

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 government and education institution, 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, which attempt to realize the process of development and success factor in OSS community, have reported[?][?]. According to Raymond[?], users have a key role in OSS community. Developers realize that they have been developing the useful software from the response of users, that is, users improve motivation of developers. And, users report bugs or requests of a new features to developers and also ask questions on usage of the products to other users or developers. Developers discuss the reported bugs and answer the questions to users. We think that communication between developers and users is important factor for success in OSS community.

In this paper, 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 social relationships between the individual actors[?][?]. In particular, we focus on 3 centralities that are measures to know the state of informal social structure. By using the centralities, 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[?]. 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 introduces 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

3 (仮) Collaborative Development in OSS Community

Open source software(OSS) is software for which the source code is available under a license, such as the public domain and GNU GPL. Development in OSS community has the following feature compared to development in a company.

3.1 Software Development in OSS Community

1. aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa
2. aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa
3. aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa
4. aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa
5. aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa

3.2 Hypothesis

Based on the feature of software development in Section 3.1, we frame 3 hypothesis that might be observed from the successful OSS community.

H1 : The communication network increases in density over time

From feature 1, the number of participants increases when the OSS community attracts attention. Then, in any OSS communities, the lowly-motivated participants which join in the OSS community easily quit the community over time, since the participants can quit freely. The number of participants decreases even if a new person join in the community (since a new person is limited). However, in the case of successful OSS community, the highly-motivated participants only remain in the community, thus, the communication network increases in density.

H2 : There are frequent communication between developers

From feature 2, frequent communication between developers need to keep their motivation. Additionally, from feature 3, software development based on bazaar style has no the strict chain of command. Therefore, we believe that each developer develop the OSS through communication with each other rather than independently without communication.

H3 : There are some coordinators which transfer information between the developers group and the users group

Therefore, we think that there are some coordinators which transfer information between the developers group and the users group コーディネータについては、次節で詳細に説明する。

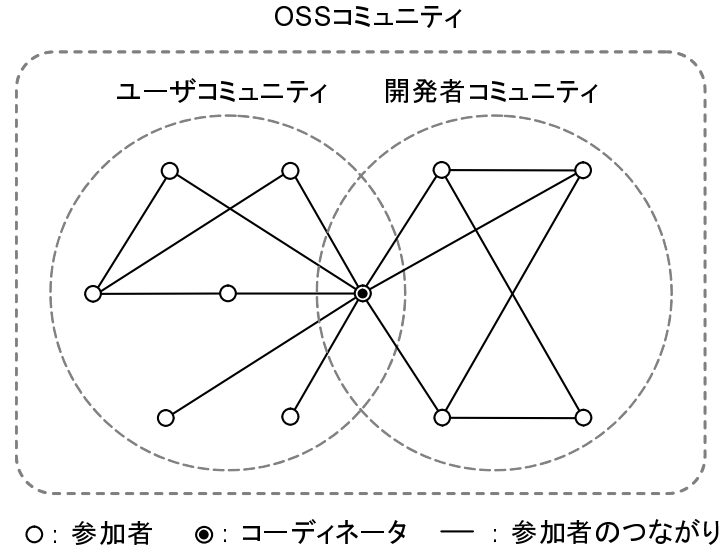
3.3 Role of Coordinator in OSS Community

Fig. 1. Organizational structure of an OSS community

4 Analysis Method

To evaluate the hypotheses in section 3.1, we use social network analysis and evaluate with communication data in some real OSS communities.

Table 1. Characteristics of target communities

Community	Num. of developers	Density of all periods	Num. of posts	Date of communities started	Num. of downloads	Num. of average downloads per a day
Community A	138	0.022	174	04-Jun-01	28,265	16.92
Community B	1	0.013	165	07-Oct-04	7,734,629	17188.06
Community C	11	0.007	766	05-Dec-99	26,000,000	11878.12
Community D	3	0.500	203	29-Dec-03	156	0.21

4.1 Definition of networks

4.2 Analysis Procedure

The followings show the procedure of our analysis using social network analysis (SNA) [?, ?].

Preparation

Calculations of network density by a certain period

Time series analysis

5 Case Study

5.1 Result of analyses

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 2. The vertical axis shows the time and horizontal axis shows statistic values or 3 centralities. The vertical dashed line shows the released period as a major version.

We used not a centrality of each participant but mean of the centrality because the use of a centrality of each participant is difficult for the large OSS community and irrelevant for comprehending the whole impage of the communication network. However, many nodes have no edge in Apache community regardless of time. Under the influence of the nodes, each mean of centralities is close in value to 0. To underline the feature of the communication network, we used mean of the centrality of top 10% centrality participants. We confirm 3 hypothesis in Section 3.2 as follows.

H1 : The communication network increases in density over time

Prediction:

Over time, the lowly-motivated participants quit the community and the highly-motivated participants only remain in the community. Therefore, the number of participants decreases but each centrality increases.

Result:

参加者は 2001 年から XX にかけて増加するものの、それ以降は減少傾向にある。各ネットワークの中心性 (A-3,4,5) は、若干の波はあるものの全体的に右肩上がりの傾向を示している。特に、分析期間の後半にかけて全ての中心性が著しく向上しており、コミュニケーションネットワークが小さくまとまっている。これらのことから、仮説の正しさを確認できる。

H2 : There are frequent communication between developers

予測：(そもそもユーザと比較することで、十分なコミュニケーションがあるといえるのか。→将来、複数のコミュニティで比較をする必要がある。) 一般的な OSS コミュニティでは、ユーザがバグやほしい機能について報告し、開発者がその報告について互いに議論を行う。そのため、開発者グループとユーザグループを比較すると、開発者グループは参加者同士の近さや親密さを表す近接中心性が高い。

結果：開発者グループはユーザグループと比べて近接中心性 (A-5) が高い。ユーザグループでは、メッセージの数が開発者と比べると圧倒的に多いものの、近接中心性が低い。これは、こういったバグがあるよ、や、こういった機能がほしいといった要求であり、ユーザ同士で議論が行われていないことが伺える。それに対し、開発者グループでは、メッセージの数が少ないものの、近接中心性が高く、それぞれの開発者同士で議論が行われていることが伺える。これらのことから、この仮説の正しさを確かめることができる。

H3 : There are some coordinators which transfer information between the developers group and the users group

予測：(コミュニティ内で高ければ十分といえるのか。→将来、複数のコミュニティで比較をする必要がある。) コーディネータは、常にある一定の人数が存在し、同時にコミュニティ内での媒介能力を表す媒介中心性が高い。

結果：コーディネータの数 (A-1) は分析期間のどこを見ても、20~40 人程度と常にある一定以上の人数が存在している。コーディネータの媒介中心性 (A-3) に着目した場合、開発者グループやユーザグループと比較すると、値が大きい。これらのことから、この仮説の正しさを確かめることができる。

6 Discussion

本論文で示すコーディネータとは、開発者と開発者またはユーザとユーザを媒介する人物ではなく、開発者とユーザを媒介する人物のことである。ただし、媒介中心性が表す値は、参加者がいずれのコミュニティ (開発者ユーザ) に属しているかを考慮に入れないため、媒介中心性では表現することができない。そこで、ある一期間におけるコミュニケーションネットワークに着目し、実際にコーディネータがどのようなコミュニケーションを行っていたかについて詳細に分析を行う。

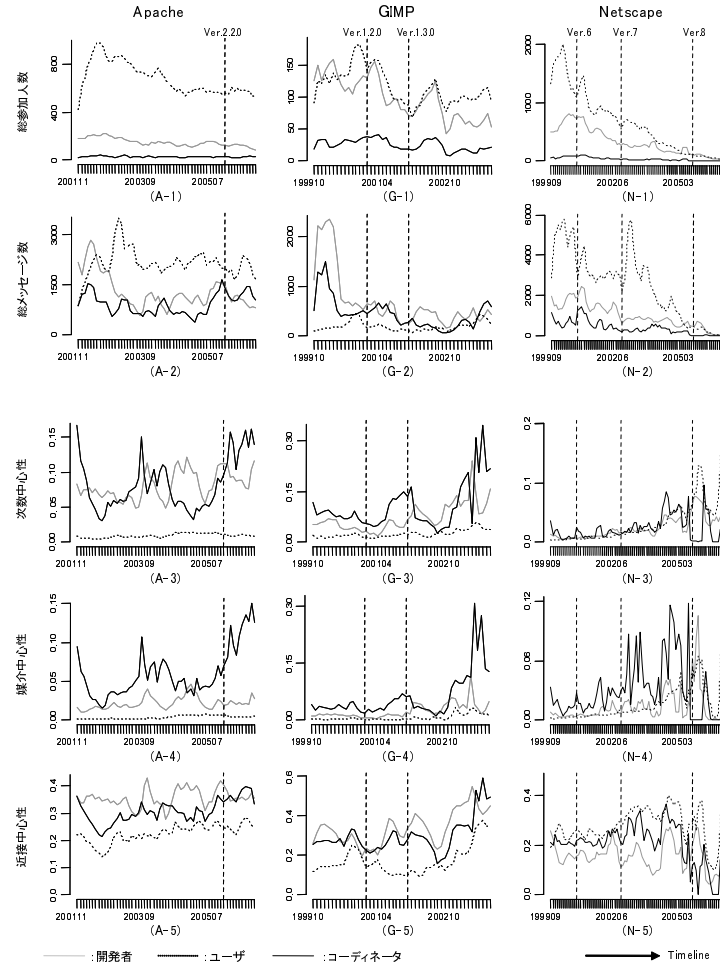


Fig. 2. Changes of average of network centralities of Apache community.

Figure 3 はメジャーバージョンのリリースが行われた一期間におけるコミュニケーションネットワークを可視化したものである。対象時期は Ver.2.2.0 とした。図の左に位置する集団は開発者コミュニティ、右はユーザコミュニティ、中央はコーディネータを表しており、次数の多い上位 5 名のコーディネータ（上位コーディネータ）のノードとエッジを強調している。Table 2 に Figure 3 の上位コーディネータの開発者コミュニティとの次数と、ユーザコミュニティとの次数を示す。

3 と 2 より、コーディネータの持つ開発者コミュニティとユーザコミュニティへの次数のバランスの違いが確認できる。上位コーディネータはバランスよく 2 つのコミュニティとコミュニケーションを取っており、2 つのコミュニティを媒

介するコーディネータとしては理想的な働きをしている。これは、仮説3の正しさを支持する根拠であるといえる。

Table 2. Degree of top 5 coordinators.

コミュニティ		1	2	3	4	5
Apache	開発者との次数	16	30	34	19	28
	ユーザとの次数	89	33	29	21	1

また、この上位5人について、リリース時期3ヶ月分のメールを確認した。そうしたところ、1, 2, 3, 4の人間は、コーディネーションしていることが明らかとなった。例えば、XXはXXである。

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

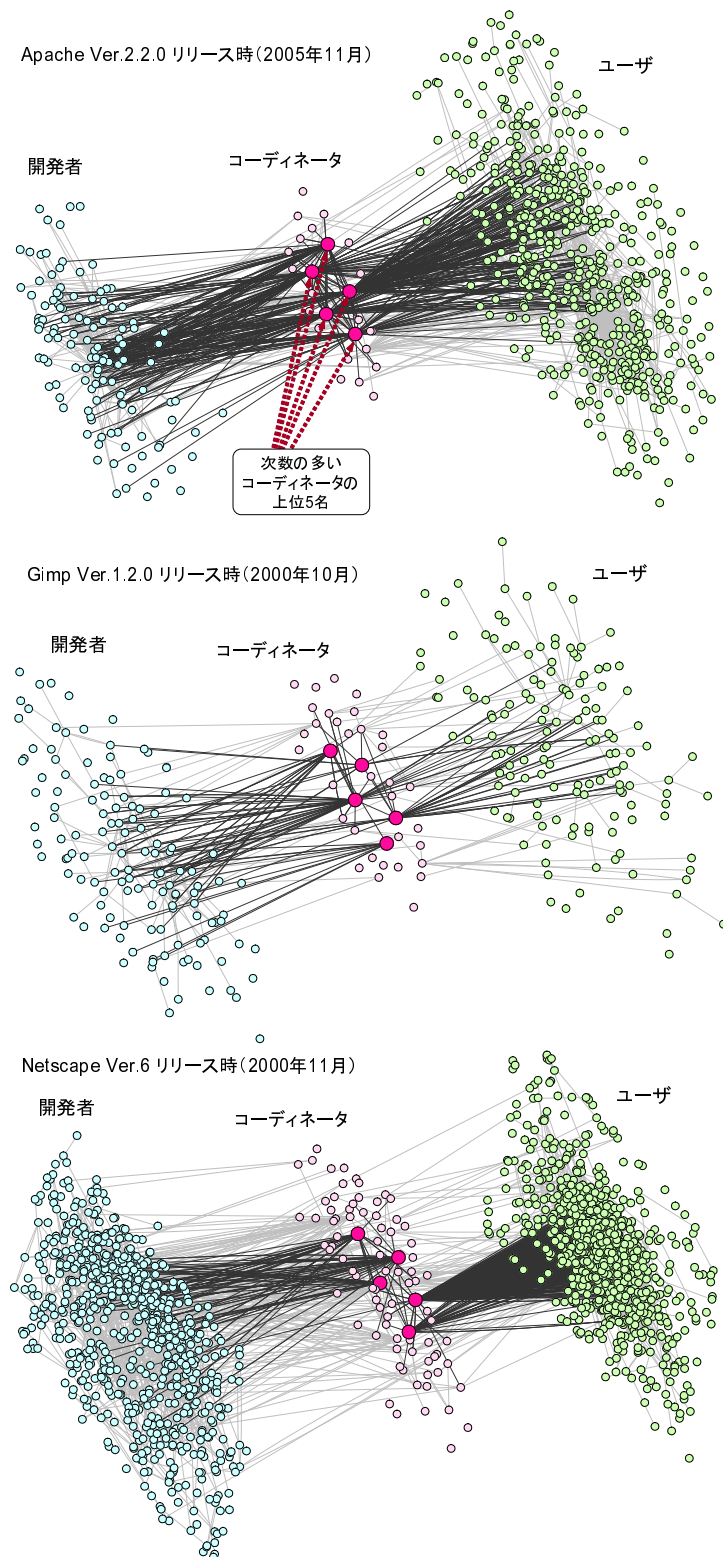


Fig. 3. Communication network of during one period of time.

Index

paragraph, 1