By the Community, for the Community

An Analysis of Open Source Software Projects

May, 2004

Yutaka Hamaoka

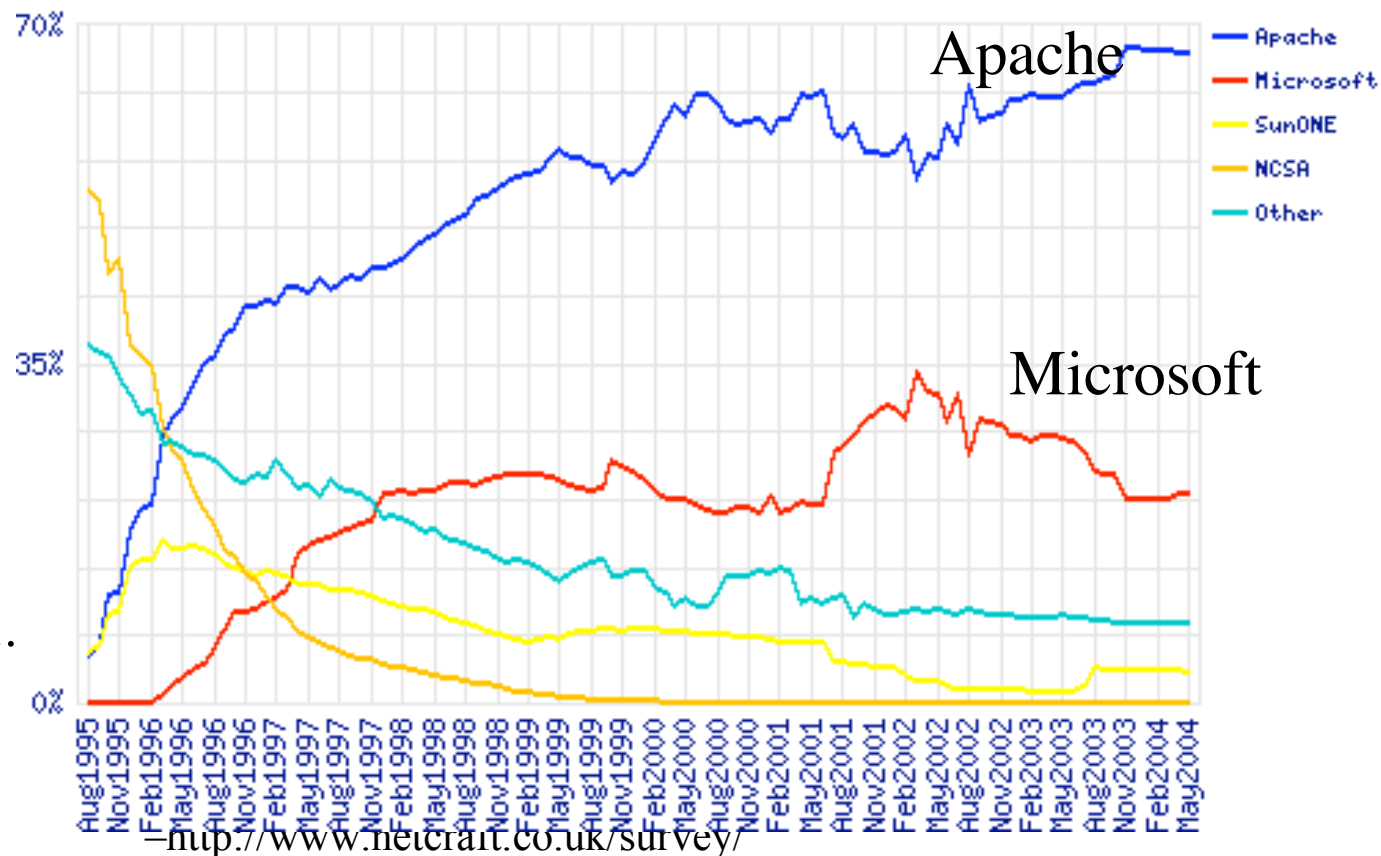Faculty of Business and Commerce

Keio University, Japan

Hamaoka@fbc.keio.ac.jp

- Some of Open Source Software achieved significant Success.

Fig. Market Share for Internet Server Software Across All Domains
August 1995 - May 2004(**50,550,965** sites)

- .....

–http://www.netcraft.co.uk/survey/

What is Open Source Software

– **The Open Source Definition** Version 1.9

1. Free Redistribution

2. Source Code

3. Derived Works

4. Integrity of The Author's Source Code

5. No Discrimination Against Persons or Groups

6. No Discrimination Against Fields of Endeavor

7. Distribution of License

8. License Must Not Be Specific to a Product

9. License Must Not Restrict Other Software

10. License Must Be Technology-Neutral

Source) http://opensource.org/

	Apache	Windows
Developed by	Core Development Team(un paid)	Microsoft (paid programmer)
Source Code	Open If you find a bug and have enough skill, you can fix by yourself! *Enables collective development.*	Closed Even if you find a bug and have enough skill, you must wait until MS fix it.
Free Redistribution	No restriction. No most of cases free of charge.	No: Paid user only

Introduction

- **Research on Open source software**
  - Case study on Development process/team
    - Apache [Mockus et.al.(2000)]
    - Gnome [Koch and Schneider(2000)]
    - Linux kernel [Tuomi(2000)]
    - Freenet [von Krough et.al.(2003)]
  - Survey on Motivation of individuals
    - Linux Developer[Hertel and Herrmann(2003)]
    - Apache help-line[Lakhani and von Hippel (2003)]

- **Limitation**
  - Focusing *single significantly succeeded* Open source software project
    - No comparison.
      - "*Why the project succeeded?*" is unexplained.

Research Question

- Q2 Are Open Source Software Projects really successful?

- Q1 What is *success* of Open Source Software Projects?

- Q3 What makes Open Source Software Project more successful?

Q1 What is *success* of OSSPs?

- Software engineering
  - Success of Software/projects
  - *Software metrics*
    - Software
    - Development Process

- Marketing
  - Success of new product at market
  - Market share, sales, recognition of product, trial, repeat,,,

- Organization behavior
  - Performance of team/individual
    - Innovativeness of product
    - Efficiency(time, cost)
    - Team member's satisfaction, learning,,

For
  - *Commercial products and service.*

Open Source Software Project as a user-centered innovation System

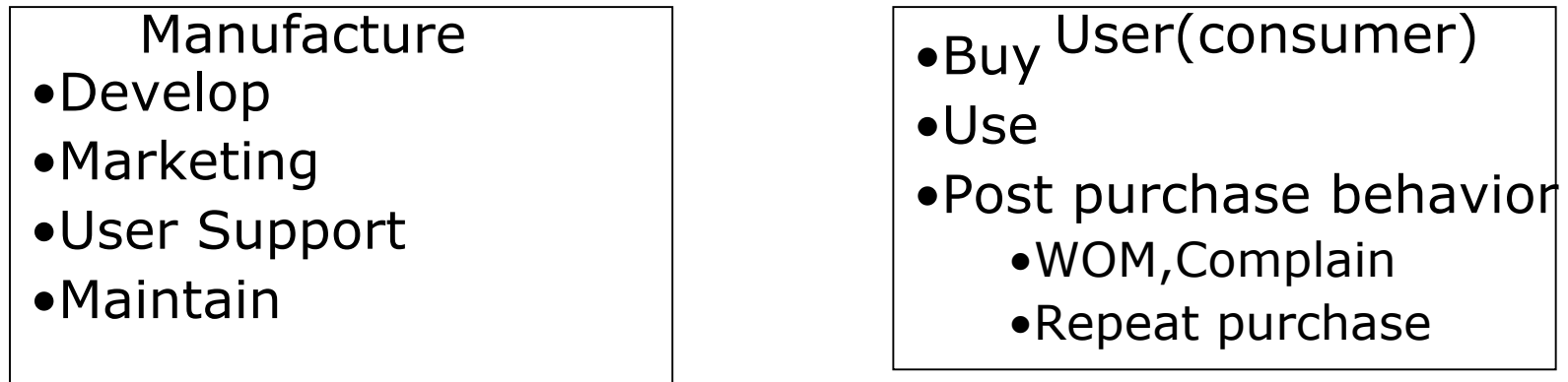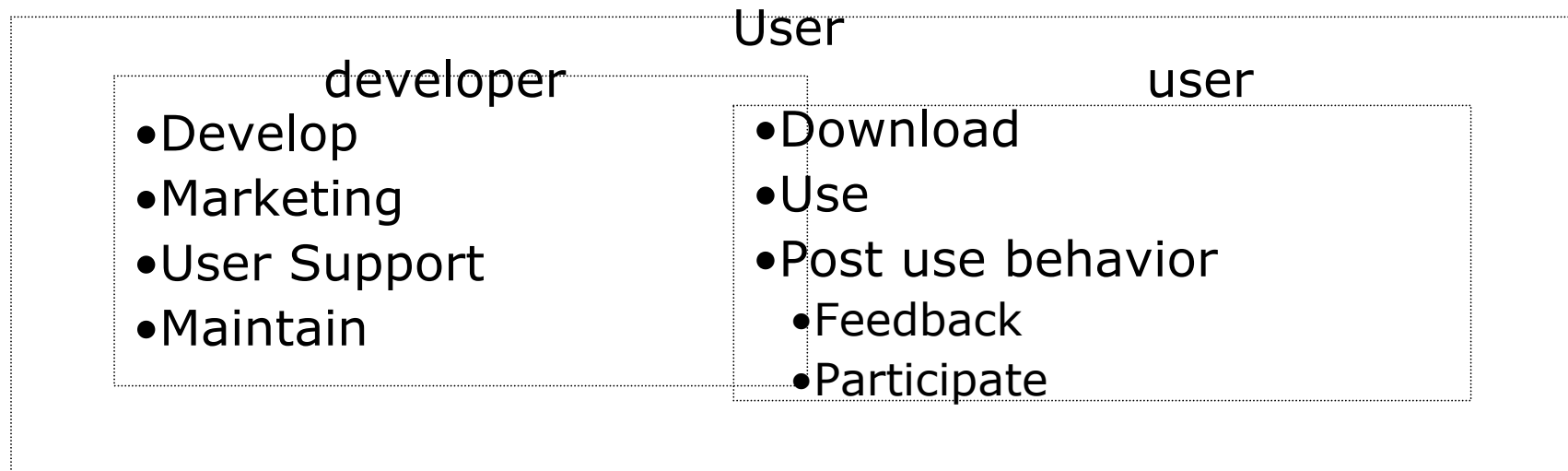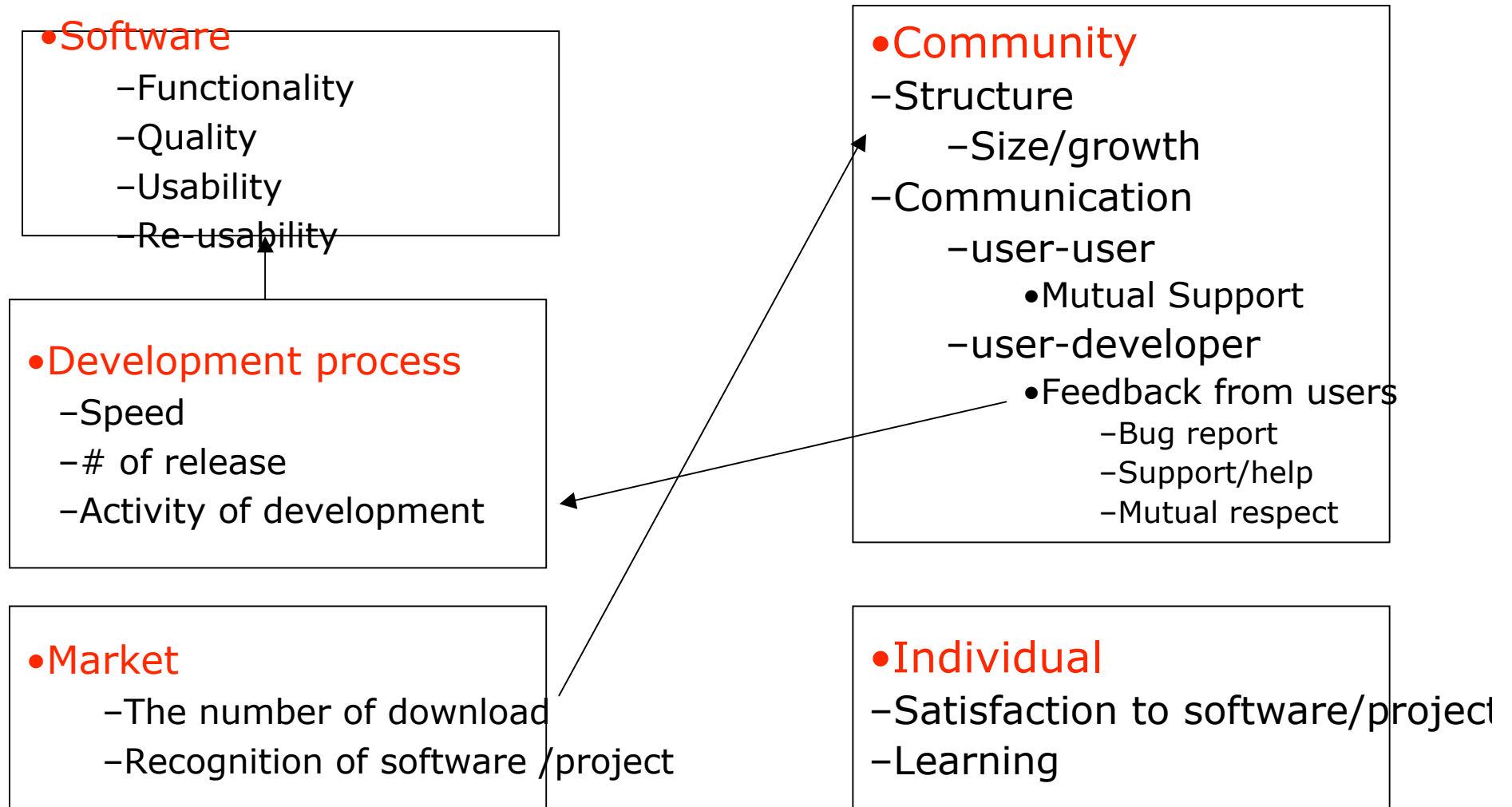## Fig. Manufacture centered Innovation system

Manufacture
- Develop
- Marketing
- User Support
- Maintain

User(consumer)
- Buy
- Use
- Post purchase behavior
  - WOM,Complain
  - Repeat purchase

Fig. User centered Innovation system

User

developer
- Develop
- Marketing
- User Support
- Maintain

user
- Download
- Use
- Post use behavior
  - Feedback
  - Participate

Proposed "Success metrics" of Open Source Software Projects

- **Software**
  - Functionality
  - Quality
  - Usability
  - Re-usability

- **Development process**
  - Speed
  - # of release
  - Activity of development

- **Market**
  - The number of download
  - Recognition of software /project

- **Community**
  - Structure
    - Size/growth
  - Communication
    - user-user
      - Mutual Support
    - user-developer
      - Feedback from users
        - Bug report
        - Support/help
        - Mutual respect

- **Individual**
  - Satisfaction to software/project
  - Learning

Q2 Are Open Source Software Projects really successful?

Data Source

- *Sourceforge.net*
  - Since Oct. 1999, *Sourceforge.net* supporting Open Source Software projects.
  - Hosting 80,671 OSS projects.
  - 842,502 registered users.
  - as of April, 2004

Sample

- Unit of Analysis
  - Project
- Sampling
  - 2,200 projects were randomly selected
    - 10% of 22,000 projects as of May,2001
  - Top 100 projects in terms of page view, download, activity were added to cover really successful projects.
  - Duplicated projects were removed.
  - 2,101 projects

- Then, publicly available data of each project were downloaded from the site.

  - Software/Project Characteristics
  - The number of released files
    - Timing of file release
  - The number of Download
  - Development activity
    - CVS(concurrent versioning system) history
  - Mailing list/Forum archives

This is a project aimed at producing a C++ version of the popular DirectConnect client, enhancing on functionality and performance. Later on, a hub (server) is planned.

- Development Status: 3 - Alpha, 4 - Beta
- Environment: Win32 (MS Windows)
- Intended Audience: End Users/Desktop
- License: GNU General Public License (GPL)
- Natural Language: English
- Operating System: Windows
- Programming Language: C++
- Topic: File Sharing

Project UNIX name: dcplusplus
Registered: 2001-11-17 11:38

Activity Percentile (last week): 99.6136%
View project activity statistics
View list of RSS feeds available for this project

Developer Info

Project Admins:
arnetheduck
tepederzani

Developers: 7
[View Members]

Latest File Releases

Package	Version	Date	Notes / Monitor	Download
Common controls 5.80	For win95/98 users	January 6, 2002	📘 - ✉	Download
dcplusplus	0.401	March 26, 2004	📘 - ✉	Download
dcplusplus (No Installer)	0.401	March 26, 2004	📘 - ✉	Download
dcplusplus source extras	Things needed for compiling	November 26, 2001	📘 - ✉	Download

[View ALL Project Files]

Public Areas   **Latest News**

Public Areas

Project Home Page

Tracker

 - Bugs (**56 open / 1398 total**)
Bug Tracking System

 - Feature Requests (**287 open / 2249 total**)
Feature Request Tracking System

 - Language files (**68 open / 332 total**)
Here you can find different translations of DC++

 - Rejected Features (**20 open / 20 total**)
The features that I definately won't add are listed here - check before posting new ones

---

DocManager: Project Documentation

Mailing Lists (**2 total**)

Latest News

New homepage!
arnetheduck - 2004-04-14 04:26
[Read More/Comment]

Notebook stolen
arnetheduck - 2003-10-13 06:04
(4 Comments) [Read More/Comment]

All welcome...
arnetheduck - 2002-12-22 09:25
(4 Comments) [Read More/Comment]

Paypal...
arnetheduck - 2002-07-28 09:09
(5 Comments) [Read More/Comment]

DC++ Weather report - dark and cloudy...

Project/software Demographics

- OS
  - Linux   45.3%
  - OS independent   32%
  - Microsoft 27.5%
  - BSD   5.5%
  - Mac OS   3.6%
- Topic
  - Internet   21.7%
  - System   20.6%
  - Software development tool
                                        17.2%
  - Communication   16.1%
  - Games   15.5%
- Development status
  - Planning   24.5%
  - Pre Alpha   18.6%
  - Alpha(prottype)   18.6%
  - Beta(before release)27.3%
  - Mature   3.4%

- Programming language
  - C   38.4%
  - C++   28.5%
  - Java   15.9%
  - Perl   13.8%
- License
  - GPL   73.3%
  - BSD   5.9%
  - MIT   1.6%
- Target audience
  - End users   57.6%
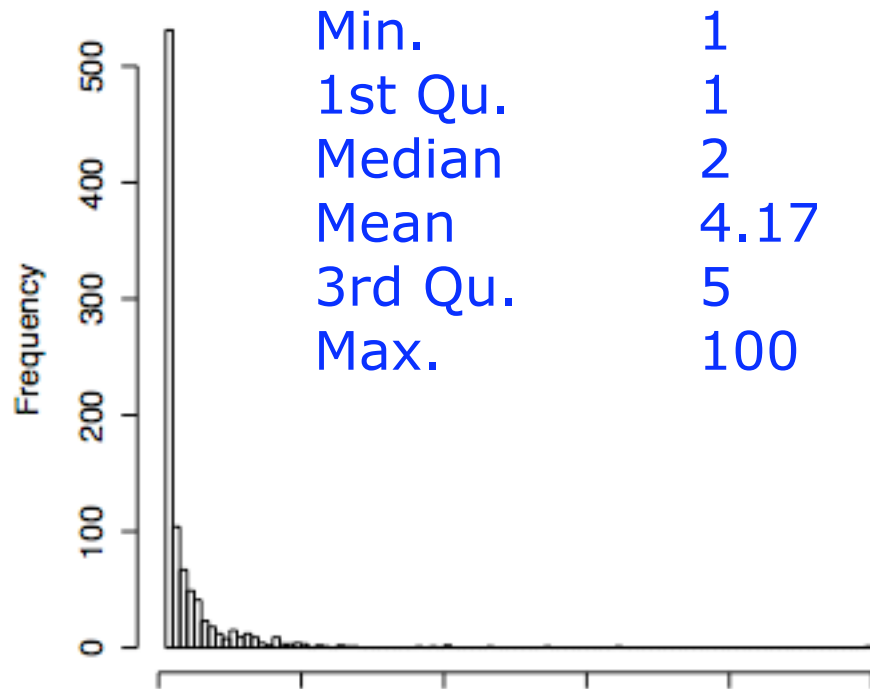 - Developers   64.0%
  - System Administrator 28.3%
  - Other   14.1%

Development metrics

Min.	1
1st Qu.	1
Median	2
Mean	4.17
3rd Qu.	5
Max.	100

Fig. Histogram of The number of Developers(Contributors) per project

N=941

Min.	0
1st Qu.	0
Median	0
Mean	4.8
3rd Qu.	5
Max.	256
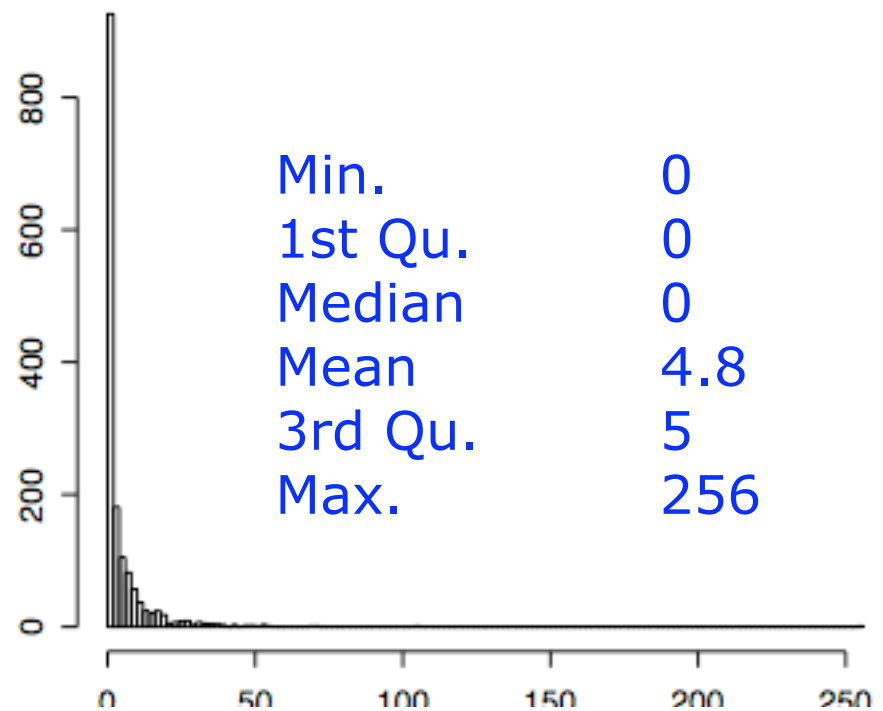
Fig. Histogram of the Number of Released Files per Project

Market Metrics


```
Min.            0
1st Qu.         158
Median          884
Mean            16,660
3rd Qu.         7,102
Max.            1,199,000
```
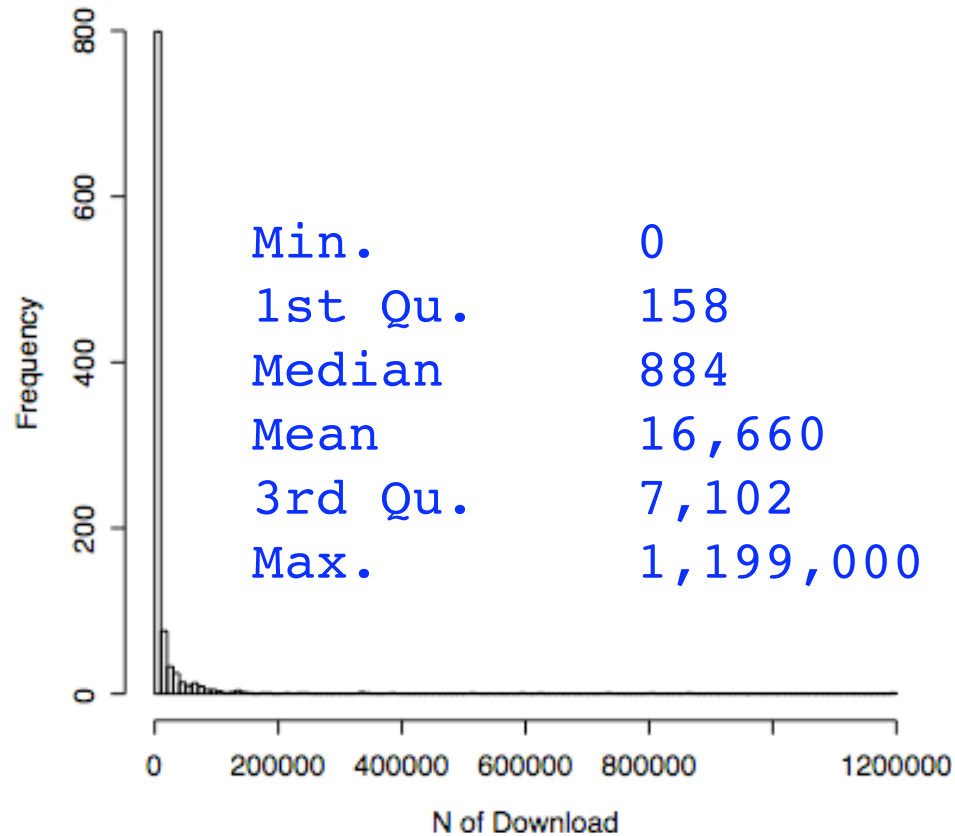
Fig. Histogram of The
Number of Download per
Project


```
Min.            9
1st Qu.         177
Median          818
Mean            39,010
3rd Qu.         8,758
Max.            5,944,000
```
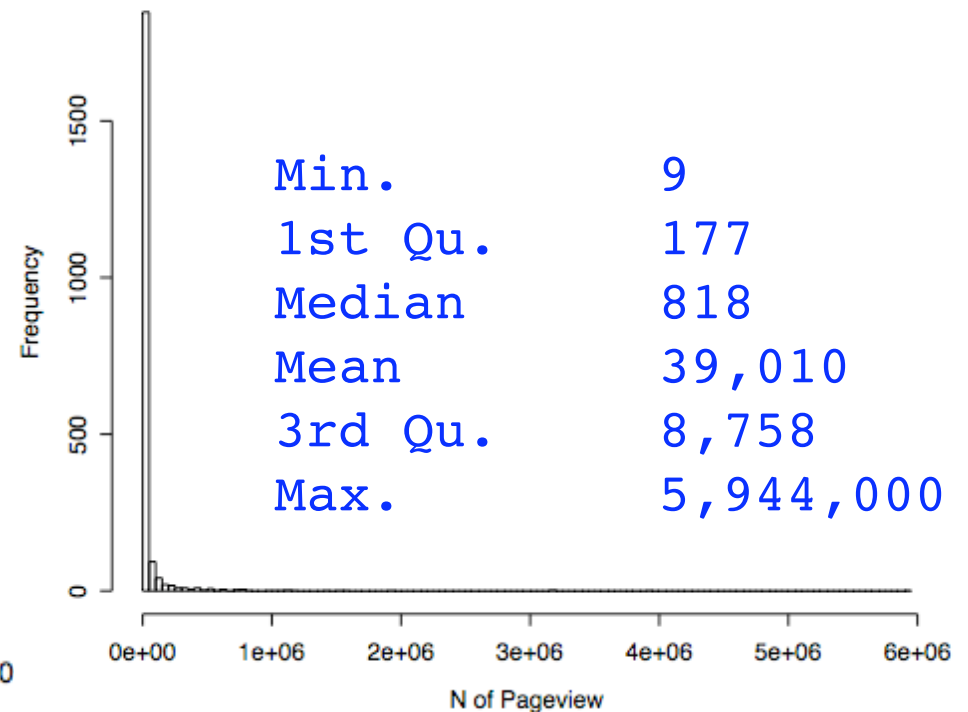
Fig. Histogram of the number
of Projects' Homepage view

Community metrics

Size of *Active User*

The number of members who Posted at least one message to Mailing list, Forum,or Tracking system.

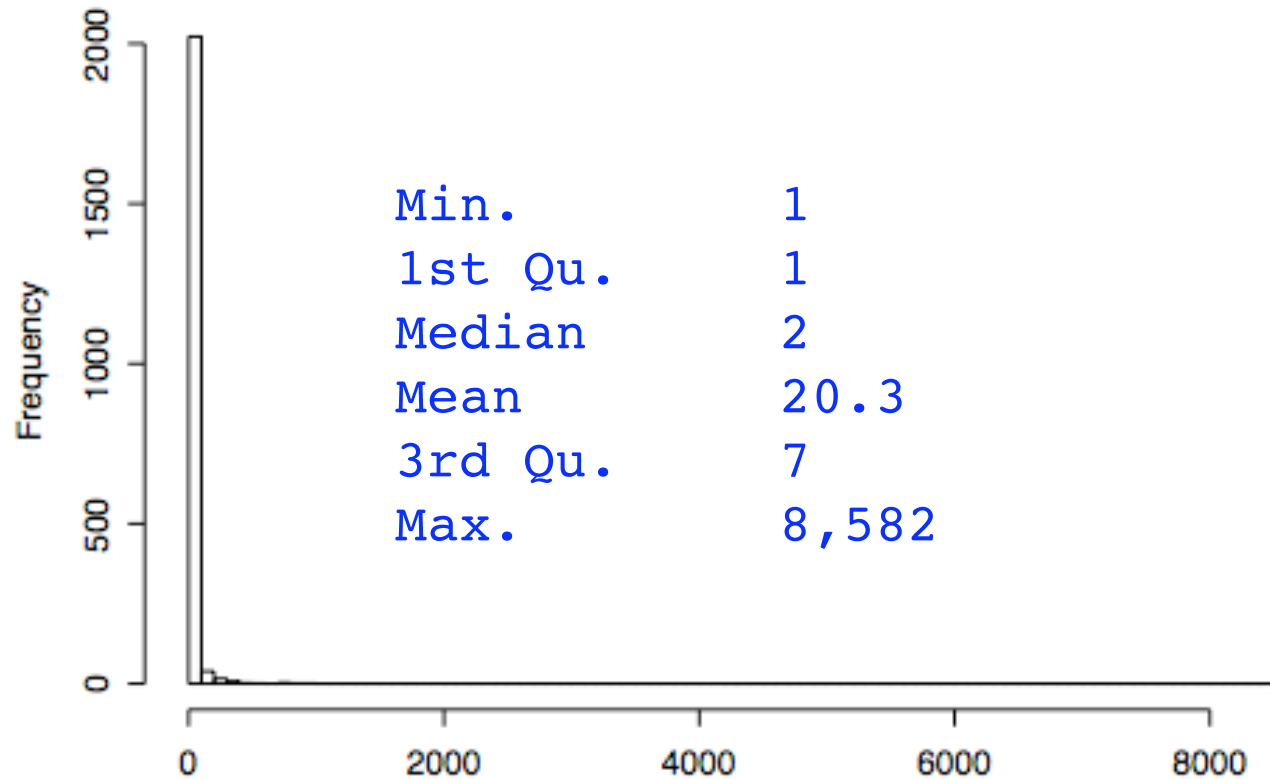|	
Min.	1
1st Qu.	1
Median	2
Mean	20.3
3rd Qu.	7
Max.	8,582

Fig. Histogram of the Number of Active Users per Project

Q3 What makes Open Source Software Project more successful?

- User feedback

"Given enough eyeballs, all bugs are shallow"

- [Raymond(1998) *The Cathedral and the Bazaar*]

- – H: Size of user community is positively related to the number of user feedback.

- User as co-developer
  - "Users are wonderful things to have,…. Properly cultivated, they can become co-developers."
    - *[Raymond(1998)]*

  H: Size of user community is positively related to the size of developers.

- Mutual Support

  H: Mutuality is positively related to posting behavior.

- etc.

Model
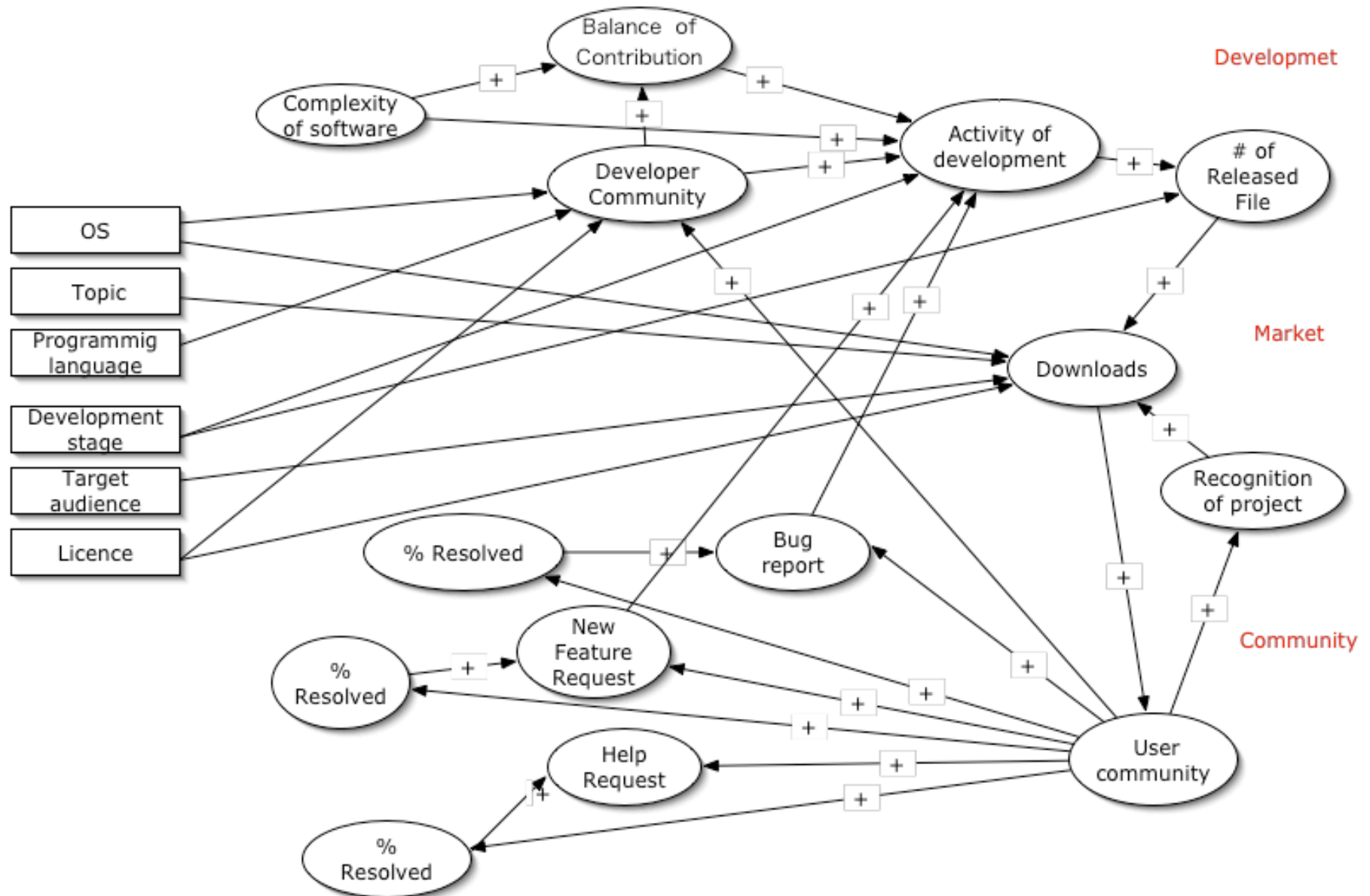
Estimation

- **Simultaneous equations**
  - 3SGLS
- **Sample**
  - Randomly selected
  - Employ CVS(concurrent versioning system)
  - Released at least one file
  - Bug Tracking system
  - Feature request Tracking system
  - Support request Tracking system

  Reduced to 500 projects

Operational definition

- Developer community
  - Log(# of contributors to CVS)

- Development activity
  - Log(# of commitment to CVS)

- Balance of contribution
  - Normalized Entropy :e

$$e = - (1/\log n)* \Sigma \ p_i \log p_i$$

$$p_i = c_i / \Sigma \ c_i$$

  c_i : contributor I's commitment to CVS

  n:# of contributor to CVS

- User community
  - Log(# of users who posted at least one messages)
- User feedback
  - # of {bug report, feature request, support request} messages
- Support
  - % of {bug report, feature request, support request} resolved

Fit

Equation	R2	p
Release	0.171	0.000
Development	0.784	0.000
Dev. Community	0.278	0.000
Balance of Development Task	0.176	0.000
Project Recognition	0.566	0.000
Download	0.504	0.000
User Community	0.164	0.000
Feature Request	0.199	0.000
Bug Rport	0.753	0.000
Support Request	0.630	0.000
Support(Feature)	0.026	0.000
Support(Bug)	0.301	0.000
Support(Support)	0.225	0.000

Results
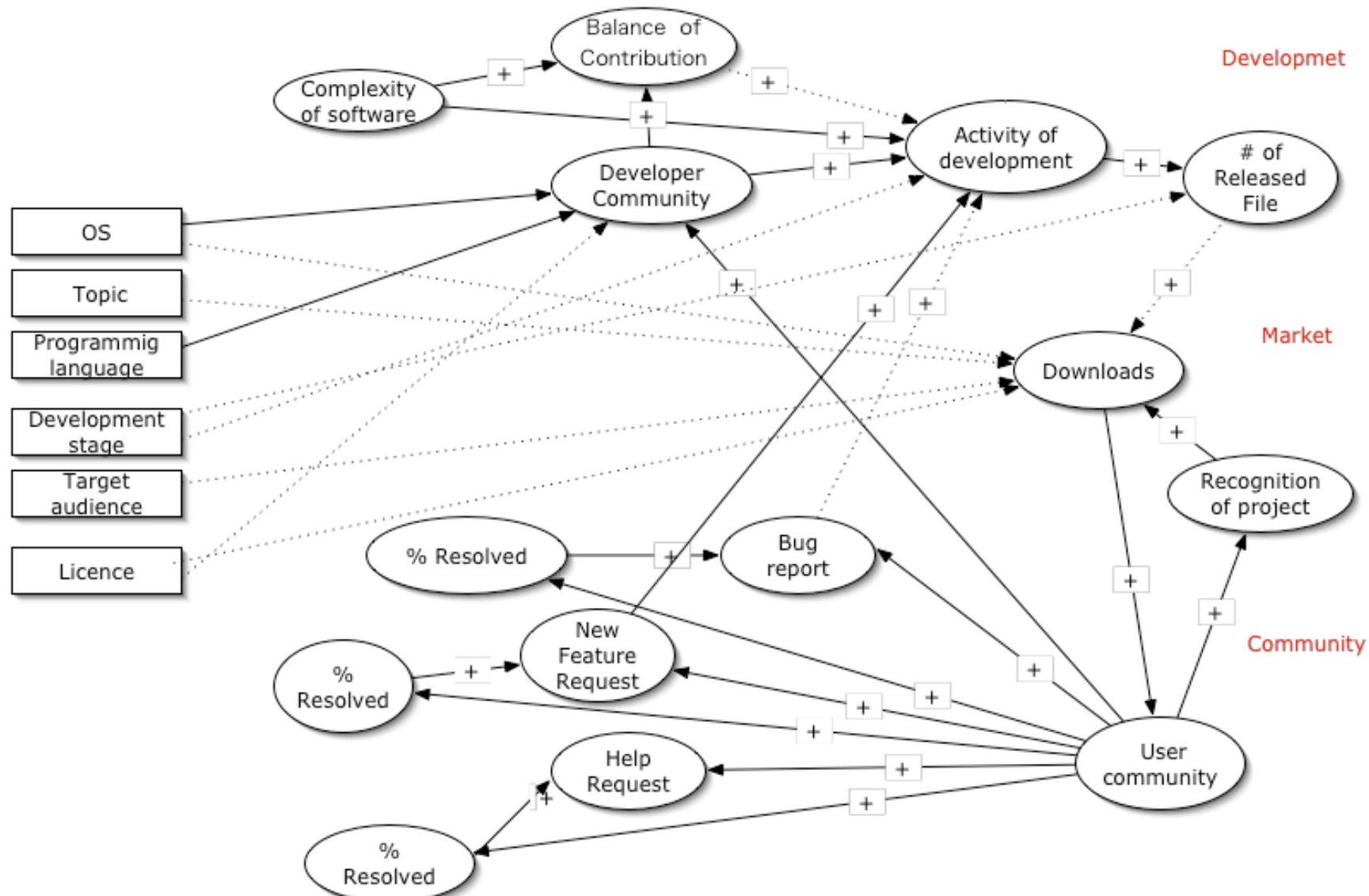
Solid line: Significant at least 5% level
Dashed line: Not significant at 5% level

Results

- User ->Developer conversion (sig)

- Positive Feedback loop
  - User > Project recognition> download>active user

- Feedback from users
  - New feature request (ns) > Development activity
  - Bug report(sig) >Development activity

- Mutual support
  - Resolve bug > bug report
  - Resolve help request > help request

Summary

- Q1 What is *success* of Open Source Software Projects?

  - Software, Development process, Market, Community, individual

  Q2 Are Open Source Software Projects really successful?

  - Most of OSSPs fail.

- Q3 What makes Open Source Software Project more successful?
  - Each variables are interdependent, so answer is not clear. But "user community" is key.

Future Research

- Qualitative analysis
  - Message content, software innovativeness,…

- Multiphase Dynamic network analysis
  - Network data: bug, feature request, help, discussion,,
  - Actor level : # of contribution

- Comparison with very large project.

- Etc.

By the Community, for the Community
An Analysis of Open Source Software Projects

May, 2004

Thank you!

Yutaka Hamaoka

Keio University, Japan

Hamaoka@fbc.keio.ac.jp