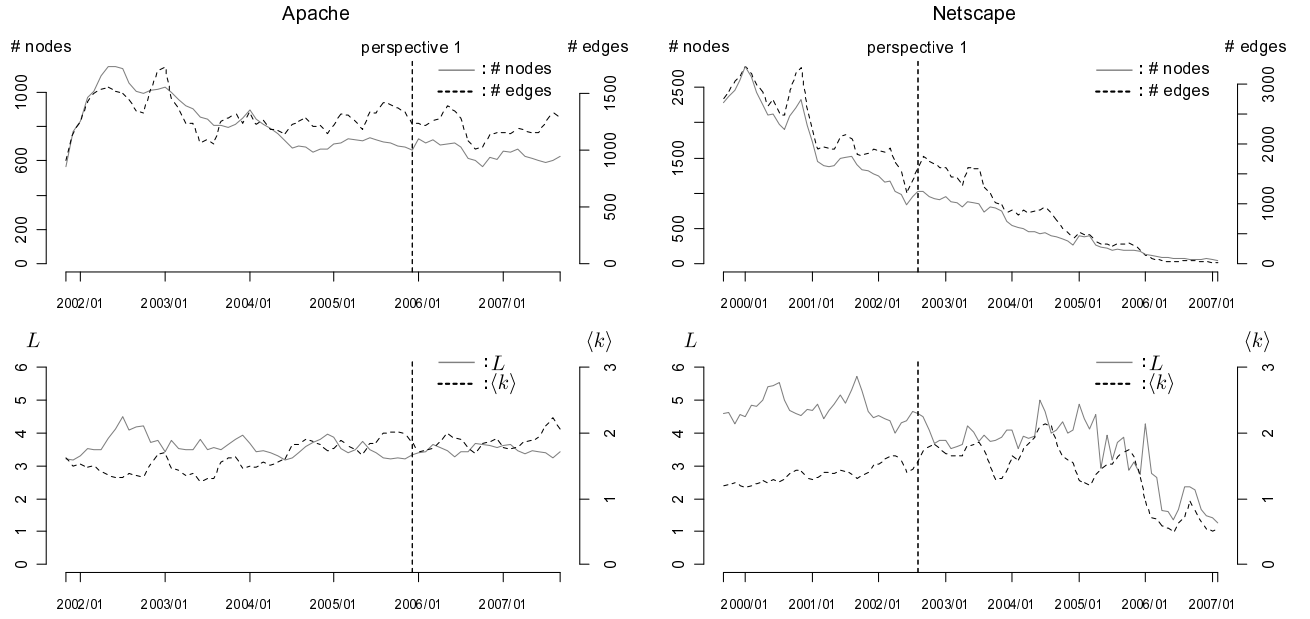


Figure 6: result graph metrics

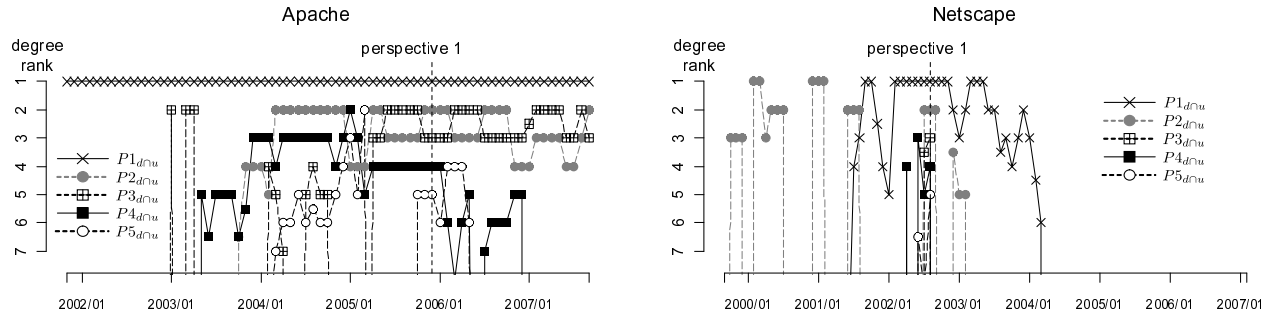


Figure 7: result degree ranking

6. DISCUSSION

6.1 ****

観点1及び観点2の分析により、ApacheとNetscapeでは、1) 開発者ネットワークはネットワーク構造そのものの違いがある、2) ユーザネットワークは中心とする参加者の種類が異なる、3) 全参加者のネットワークでは $P_{d \cap u}$ の働きに大きな違いがある、という結果が得られた。

開発者ネットワークの違いについては、Apacheの開発者ネットワークは、クラスタリング係数 C が高く、特に30人ほどの開発者同士の密度が非常に高くことから、モチベーションの高い開発者同士が活発に議論を交えているといえる。Netscapeはクラスタリング係数 C が低く、開発者同士が互いに議論し合っているというよりは、Q&Aのような（つまり極端に発言の多い参加者に次数1の参加者が複数繋がるような）コミュニケーション構造であるといえる。また強力なリーダーシップを持つ開発者もないことから、議論が一点に集中しておらず、開発者間での協調作業が難しい構造であると考えられる。

ユーザネットワークの違いについては、Apacheではユーザの質問に積極的に答える人物（次数の多い人物）が $P_{d \cap u}$ である割合が高く、ユーザからの意見を $P_{d \cap u}$ が吸収し開発者コミュニティへ伝達しやすい構造であるといえる。Netscapeはweb

ブラウザであるというツールの特性上、 P_{usr} の数が非常に多い。さらに次数の多い $P_{d \cap u}$ が少なく、2人の P_{usr} が極端に次数が高いことから、ユーザ中心の informal social structure を形成しているといえるが、ユーザの意見を開発者へ伝えるという点ではApacheのほうが優れた構造をしていると考えられる。

全参加者のネットワークでは、Apacheの $P_{d \cap u}$ はNetscapeの $P_{d \cap u}$ よりも、が開発者、ユーザの両方の参加者と積極的にコミュニケーションをとっており、Apacheでは開発者とユーザ間の協調作業が円滑に進められていると考えられる。また観点2の結果から、Apacheでは $P1_{d \cap u}$ は常に次数がトップであり、非常に高いモチベーションを安定して保ち続けている強力なリーダーが存在しているといえる。この結果は開発者コミュニティを分析したHowisonらの結果[?]と一致する。このような、常に高いモチベーションとリーダーシップを持つLinusのような存在がOSSコミュニティの発展と継続のために重要であると考えられる。

6.2 Threats of Validity

本稿では、投稿メッセージのメール処理については、履歴データ中のメッセージのうちメールアドレスが完全に一致する送信者を同一人物と見なしたが、複数のアドレスを使い分ける参加者も存在する。そのため、@以前の文字列を用いて first, Last

を考慮するような処理方法も存在しており [], 今後これらの方法を用いた分析を行う必要がある.

分析の際には参加者の形成するネットワークを無向グラフと見なし分析を行ったが, 実際にはメッセージの送信者と受信者という向きを持った関係が存在する. Bird らの研究 [?] と同様にエッジの向きを考慮した分析を今後行う予定である.

また, 異なる種類のソフトウェアを開発する OSS コミュニティを比較した点についても注意されたい. Apache は HTTP Server ソフトウェアであり, その利用者はある程度の高度な専門的な知識を持っている割合が高いが, Netscape は Web ブラウザであるため, そのユーザ層は非常に幅広い. これらを単純に比較することはよくないけど, ??

ML, ニュースグループの違いについて 1. 後ろに回す 2. このまま 3. ポインタ?

7. CONCLUSION AND FUTURE WORK

8. ACKNOWLEDGEMENTS

This research is being conducted as a part of the EASE project in Comprehensive Development of e-Society Foundation Software program, and Stage Project, the Development of Next Generation IT Infrastructure, and Grant-in-aid for Scientific Research (B) 17300007, 2007 and for Young Scientists (B), 17700111, 2007, by the Ministry of Education, Culture, Sports, Science and Technology of Japan.

9. REFERENCES

- [1] C. Bird, A. Gourley, P. Devanbu, M. Gertz, and A. Swaminathan. Mining email social networks. In In Proceedings of the 2006 International Workshop on Mining Software Repositories (MSR'06), pages 137–143, 2006.
- [2] L. Hossain, A. Wu, and K. K. S. Chung. Actor centrality correlates to project based coordination. In In Proceedings of the 2006 Conference on Computer Supported Cooperative Work (CSCW'06), pages 363–372, 2006.
- [3] J. Howison, K. Inoue, and K. Crowston. Social dynamics of free and open source team communications. In In Proceedings of the 2nd International Conference on Open Source Systems (OSS'06), pages 319–330, 2006.
- [4] E. S. Raymond. The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary. O'Reilly and Associates, 1999.
- [5] Y. Ye and K. Kishida. Toward an understanding of the motivation open source software developers. In In Proceedings of the 25th International Conference on Software Engineering (ICSE'03), pages 419–429, 2003.