Developers' Contribution to Structural Complexity in Free Software Projects

Antonio Terceiro

Supervisor: prof. Christina Chavez

Co-supervisor: prof. Manoel Mendonça

Introduction

Developers rewriting entire systems

- **EOG** rewritten from scratch
- **GNOME-session** rewritten from scratch

Why rewriting?

The code became so **complex** that **rewriting pays off.**

Why so much complexity?

- Conventional setting: appointed designers.
- Free software: **evolutionary design**

Complexity is added, little by little, by the developers themselves.

Goal

Determine whether the variations in structural complexity can be explained by attributes of the developers, and under which conditions

Research Questions

Does the developers' level of participation affect structural complexity? (RQ1)

Does individual developers' experience in the project affect structural complexity? (RQ2)

Does individual developers’ experience in specific parts of the project affect structural complexity? (RQ3)

Does specialisation and generalism affect structural complexity? (RQ4)

Background

Structural complexity

- Architectural concern
- Coupling and Cohesion [Darcy et al, 2005]

SC definition

$$SC(p) = \frac{\sum_{m \in M(p)} CBO(m) \times LCOM4(m)}{|M(p)|}$$

[Chidamber and Kemerer, 1994] (CBO)

[Hitz and Montazeri, 1995] (LCOM4)

⬆ Structural complexity

⬆ Maintenance effort
[Darcy et al, 2008]

⬆ Structural complexity

⬆ Maintenance effort

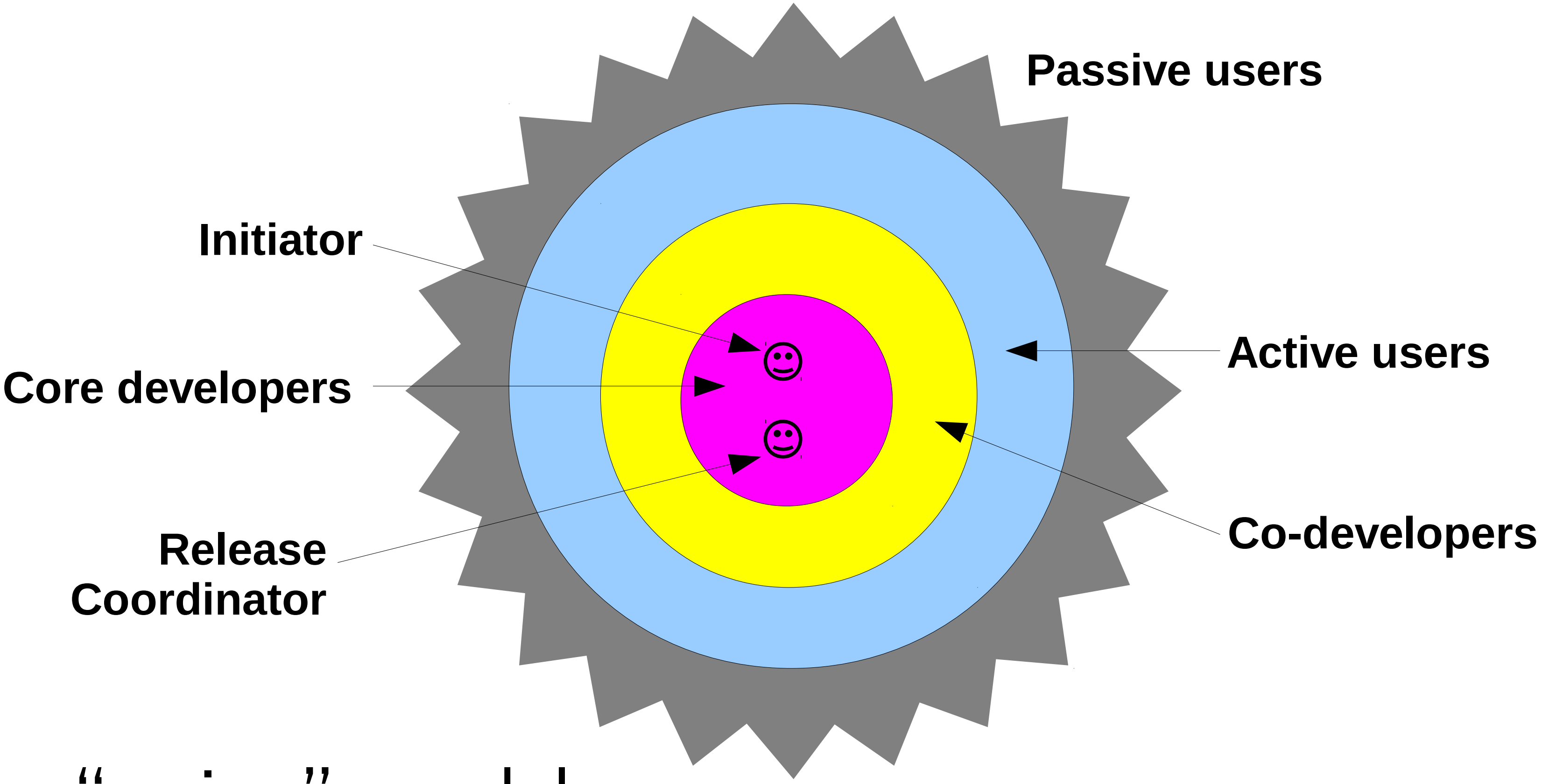
Number of bugs

[Midha, 2008]

⬆ Structural complexity

⬇ Contributions from new developers [Midha, 2008]

Core and periphery in free software projects



The “onion” model.

Adapted from [Crowston and Howison, 2005]

Developer attributes

☑ Level of participation

□ Experience in the project

□ Experience in specific parts

□ Specialization/generalism

Methodology

Research design

analyse changes made to the source code of free software projects

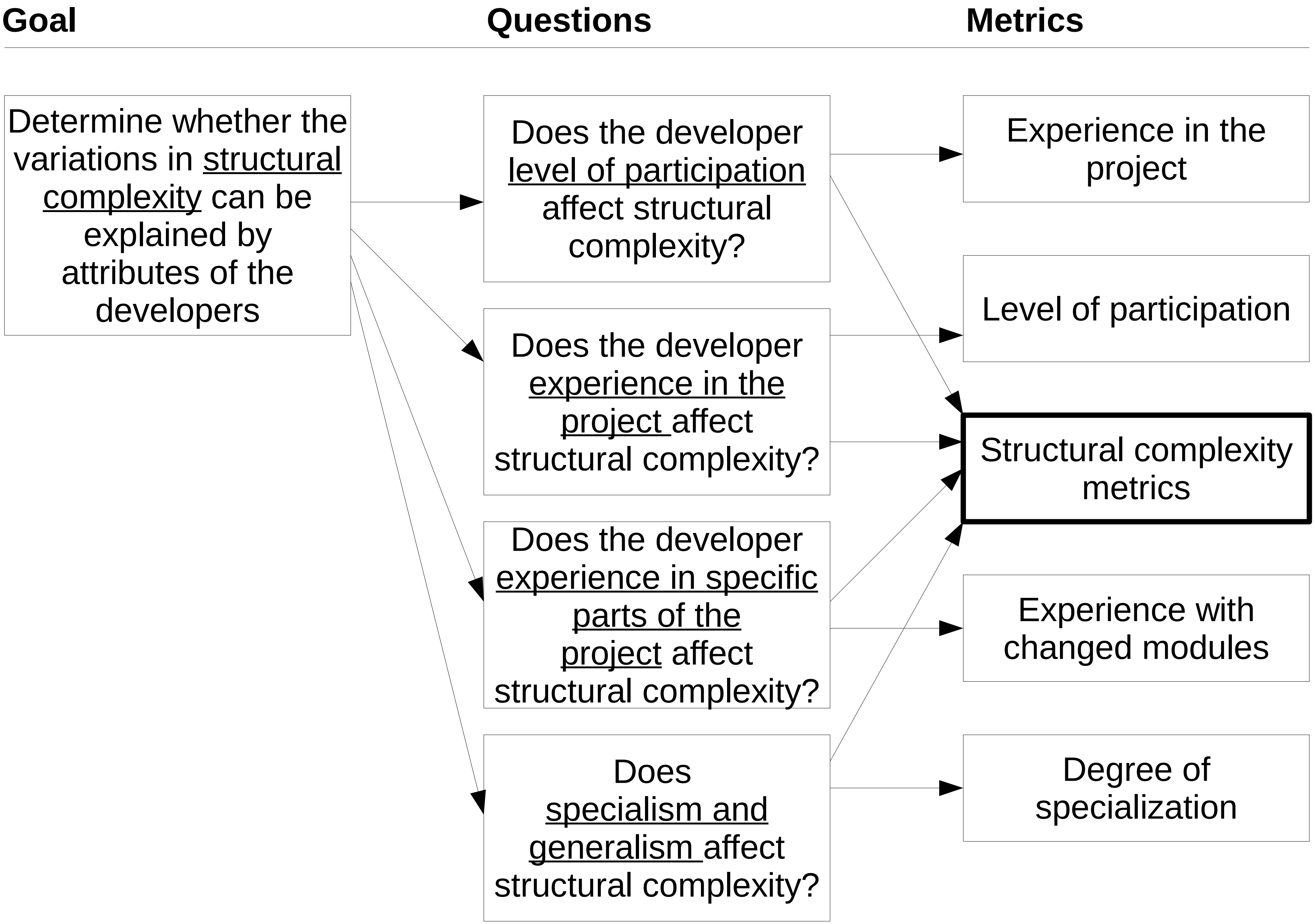
as stored in their version control repositories

for the purpose of characterization

with respect to structural complexity added or removed, level of developer engagement, developer experience in the project, developer experience with the modules changed and developer specialisation

from the perspective of the researcher

in the context of free software projects



Goal

Questions

Metrics

Determine whether the variations in structural complexity can be explained by attributes of the developers

Does the developer level of participation affect structural complexity?

Does the developer experience in the project affect structural complexity?

Does the developer experience in specific parts of the project affect structural complexity?

Does specialism and generalism affect structural complexity?

Experience in the project

Level of participation

Structural complexity metrics

Experience with changed modules

Degree of specialization

Population: changes made to the source code of the projects (“commits”, “checkins”)

Independent variables

- Level of participation, _*L*_
- Experience in the project, _*Ep*_.
- Experience with the modules being changed, _*Em*_
- Degree of Specialisation, _*S*_

Dependent variables

- Structural complexity, SC
- Structural complexity variation, ΔSC
- Absolute structural complexity variation, $|\Delta \mathrm{SC}|$

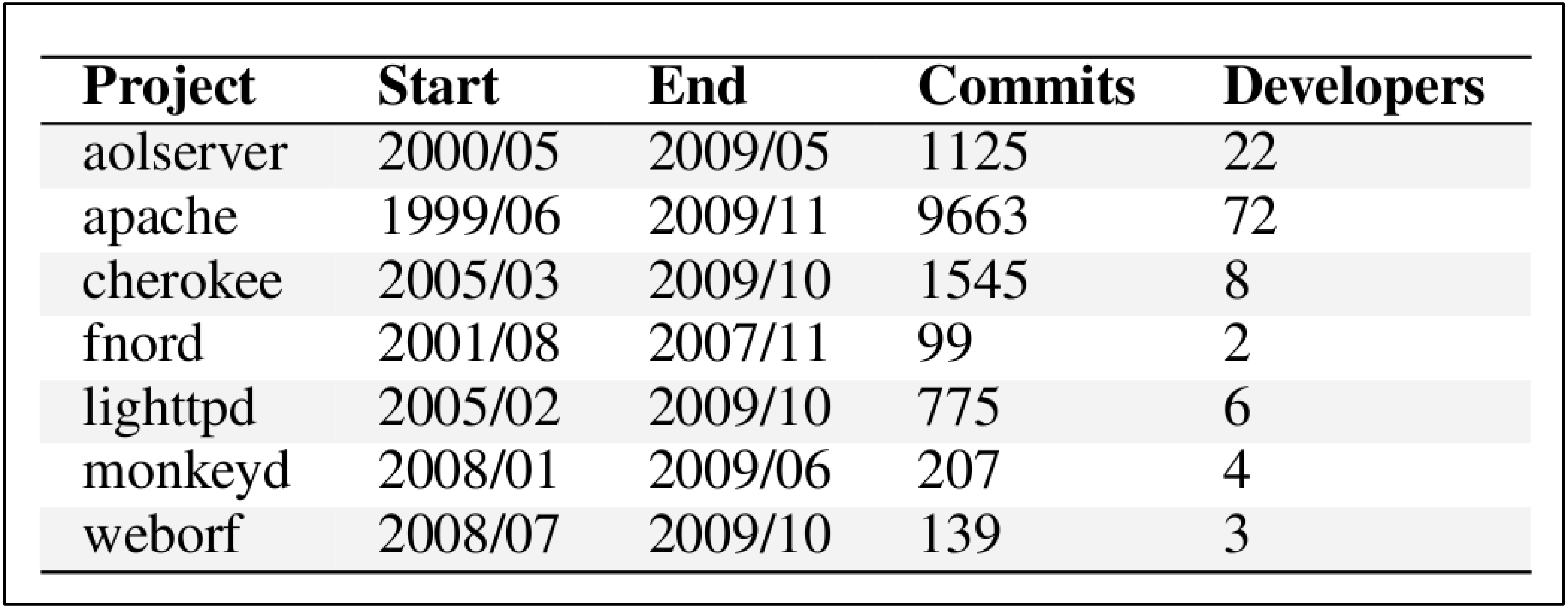
Sample

- Available in Debian GNU/Linux.
- Written in C, C++ or Java
- Publicly accessible version control repository
- Selected application domains

Data collection

- Version control repository mining
- Determine list of relevant changes (those that actually change source code)
- Extract source metrics and change metadata (author, changed files, date etc)
- Load the data in a relational database for further calculations

e.g. first sample



Project	Start	End	Commits	Developers
aolserver	2000/05	2009/05	1125	22
apache	1999/06	2009/11	9663	72
cherokee	2005/03	2009/10	1545	8
fnord	2001/08	2007/11	99	2
lighttpd	2005/02	2009/10	775	6
monkeyd	2008/01	2009/06	207	4
weborf	2008/07	2009/10	139	3

Current state of the work

Planned activities

☑ RQ1 (Level of participation)

□ RQ2 (experience in the project)

□ RQ3 (experience in specific parts)

□ RQ4 (specialisation/generalism)

□ Thesis

Results

Exploratory study on structural complexity evolution in free software projects

[Terceiro and Chavez, 2009]

Results: core developers introduce less SC, and reduce more SC, than periphery ones

[Terceiro et al, 2010b] (submitted)

⬆ Structural complexity

⬇ Attractiveness

[Meirelles et al, 2010]

(submitted)

Collaboration with ***CCSL - IME/USP*** *(Paulo Meirelles, João Miranda, Carlos Santos Jr., Fabio Kon)*

Expected contributions

Models of the relationship between developer characteristics and structural complexity in free software projects

A multi-language source code analysis and visualization toolkit

http://softwarelivre.org/mezuro/analizo

Pending issues

Which operational definitions?

Which analysis technique?

Explain better the context and clarify the boundaries of my research.

Questions?

References (1/3)

S.R. Chidamber and C.F. Kemerer. A metrics Suite for Object Oriented Design. IEEE Trans. Sftware Eng., 20(8):476–493, 1994.

Kevin Crowston and James Howison. The Social Structure of Free and Open Source Software Development. First Monday, 10(2), 2005.

D. P. Darcy, C. F. Kemerer, S. A. Slaughter, and J. E. Tomayko. The Structural Complexity of Software: An Experimental Test. IEEE Transactions on Software Engineering, 31(11):982–995, Nov. 2005.

M. Hitz and B. Montazeri. Measuring coupling and cohesion in object-oriented systems. In Proceedings of the International. Symposium on Applied Corporate Computing, 1995.

M. M. Lehman, J. F. Ramil, P. D. Wernick, and D. E. Perry. Metrics and Laws of Software Evolution-The Nineties View. In Proceedings of the 4th International Symposium on Software Metrics, 1997.

Vishal Midha. Does Complexity Matter? The Impact of Change in Structural Complexity on Software Maintenance and New Developers' Contributions in Open Source Software. In ICIS 2008 Proceedings, 2008.

References (2/3)

Katherine J. Stewart, David P. Darcy, and Sherae L. Daniel. Opportunities and Challenges Applying Functional Data Analysis to the Study of Open Source Software Evolution. Statistical Science, 21:167, 2006.

Antonio Terceiro and Christina Chavez. Structural Complexity Evolution in Free Software Projects: A Case Study. In Muhammad Ali Babar, Bj̈ rn Lundell, and Frank van der Linden, editors, QACOS-OSSPL 2009: Proceedings of the Joint Workshop on Quality and Architectural Concerns in Open Source Software (QACOS) and Open Source Software and Product Lines (OSSPL), 2009.

A. Terceiro, “Developers Contribution to Structural Complexity in Free Software projects,” in Proceedings of the OSS 2010 Doctoral Consortium, W. Scacchi, K. Ven, and J. Verelst, Eds., 2010.

Antonio Terceiro, Luiz Romá́rio Rios, and Christina Chavez. An Empirical Study on the Structural Complexity introduced by Core and Peripheral Developers in Free Software projects, 2010 (***unpublished***)

Paulo Meirelles, Carlos Santos Jr., João Miranda, Fabio Kon, Antonio Terceiro, Christina Chavez. A Study of the Relationships between Source Code Metrics and Attractiveness in Free Software Projects, 2010 (***unpublished***)

References (3/3)

D. L. Parnas, “Software aging,” in ICSE ’94: Proceedings of the 16th international conference on Software engineering. Los Alamitos, CA, USA: IEEE Computer Society Press, 1994, pp. 279–287.

E. Capra, C. Francalanci, and F. Merlo, “An Empirical Study on the Relationship Between Software Design Quality, Development Effort and Governance in Open Source Projects,” IEEE Transactions on Software Engineering, vol. 34, no. 6, pp. 765–782, Nov.-Dec. 2008.

D. Barbagallo, C. Francalenei, and F. Merlo, “The Impact of Social Networking on Software Design Quality and Development Effort in Open Source Projects,” in ICIS 2008 Proceedings, 2008. [Online]. Available: http://aisel.aisnet.org/icis2008/201