AN INFORMALISMS-CENTRIC APPROACH  
FOR  
OPEN-SOURCE PROJECTS HEALTH INSPECTION <?>

# *Thesis proposal*

***Amir Azarbakht***

***Major advisor: Prof. Carlos Jensen***

## ABSTRACT

## 1. INTRODUCTION

### *1.1. PROPOSED CONTRIBUTIONS*

## 2. BACKGROUND AND RELATED WORK

***3. STATEMENT OF THESIS***

## 4. RESEARCH GOALS AND METHODS

## 4.1. Statistical Analysis Study

***\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_***

***Research Objective 1: To identify relevant metrics associated with unhealthy dynamics in OSS projects' communities.***

Data selection criteria:

Projects with a sizable (>50) developer community.

Forked, as listed on the forked projects database of Gregorio & Jesus.

Availability of data

Bigger view:

hard to identify

potentially ahead of time

if we can objectively identify those interaction patterns in project

to determine health and stability of the project

if we can predict it, we can engineer it.

most prominent markers

how early can you detect when things are going out of sync

influx of people new user

message spikes

behavioral metrics into ML algorithms

breaking of the bonds

message pairs

any kinda behavior that

WHAT other features

do I need normalization?

Social interaction theory

Flaming

shorter .. angrier emails<?>

practical outcomes of project separating

visualization to show those 3 metrics

can be artifact of that behavior

ML algorithm can identify

contact ML

Stat

**Table** 0 **Projects forked with "Differences among developer team" as forking reason.**

| ***Original Project*** | ***Forked Project*** | ***Fork Year*** | ***Fork Month*** | ***Data Available?*** | ***Data Collected?*** |
| --- | --- | --- | --- | --- | --- |
| ***GNU Emacs*** | ***X Emacs*** | ***1991*** | ***Unknown*** | ***No*** | ***N/A*** |
| ***NetBSD*** | ***OpenBSD*** | ***1995*** | ***October*** | ***Yes, scarce*** | ***N/A*** |
| ***lMule*** | ***xMule*** | ***2003*** | ***June*** | ***No*** | ***N/A*** |
| ***xMule*** | ***aMule*** | ***2003*** | ***August*** | ***No*** | ***N/A*** |
| ***Sodipodi*** | ***Inkscape*** | ***2003*** | ***November*** | ***Yes, SF.net*** | ***<requested>*** |
| ***Nucleus CMS*** | ***Blog:CMS*** | ***2004*** | ***May*** | ***No*** | ***N/A*** |
| ***Beep Media Player*** | ***Audacious*** | ***2005*** | ***October*** | ***Yes, SF.net*** | ***<requested>*** |
| ***OpenWRT*** | ***FreeWRT*** | ***2006*** | ***May*** | ***No*** | ***N/A*** |
| ***ntfsprogs*** | ***NTFS-3G*** | ***2006*** | ***July*** | ***No*** | ***N/A*** |
| ***QtiPlot*** | ***SciDavis*** | ***2007*** | ***August*** | ***No*** | ***N/A*** |
| ***Kamailio*** | ***OpenSIPS*** | ***2008*** | ***August*** | ***Yes*** | ***Yes*** |
| ***Blastwave.org*** | ***OpenCSW*** | ***2008*** | ***August*** | ***No*** | ***N/A*** |
| ***jMonkeyEngine*** | ***Ardor3D*** | ***2008*** | ***September*** | ***Yes, scarce*** | ***N/A*** |
| ***Frog CMS*** | ***Wolf CMS*** | ***2009*** | ***July*** | ***No*** | ***N/A*** |
| ***Aldrin*** | ***Neil*** | ***2009*** | ***Unknown*** | ***No*** | ***N/A*** |
| ***Ffmpeg*** | ***libav*** | ***2011*** | ***March*** | ***Yes*** | ***Yes*** |

***RQ1.1.:***

***RQ1.2.:***

**Table** 0**: Projects forked with "More community-driven development" as forking reason**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ***Original Project*** | ***Forked Project*** | ***Fork Year*** | ***Fork Month*** | ***Data Available?*** | ***Data Collected?*** |
| ***Nethack*** | ***Slash’EM*** | ***1996*** | ***Unknown*** | ***No*** | ***N/A*** |
| ***GCC*** | ***EGCS*** | ***1997*** | ***Unknown*** | ***No*** | ***N/A*** |
| ***PHPNuke*** | ***PostNuke*** | ***2001*** | ***Summer*** | ***No*** | ***N/A*** |
| ***SourceForge*** | ***Savane*** | ***2001*** | ***October*** | ***No*** | ***N/A*** |
| ***QTExtended*** | ***OPIE*** | ***2002*** | ***May*** | ***Not found*** | ***N/A*** |
| ***GraphicsMagick*** | ***Graphics*** | ***2002*** | ***November*** | ***No*** | ***N/A*** |
| ***freeglut*** | ***OpenGLUT*** | ***2004*** | ***March*** | ***Yes, SF.net*** | ***Yes*** |
| ***SER*** | ***Kamailio*** | ***2005*** | ***June*** | ***Yes*** | ***Yes*** |
| ***Mambo*** | ***Joomla!*** | ***2005*** | ***August*** | ***No*** | ***N/A*** |
| ***PHPNuke*** | ***RavenNuke*** | ***2005*** | ***November*** | ***Not found*** | ***N/A*** |
| ***Compiere*** | ***A Dempiere*** | ***2006*** | ***September*** | ***N/A*** | ***N/A*** |
| ***Compiz*** | ***Beryl*** | ***2006*** | ***September*** | ***No*** | ***N/A*** |
| ***SQL-Ledger*** | ***LedgerSMB*** | ***2006*** | ***September*** | ***N/A*** | ***N/A*** |
| ***Hula*** | ***Bongo*** | ***2006*** | ***December*** | ***No*** | ***N/A*** |
| ***CodeIgniter*** | ***KohanaPHP*** | ***2007*** | ***May*** | ***Not found*** | ***N/A*** |
| ***Asterisk*** | ***Callweaver*** | ***2007*** | ***June*** | ***Yes*** | ***Yes*** |
| ***OpenOffice.org*** | ***Go-oo.org*** | ***2007*** | ***October*** | ***No*** | ***N/A*** |
| ***Mambo*** | ***MiaCMS*** | ***2008*** | ***May*** | ***No*** | ***N/A*** |
| ***TORCS*** | ***Speed Dreams*** | ***2008*** | ***November*** | ***Yes, scarce*** | ***N/A*** |
| ***MySQL*** | ***MariaDB*** | ***2009*** | ***January*** | ***Yes*** | ***No, not in mbox format*** |
| ***Nagios*** | ***Icinga*** | ***2009*** | ***May*** | ***Yes*** | ***<requested>*** |
| ***Dokeos*** | ***Chamilo*** | ***2010*** | ***January*** | ***Not found*** | ***N/A*** |
| ***Project Darkstar*** | ***RedDwarf*** | ***2010*** | ***February*** | ***Yes, scarce*** | ***N/A*** |
| ***SysCP*** | ***Froxlor*** | ***2010*** | ***February*** | ***No*** | ***N/A*** |
| ***rdesktop*** | ***FreeRDP*** | ***2010*** | ***March*** | ***Yes*** | ***Yes*** |
| ***GNU Zebra*** | ***Quagga*** | ***2010*** | ***July*** | ***No*** | ***N/A*** |
| ***OpenOffice.org*** | ***LibreOffice*** | ***2010*** | ***September*** | ***No*** | ***N/A*** |
| ***Redmine*** | ***ChiliProject*** | ***2011*** | ***February*** | ***Yes, scarce*** | ***N/A*** |

# ***4.2. Histogram study***

* ***A study to validate the findings of the statistical analysis study.***

***\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_***

***Research Objective 2: To validate the findings of statistical analysis of SNA metrics associated with OSS project viability.***

***RQ2.1:***

***RQ2.2:***

# ***4.3. Social Graph study***

***\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_***

***Research Objective 3: To validate that a social graph representation of the OSS contribution network is an informative representation for project managers.***

***RQ3.1:***

***RQ3.2:***

## 5. TIMELINE

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | ***RQs*** | ***2nd Half 2012*** | ***1st Half 2013*** | ***2nd Half 2013*** | ***1st Half 2014*** | ***2nd Half 2014*** |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

|  |  |
| --- | --- |
| ***Background Research*** | Feb. 2012 – May 2013 |
| ***Data Collection*** | Jul 2013 - Sept 2013 |
| ***Data Cleaning*** | Oct 2013 – Dec 2013 |
| ***Statistical Analysis*** | Jan 2014 – Nov 2014 |
| ***Visualization*** | Mar 2014 – Sept. 2014 |
| ***Initial Implementation*** | ***Nov 2014 – Jan 2015*** |
| ***Pilot Study*** | ***Feb 2015*** |
| ***Full Implementation*** | ***Mar 2015 – May 2015*** |
| ***Main Study*** | ***May 2015 – July 2015*** |
| ***Writing*** |  |
| ***Defense*** |  |

## 6. CONCLUSIONS

## REFERENCES

[1] Azarbakht, A. and C. Jensen, ”Drawing the Big Picture: Temporal Visualization of Dynamic Collaboration Graphs of OSS Software Forks,” in Proceedings of the 10th International Conference on Open Source Systems (OSS 2014), San Jose, Costa Rica, 2014.

[2] Azarbakht, A. and C. Jensen, ”Temporal Visualization of Dynamic Collaboration Graphs of OSS Software Forks,” in Proceedings of the International Network for Social Network Analysis (INSNA) Sunbelt XXXIV conference, St. Pete Beach, FL, USA, 2014.

[3] Azarbakht, A., ”Drawing the Big Picture: Analyzing FLOSS Collaboration with Temporal Social Network Analysis,” in Proceedings of the 9th International Symposium on Open Collaboration (WikiSym + OpenSym 2013). ACM, 2013.

[4] Azarbakht, A. and C. Jensen, ”Analyzing FOSS Collaboration & Social Dynamics with Temporal Social Networks,” in Proceedings of the 9th International Conference on Open Source Systems (OSS 2013) Doctoral Consortium, Koper Capodistria, Slovenia, 2013.

[5] Davidson, J, R. Naik, A. Mannan, A. Azarbakht, C. Jensen, ”Investigating Older Adults’ Experiences with Contributing to Free/Open Source Software,” in Proceedings of the IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC 2014), Melbourne, Australia, 2014.

[6] Azarbakht, A., ”Temporal Visualization of Collaborative Software Development in FOSS Forks,” in Proceedings of the IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC 2014), Melbourne, Australia, 2014.