

Caracterização da Complexidade Estrutural em Sistemas de Software

Tese de Doutorado em Ciência da Computação
por Antonio Terceiro

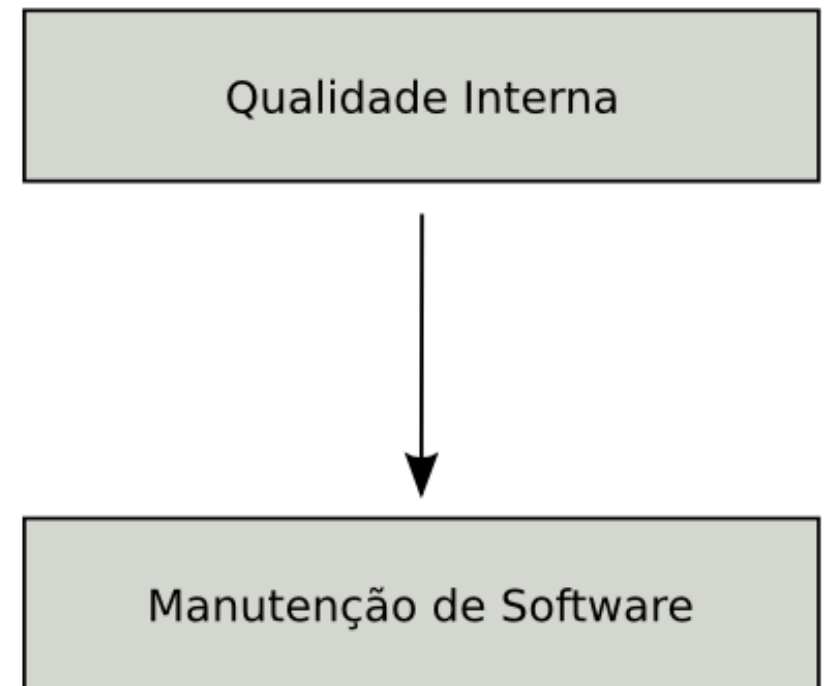
Orientadora: Dr^a. Christina Chavez

Co-Orientador: Dr. Manoel Mendonça

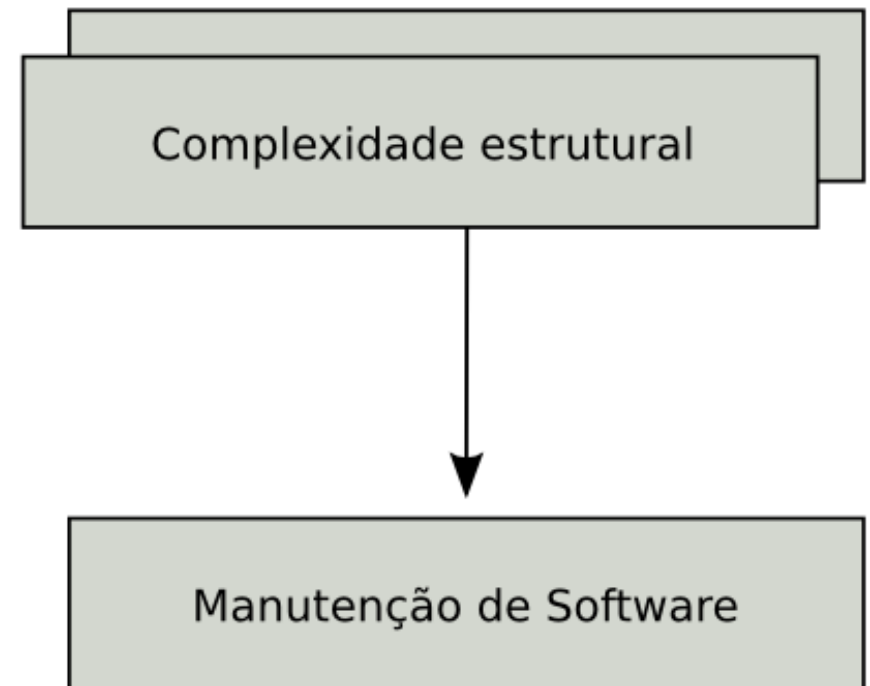
Motivação

Manutenção de Software

Motivação



Motivação



Efeitos da complexidade estrutural em sistemas de software

↑ Complexidade estrutural

↑ Esforço de manutenção

[Darcy et al. 2005]

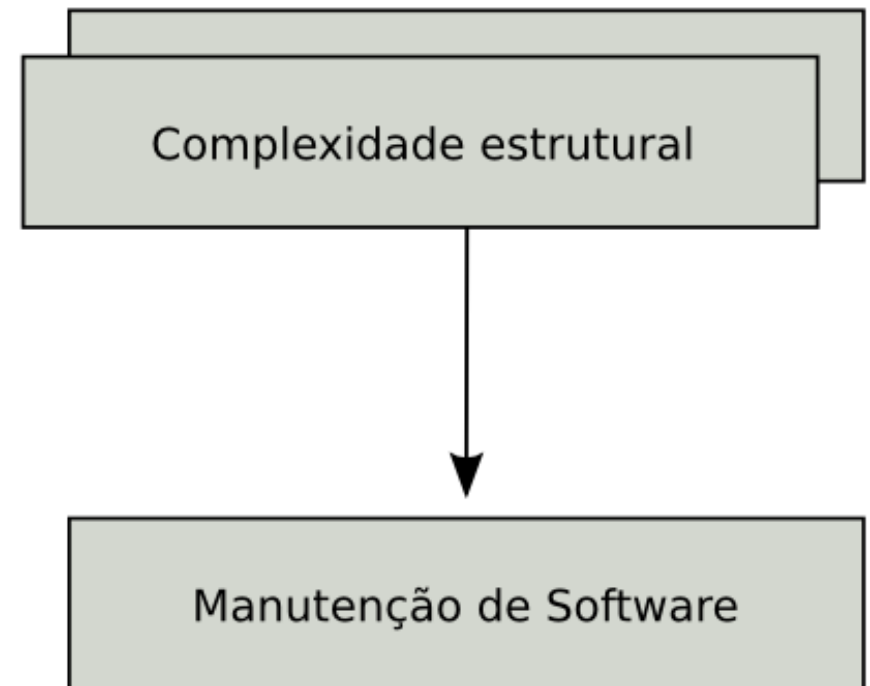
↑ Complexidade estrutural

↓ Novos desenvolvedores

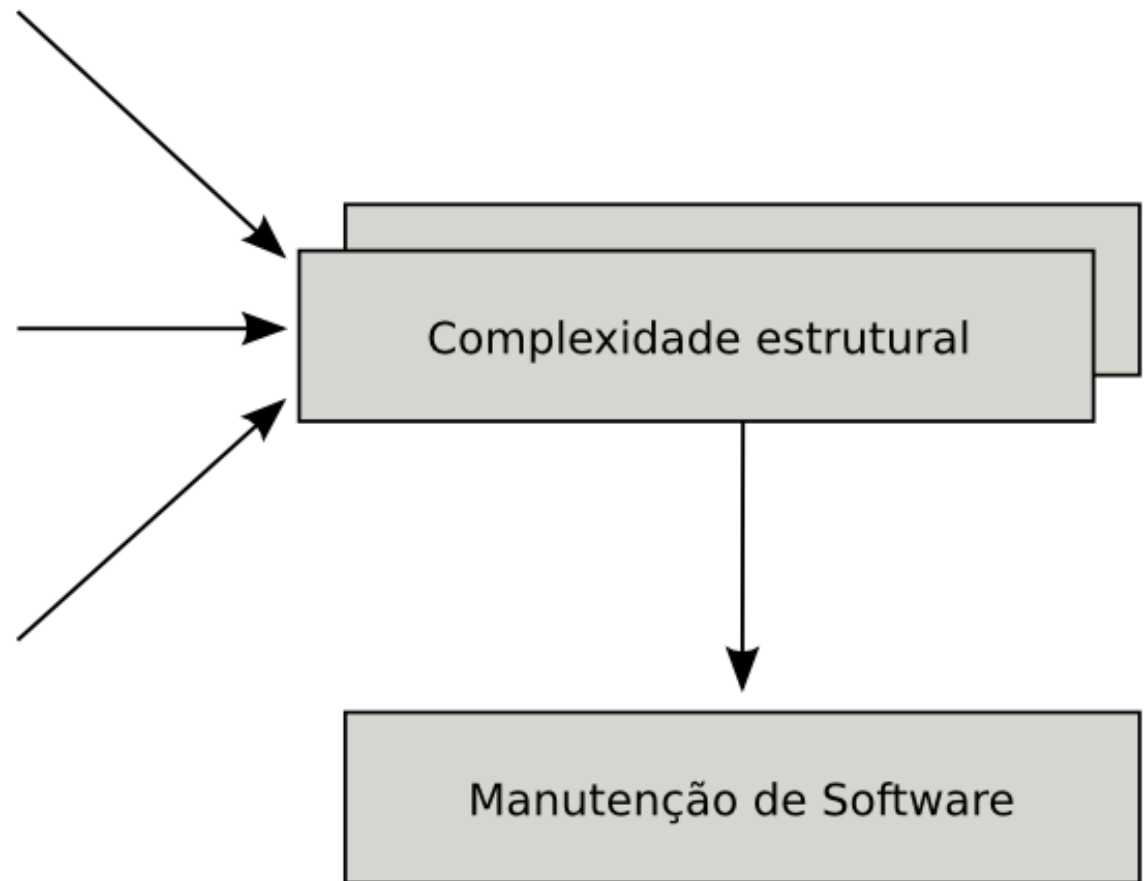
(em projetos de Software Livre)

[Meirelles et al. 2010]

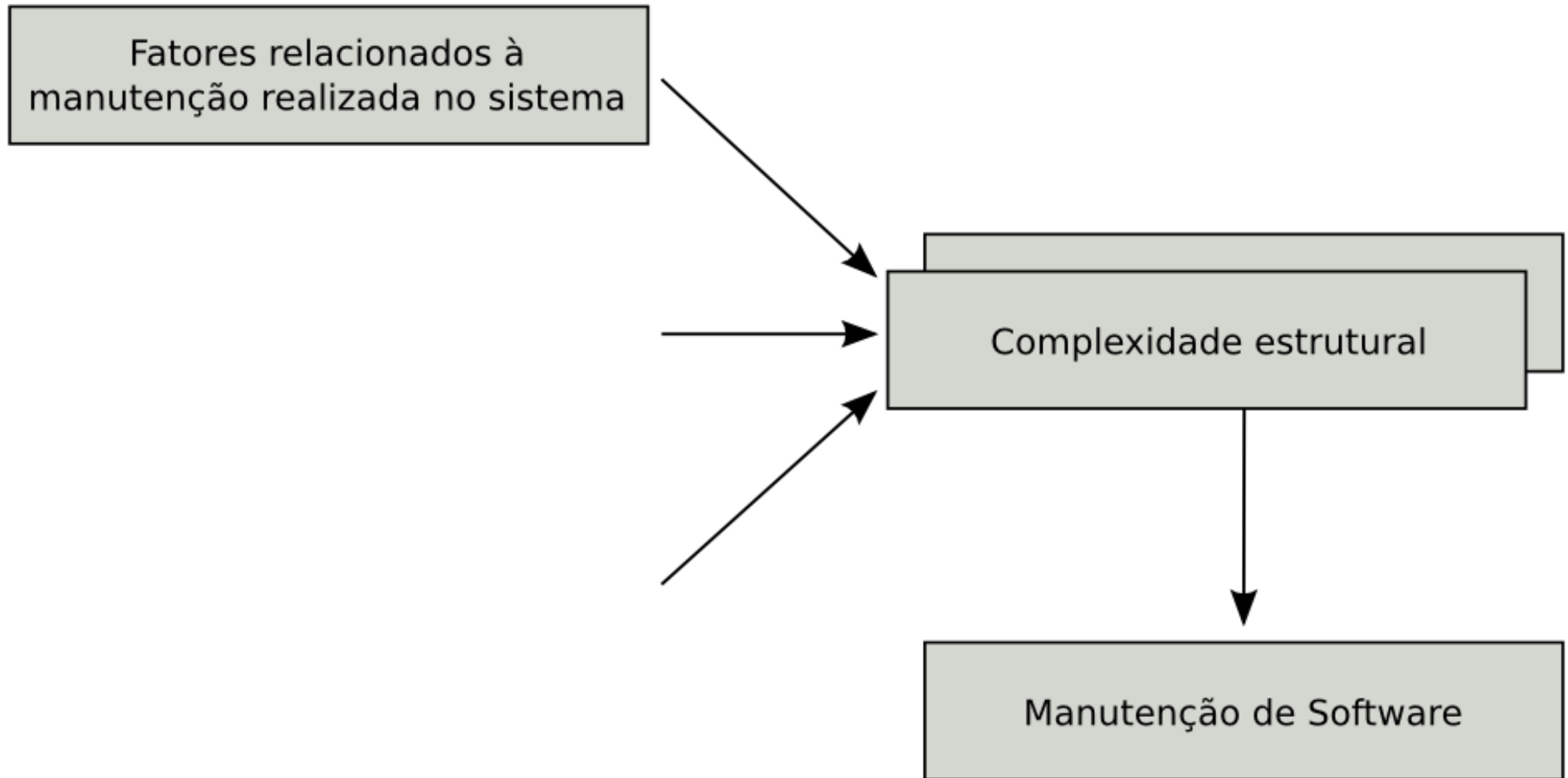
Motivação



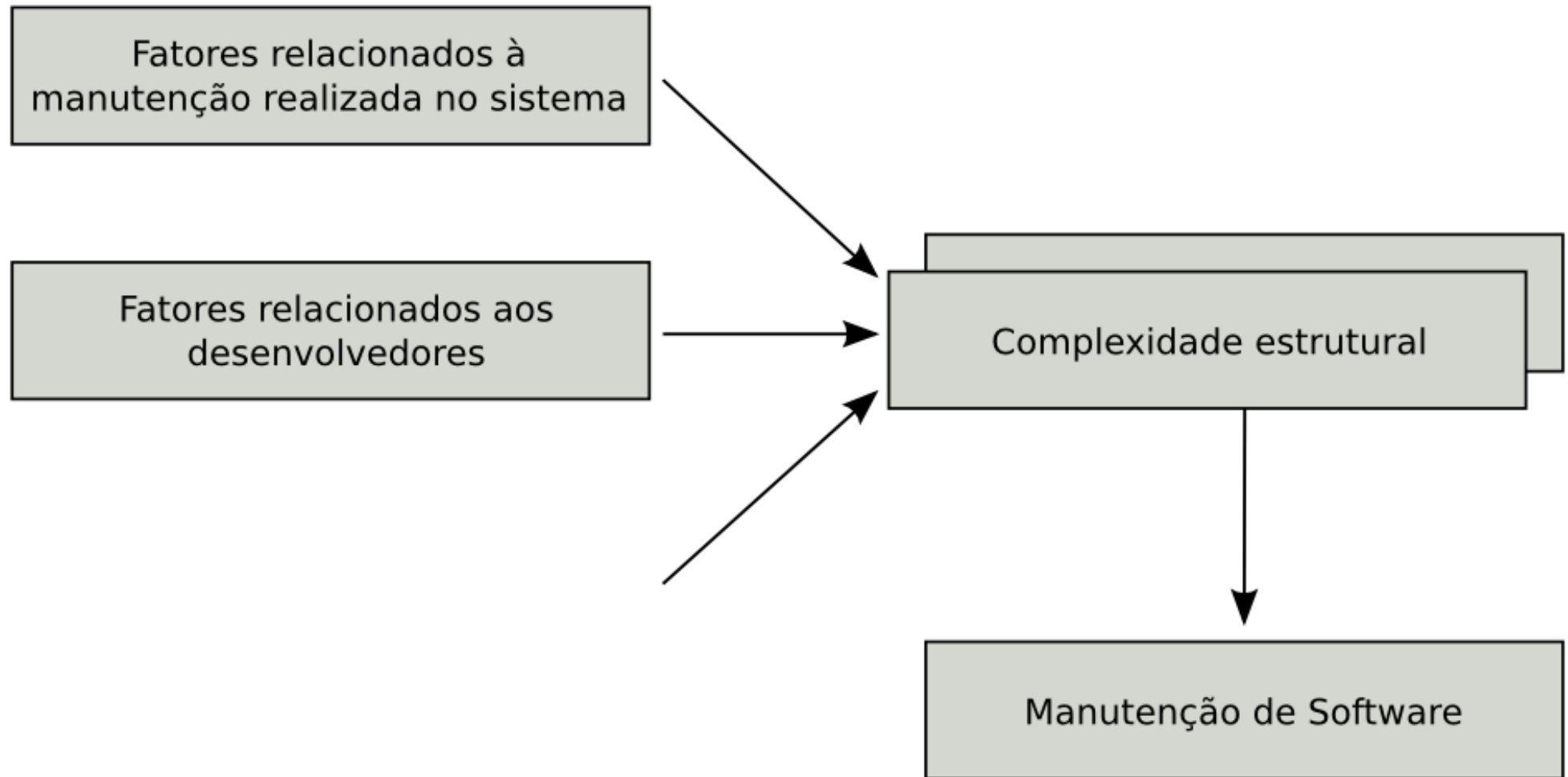
Motivação



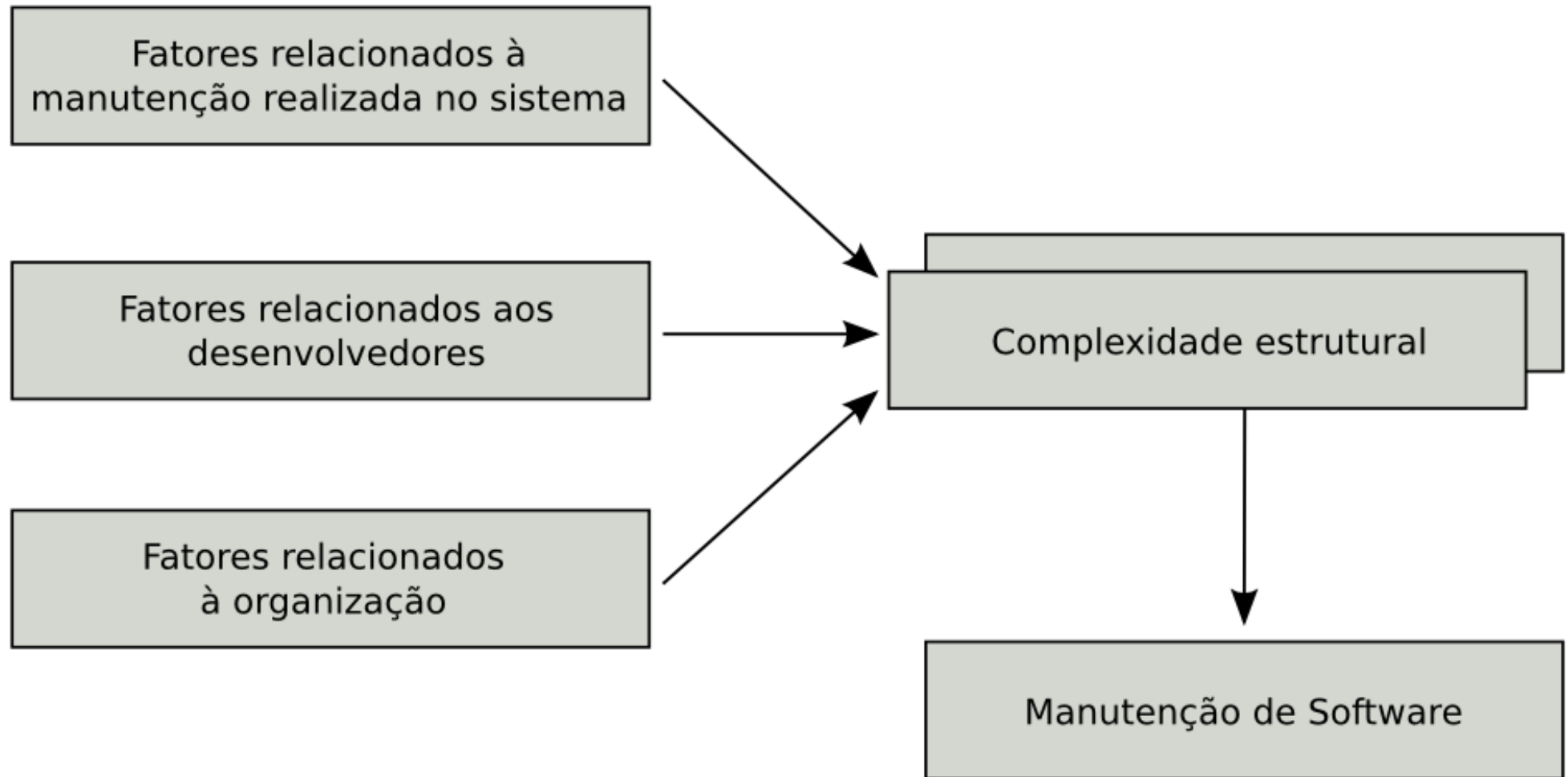
Motivação



Motivação



Motivação



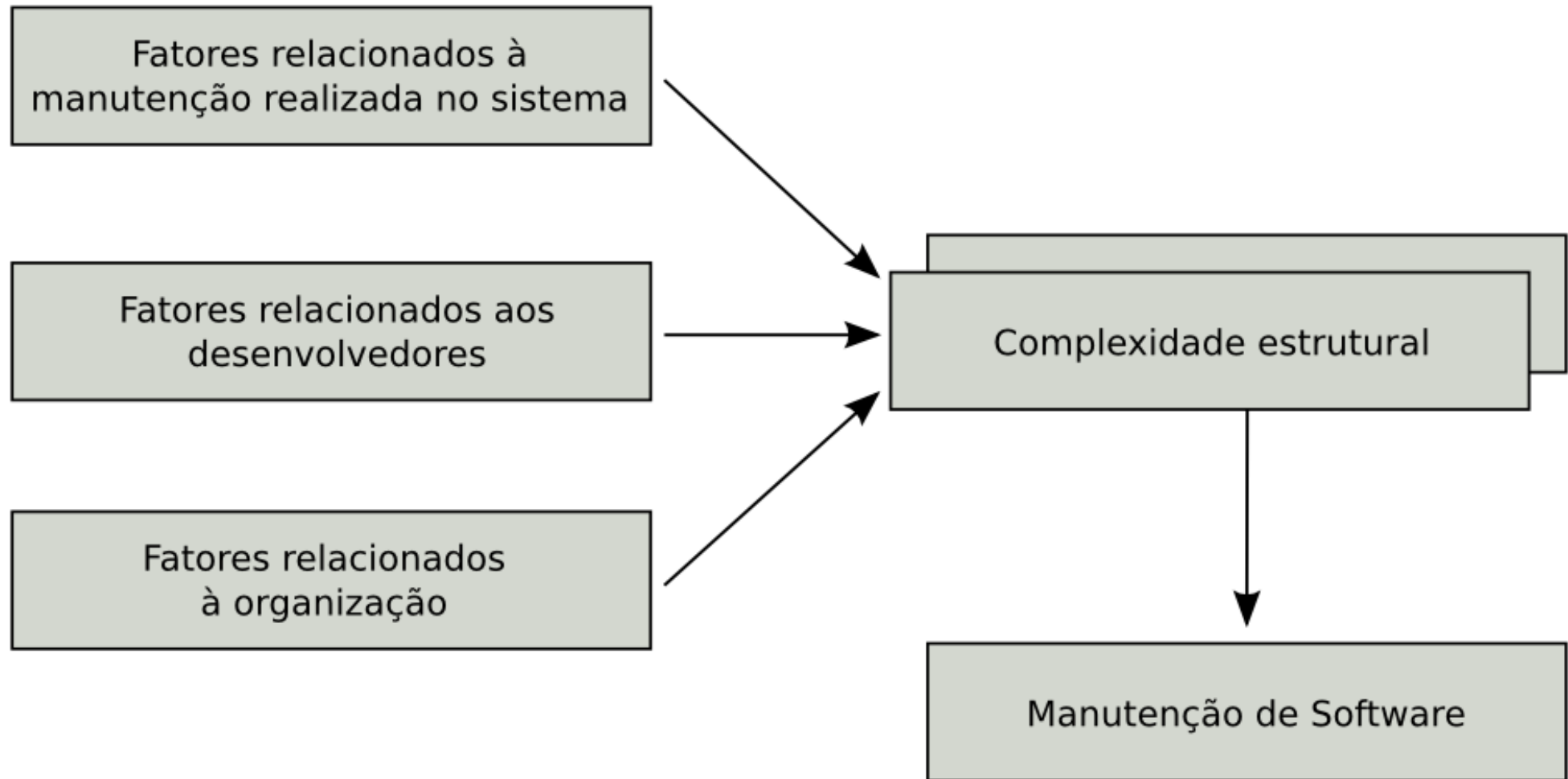
Objetivo geral

Melhorar entendimento sobre a complexidade estrutural e os fatores que a influenciam ao longo do tempo.

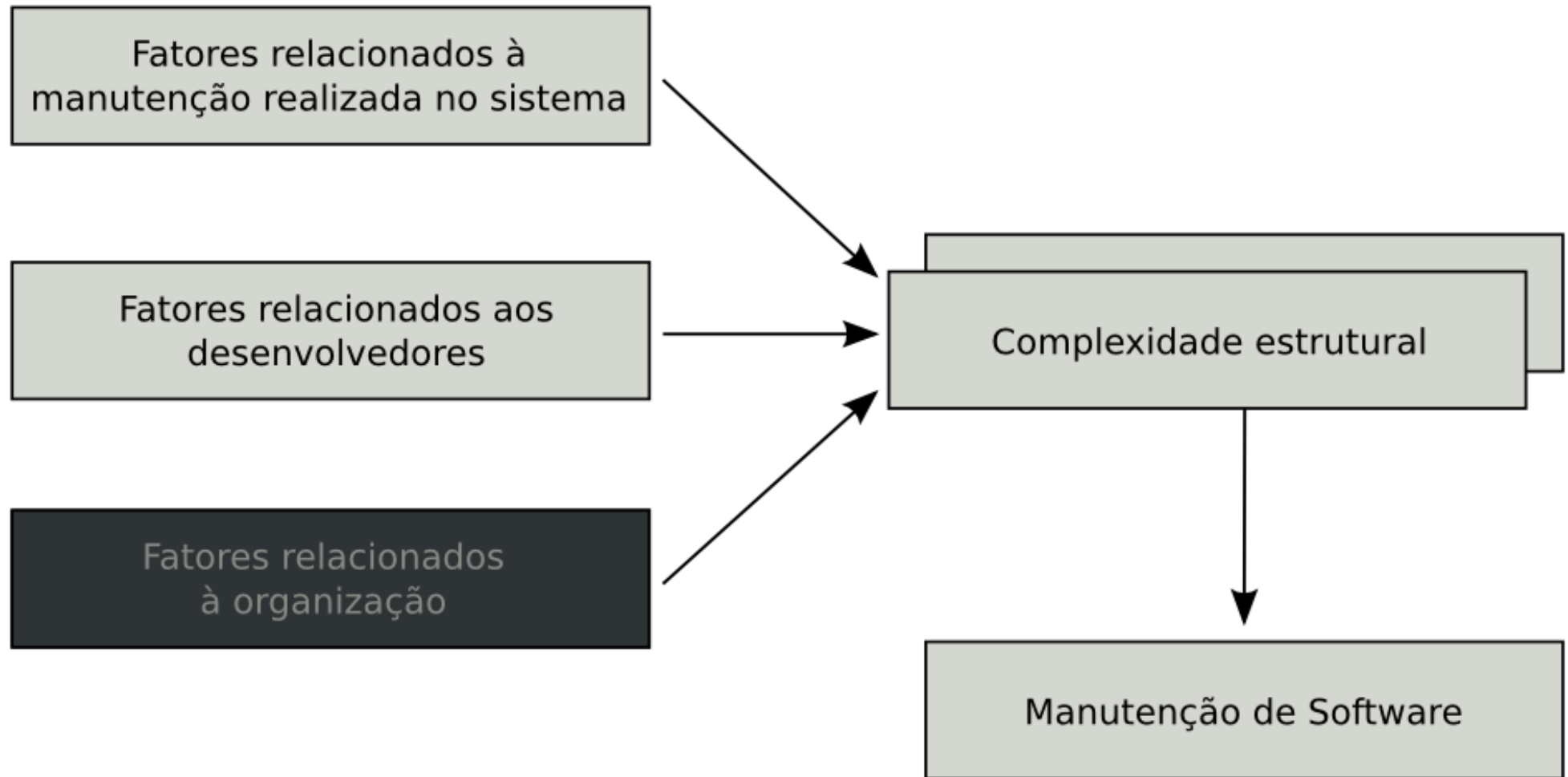
Objetivos específicos

- Identificar *fatores* que influenciam a complexidade estrutural
- Formular uma *teoria* que identifique
 - contribuição dos fatores identificados para à variação da complexidade estrutural
 - efeitos da complexidade estrutural
- Testar a validade das proposições desta teoria

Escopo



Escopo



Metodologia

- Engenharia de Software Experimental
- Mineração de repositórios de código fonte de projetos de software livre
- 4 estudos experimentais
- Análise quantitativa

Metodologia (2)

- 13 projetos
- 1 a 164 desenvolvedores
- 1,5 a 5 anos de existência
- 19102 commits analisados no total

Resultados

- Proposta de teoria para complexidade estrutural em sistemas de software
- Série de resultados experimentais sobre complexidade estrutural
- Ferramenta de análise de código-fonte

Complexidade Estrutural

Fatores

Teoria

Estudos experimentais

Discussão

Conclusões

Complexidade em sistemas de software

Propriedade relacionada a:

- Organização interna dos módulos
- Relacionamento entre módulos

3 medidas de complexidade de software
foram identificadas na literatura

Critérios para avaliação de medidas para complexidade de sistemas de software

C1: a medida leva em consideração tanto a complexidade interna a um módulo quanto a complexidade do relacionamento entre módulos?

C2: a medida representa uma dimensão independente do tamanho?

C3: a medida é independente de linguagem de programação ou paradigma de desenvolvimento?

Avaliação das medidas de complexidade

		<div>Complexidade intra-módulo</div> <div>complexidade inter-módulos</div> <div>Independência do tamanho</div> <div>Independência do paradigma</div>		
Proposta		C1	C2	C3
1)	CC [McCabe 1976]	Não	Não	Sim
Graus de violação e tamanho				
2)	[Sangwan, Vercellone-Smith e Laplante 2008]	Sim	?	Não
3)	SC [Darcy et al. 2005]	Sim	Sim	Sim

Complexidade Estrutural

- Combinação de acoplamento e coesão
[Darcy et al 2005]
- Associada a efeitos negativos em sistemas de software.
[Darcy et al 2005, Meirelles et al 2010]

Complexidade estrutural: definição

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

$SC(p)$: complexidade estrutural do sistema p

$A(m)$: acoplamento do módulo m

$FC(m)$: falta de coesão do módulo m

$M(p)$: conjunto de módulos do sistema p

Comportamento evolucionário da complexidade estrutural

- Tendência de crescimento [Stewart, Darcy e Daniel 2006, Darcy, Daniel e Stewart 2010, Terceiro e Chavez 2009,]
- Lei da Complexidade Crescente [Lehman e Belady 1985, Lehman et al. 1997]
- Aumento no tamanho não está necessariamente associado a aumento na complexidade [Darcy, Daniel e Stewart 2010]

Complexidade Estrutural

Fatores

Teoria

Estudos experimentais

Discussão

Conclusões

Fatores

Relacionados aos desenvolvedores

Relacionados à manutenção

Experiência dos desenvolvedores

- Diferentes formas de medir
- Número de modificações realizadas no sistema
- Usada para
 - Identificar experts [Mockus e Herbsleb 2002]
 - Preservar equipes de manutenção
[Banker, Datar e Kemerer 1991]
 - Predizer defeitos
[Mockus e Weiss 2000][Matsumoto et al. 2010]

Grau de autoria

[Fritz et al. 2010]

Componentes:

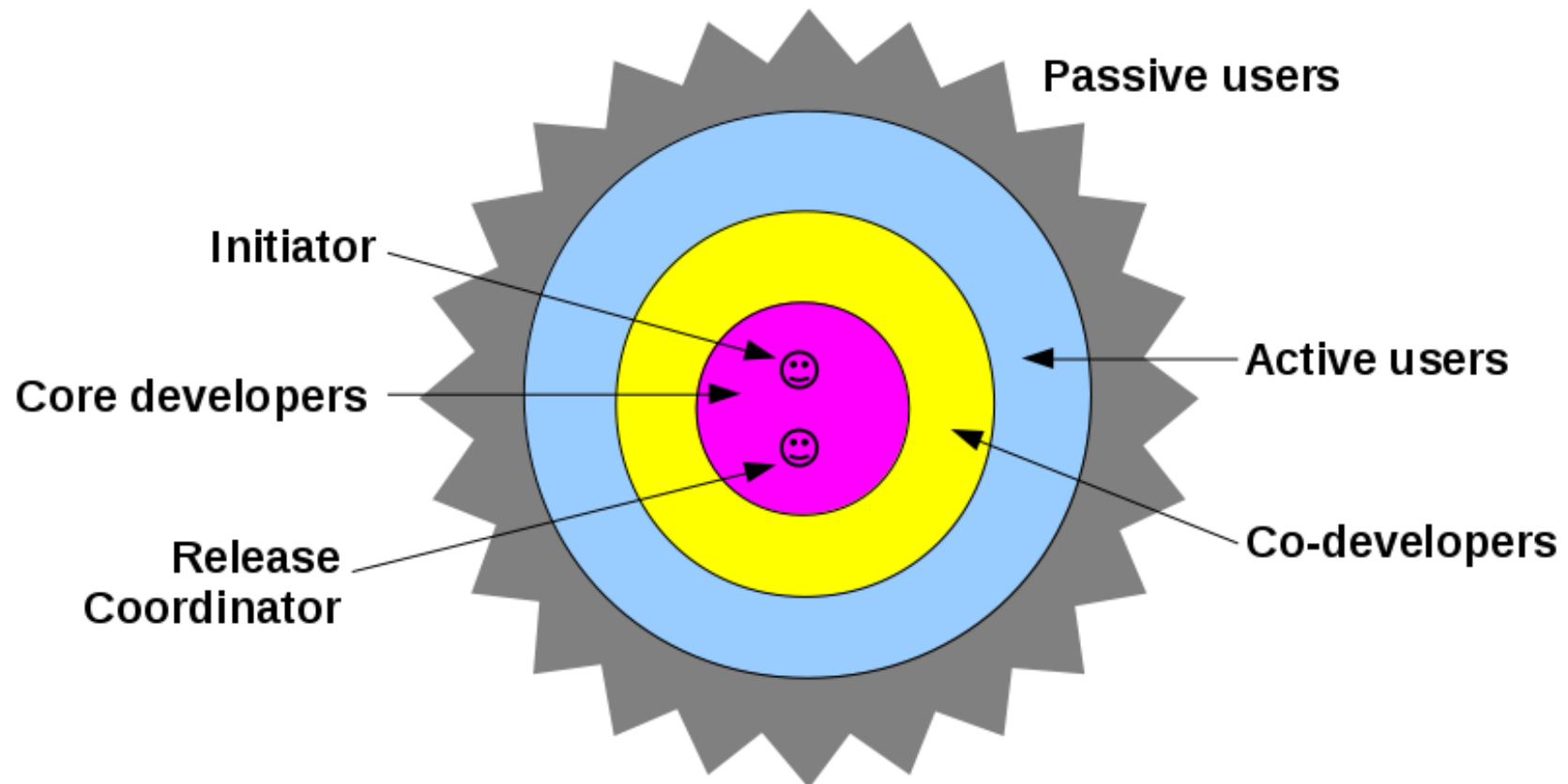
- FA – *first authorship* (primeira autoria)
- DL – *deliveries* (alterações)
- AC – *acceptances* (alterações por outros)

Usado para

- Identificar experts [Mockus, Fielding e Herbsleb 2002]
- Identificar interesse/conhecimento [Fritz et al. 2010]

Nível de participação em projetos de Software Livre.

O modelo “cebola” [Crowston e Howison 2005]



Fatores

Relacionados aos desenvolvedores

Relacionados à manutenção

Variação de tamanho

- Tamanho como medida de complexidade
- Tamanho dificulta mudanças
[Parnas 1994]
- "Crescimento Contínuo" e "Complexidade Crescente" [Lehman et al. 1997]
- Quando um sistema cresce, ele necessariamente fica mais complexo?

Difusão de mudança

- Medida do espalhamento da uma mudança
- Maior difusão de mudança pode indicar uma mudança incompatível com o *design* existente [Mockus e Weiss 2000, Hassan 2009]
- Maior probabilidade de introdução de defeitos [Mockus e Weiss 2000, Hassan 2009]

Complexidade Estrutural
Fatores

Teoria

Estudos experimentais

Discussão

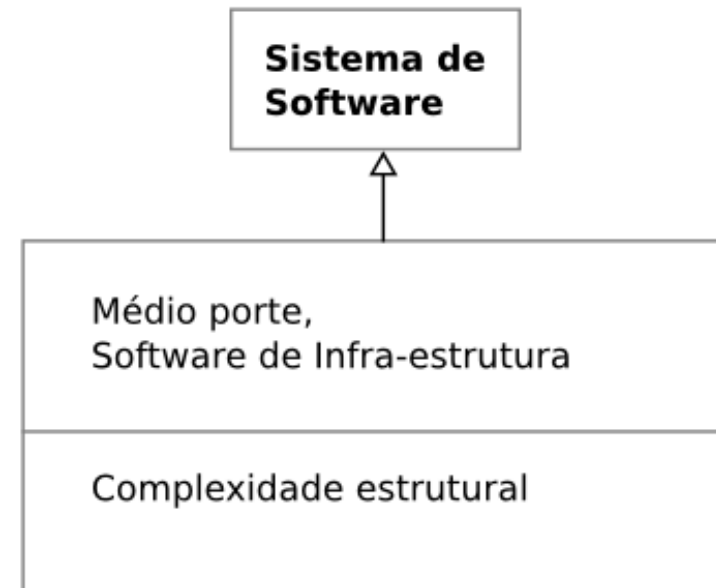
Conclusões

Teorias em Engenharia de Software

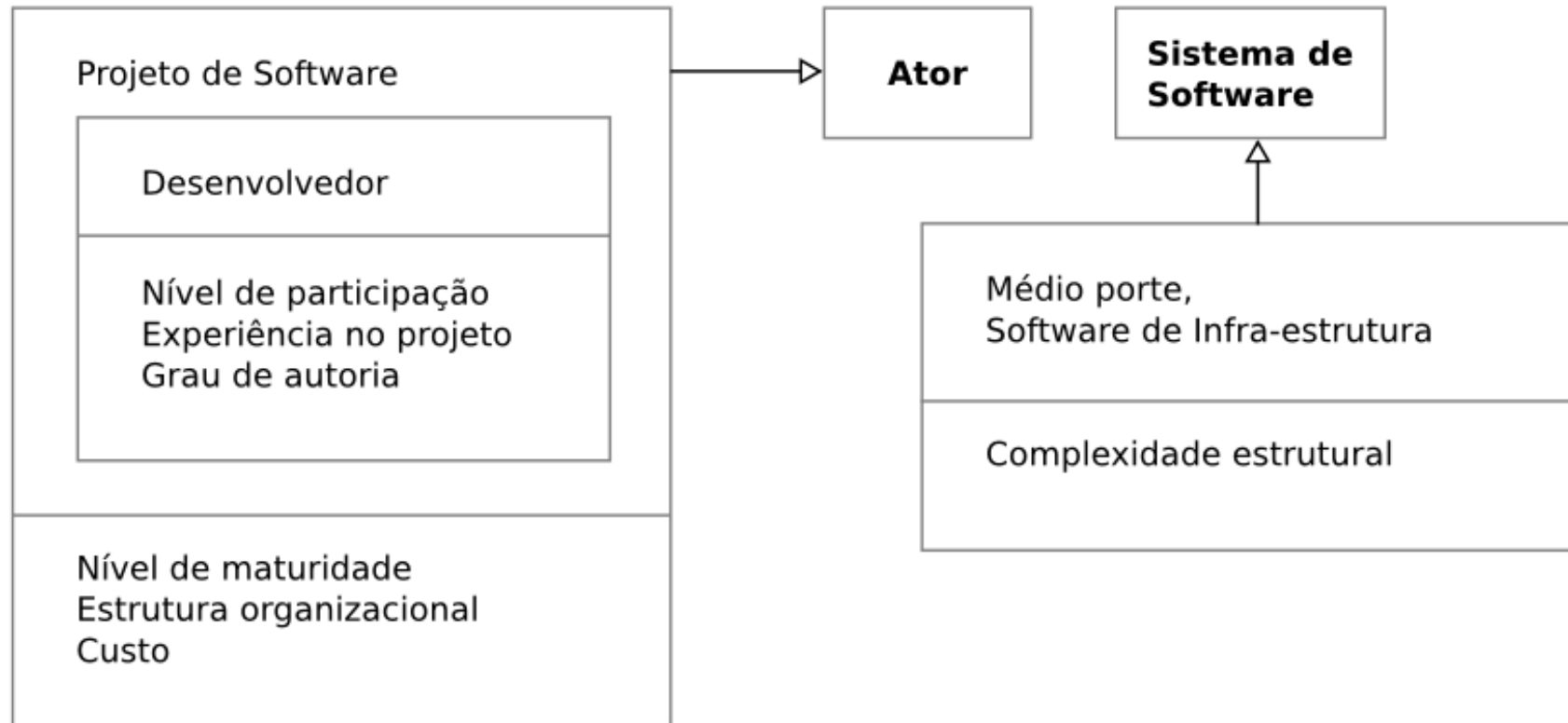
[Sjøberg et al. 2008]

- Permitem combinar o conhecimento obtido em múltiplos estudos individuais
- Relacionam *construtos* através de *proposições*

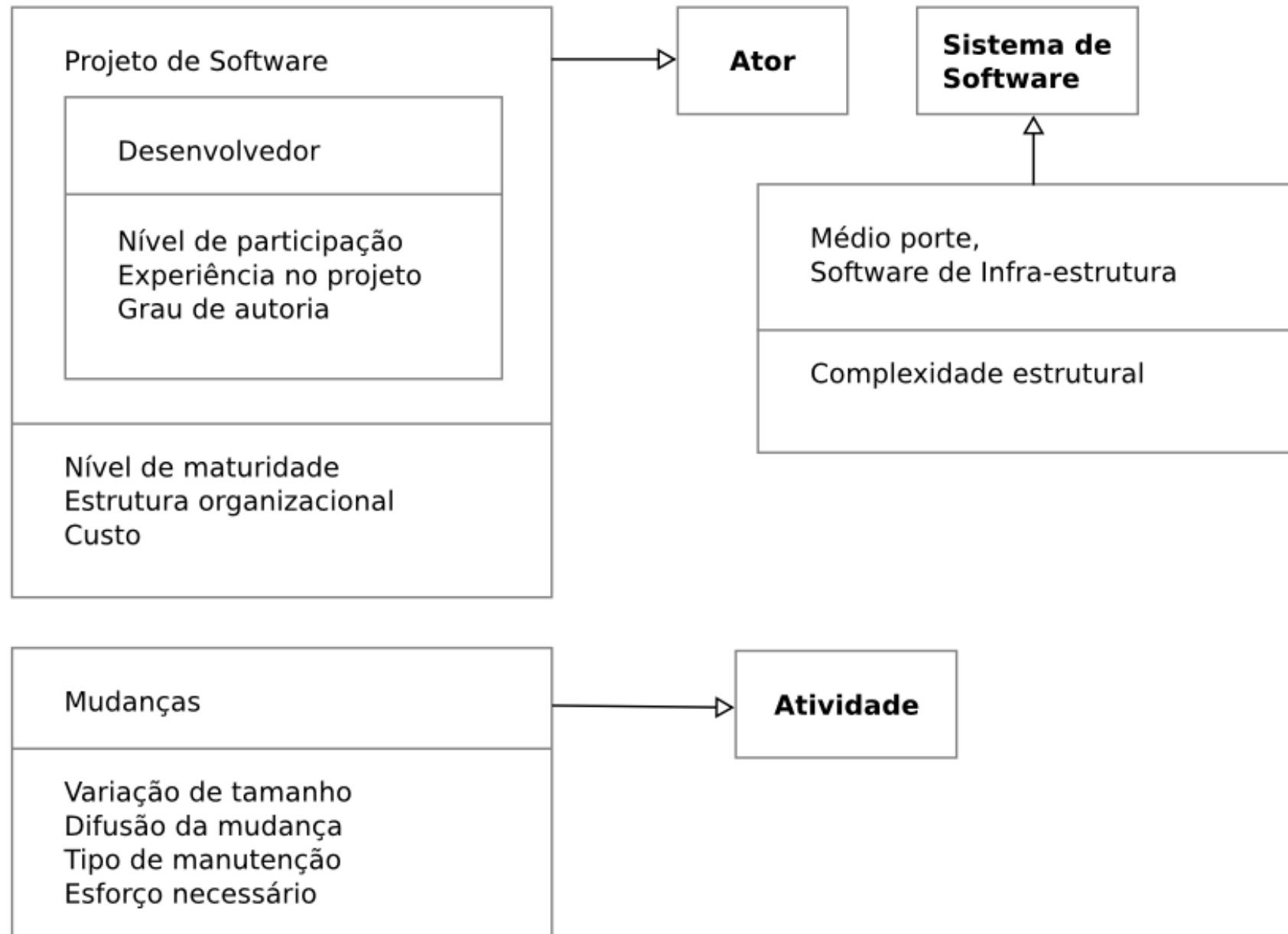
Teoria – visão geral



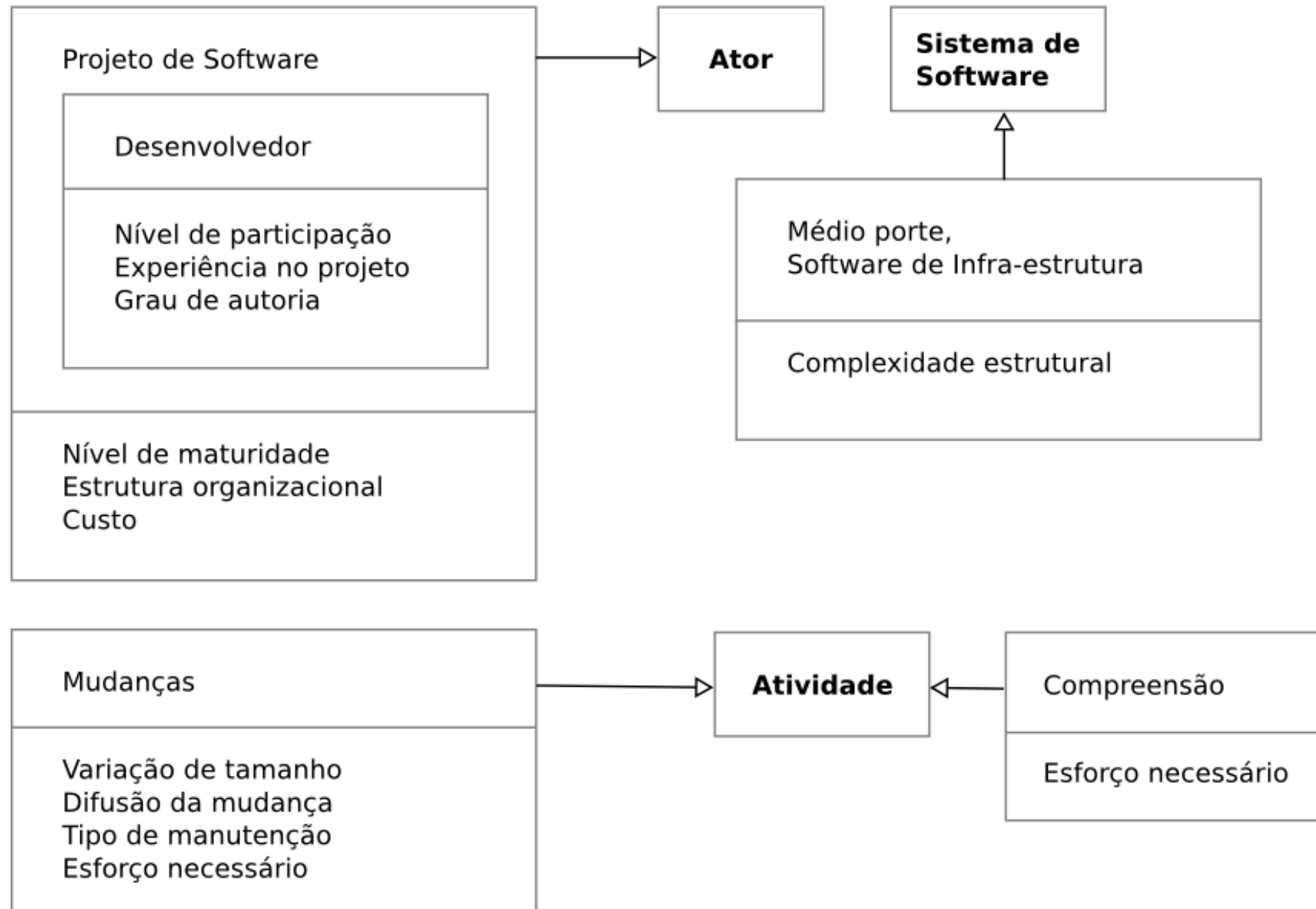
Teoria – visão geral



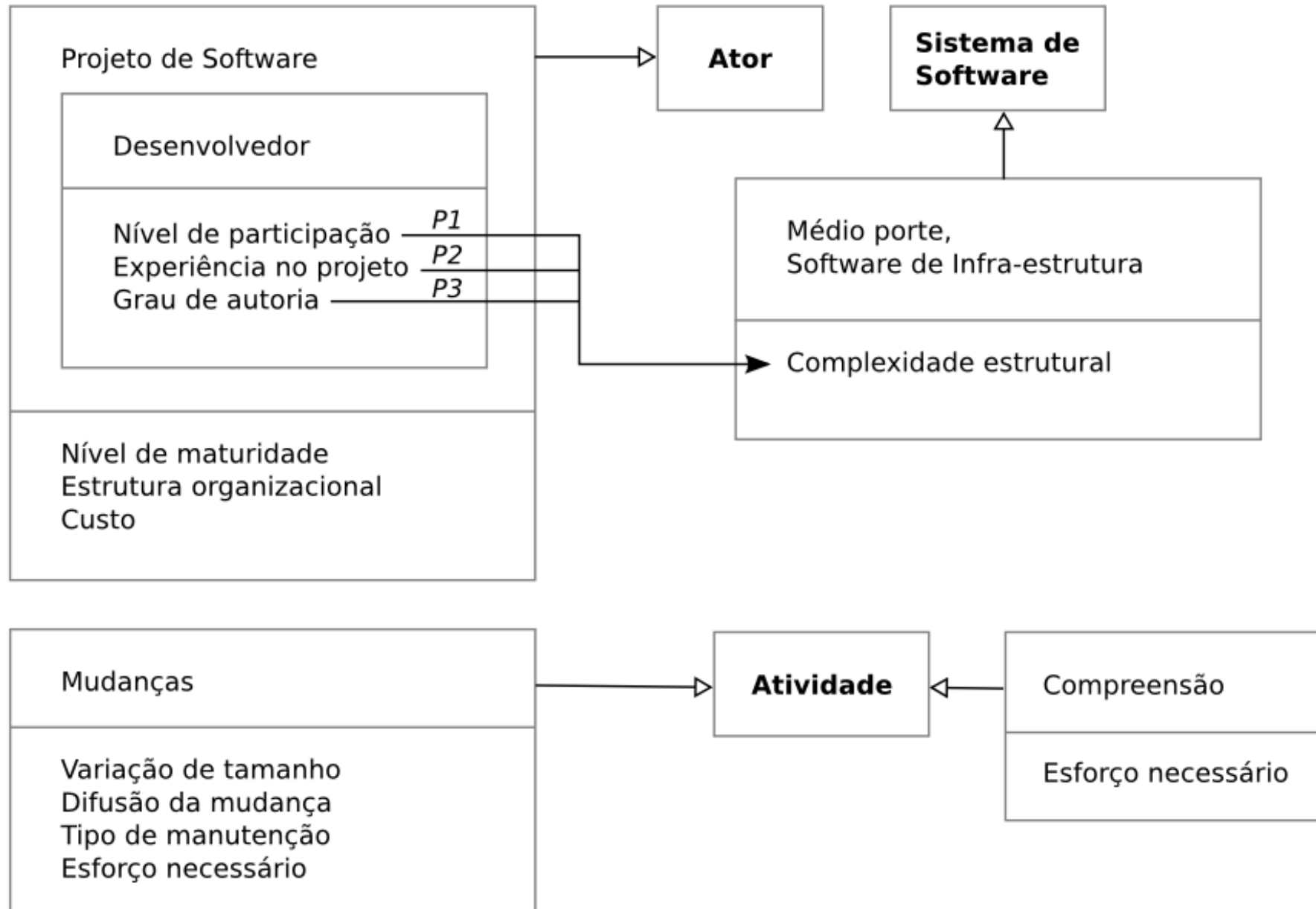
Teoria – visão geral



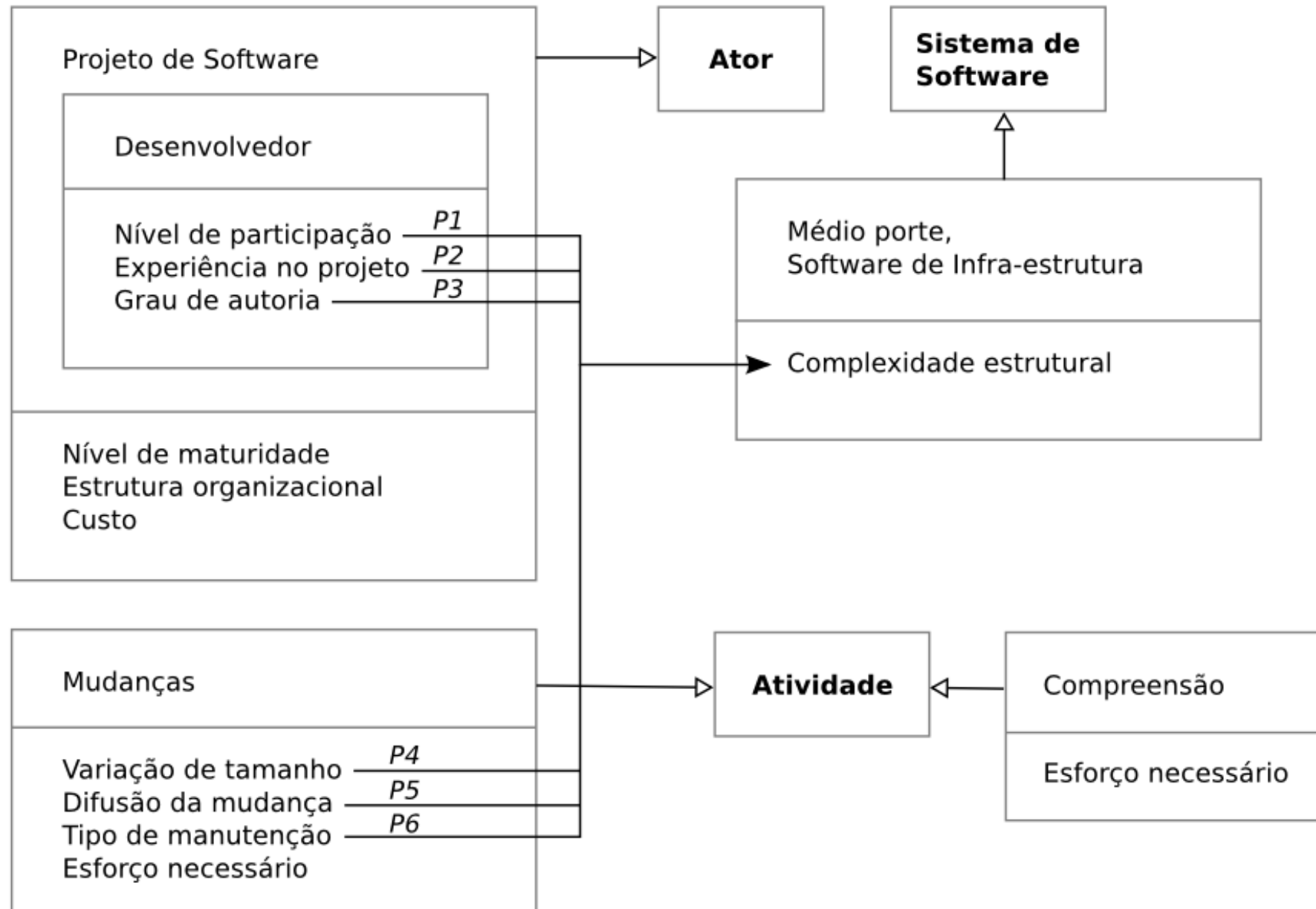
Teoria – visão geral



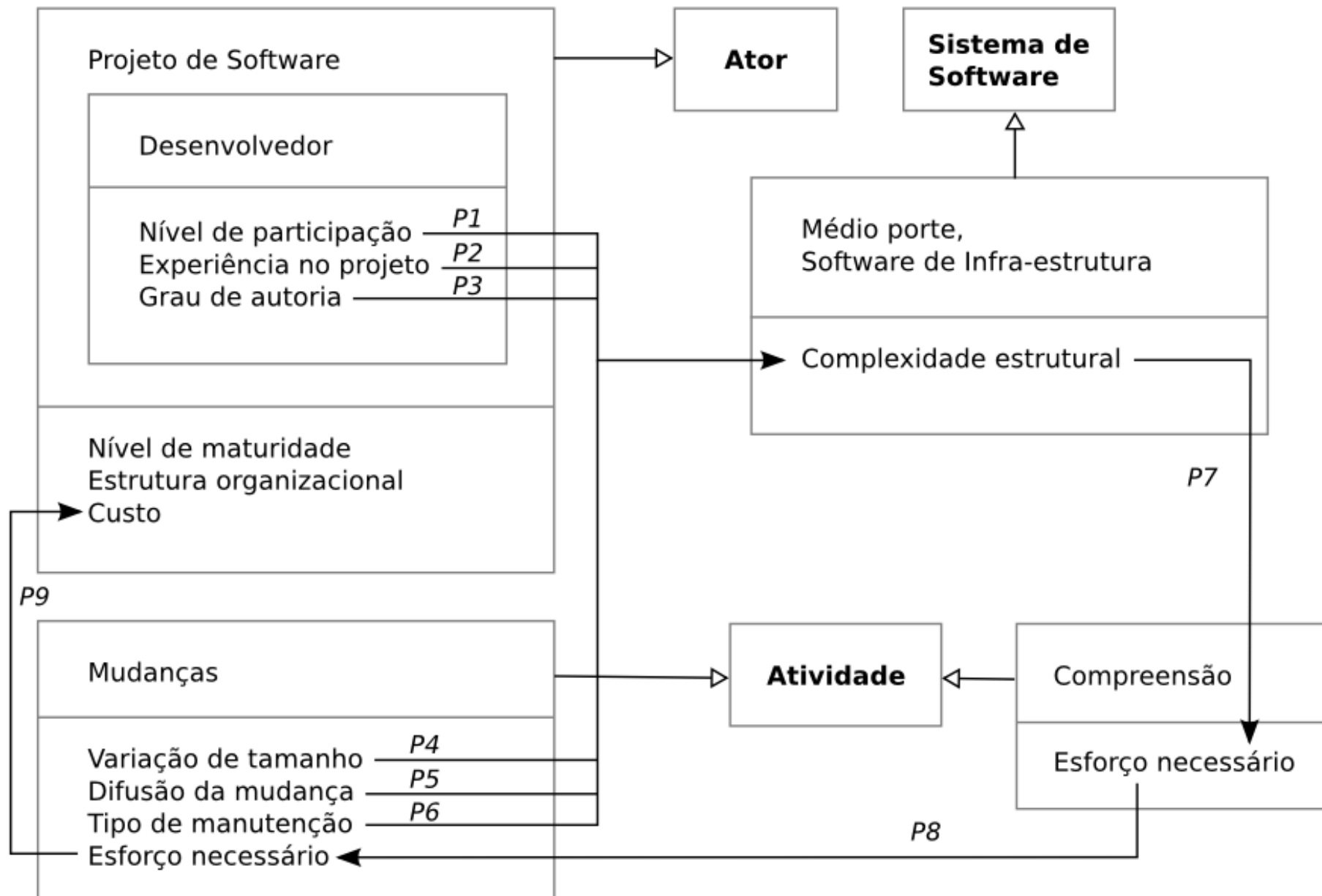
Teoria – visão geral



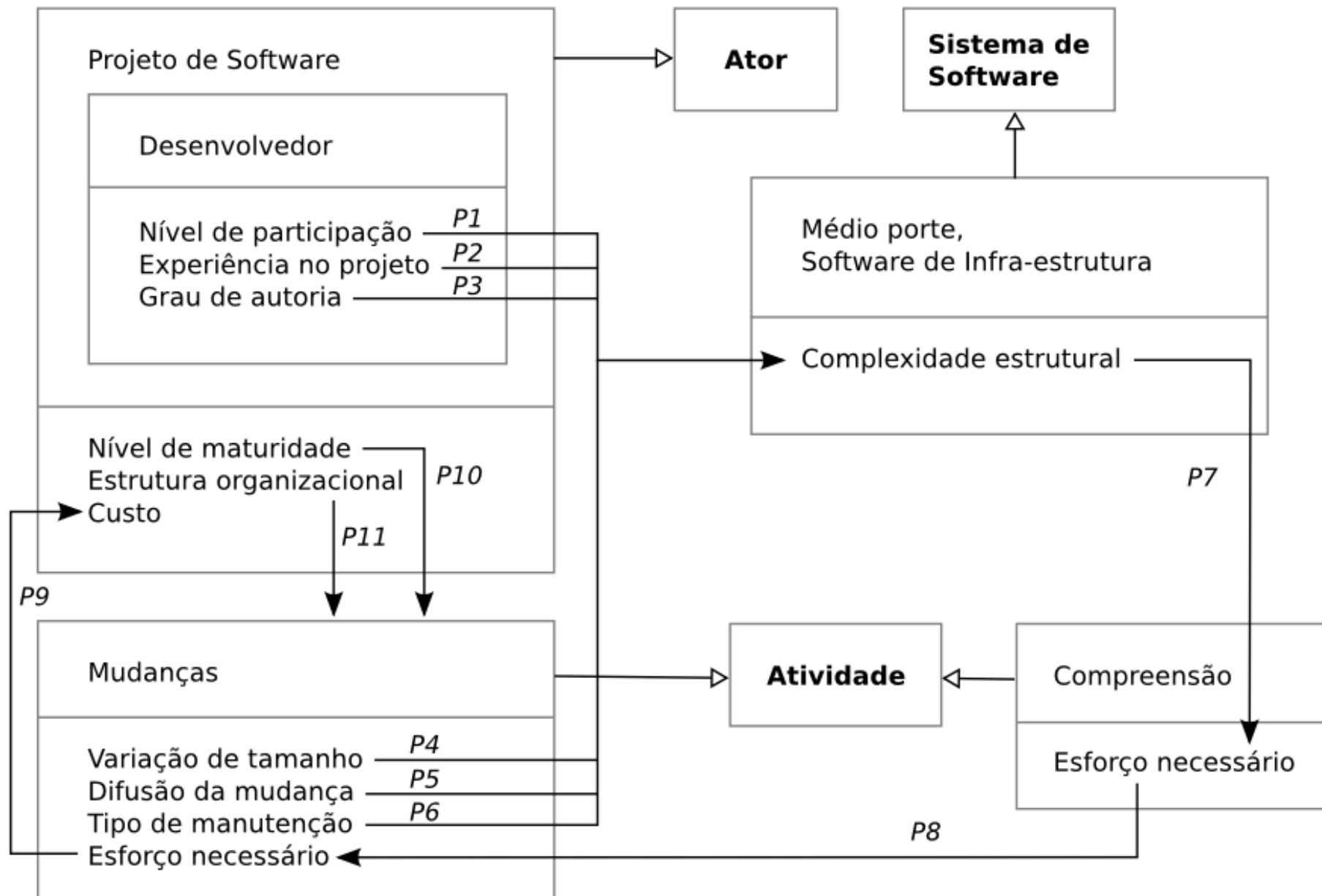
Teoria – visão geral



Teoria – visão geral



Teoria – visão geral



Teoria: proposições

Proposição	Situação
P1	Explorada (Estudo 2)
P2	Explorada (Estudos 3 e 4)
P3	Explorada (Estudo 4)
P4	Explorada (Estudos 3 e 4)
P5	Explorada (Estudos 3 e 4)
P6	Não explorada
P7	Validada por outros pesquisadores
P8	Axioma
P9	Axioma
P10	Não explorada
P11	Não explorada

Escopo da teoria

- Projetos de Software Livre
- C/C++/Java
- Software de infraestrutura
- Até 35000 LOC

Complexidade Estrutural

Fatores

Teoria

Estudos experimentais

Discussão

Conclusões

Estudos experimentais

Considerações gerais

Estudo 1

Estudo 2

Estudo 3

Estudo 4

Complexidade Estrutural: cálculo

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

A : CBO [Chidamber e Kemerer 1994]

FC : LCOM₄ [Hitz e Montazeri 1995]

nalizo

- Análise e visualização de código-fonte
- Multi-linguagem: C/C++/Java
- Software livre
- 6 colaboradores

Reprodutibilidade

- <http://deb.li/tt>
- Ferramentas livres
- Projetos de Software Livre
- Metodologia documentada

Estudos experimentais

Considerações gerais

Estudo 1

Estudo 2

Estudo 3

Estudo 4

Estudo 1: objetivos

- Experimentar uma abordagem para estudo da evolução da complexidade estrutural
 - Prova de conceito da analizo
 - Estender resultados existentes sobre evolução da complexidade estrutural
- [Stewart, Darcy e Daniel 2006]

Estudo 1: hipóteses

H_1 : o tamanho do projeto aumenta com a passagem do tempo

H_2 : a complexidade estrutural do projeto aumenta com a passagem do tempo.

Estudo 1: projeto experimental

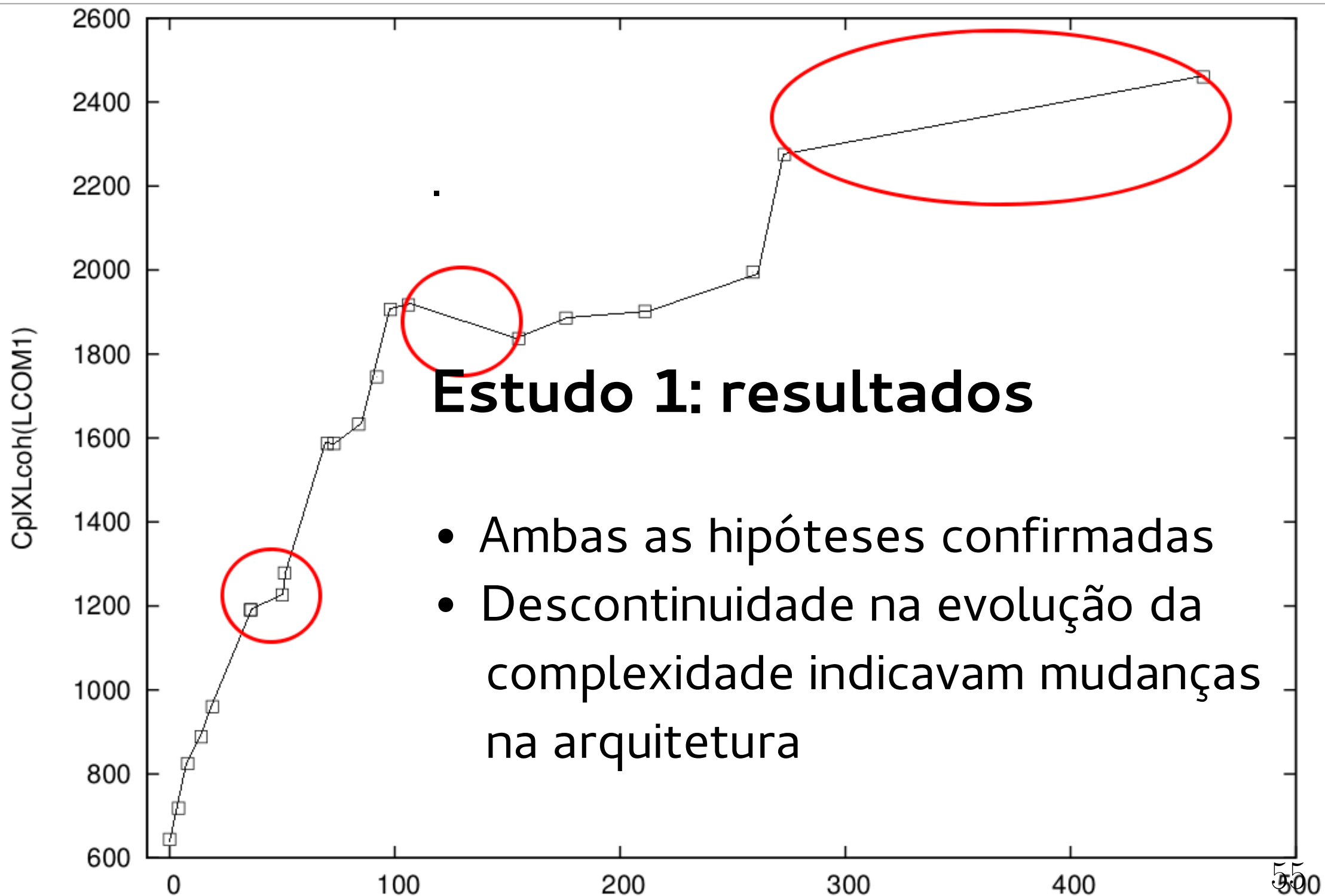
Amostra: Projeto Ristretto (pequeno, C)

Unidade de análise: *releases*

(21 *releases*, 15 meses de desenvolvimento)

Variáveis:

- *RD* – dia de lançamento
- *SLOC* – tamanho do sistema
- *SC* – complexidade estrutural



Estudo 1: conclusões

- Abordagem promissora para estudos experimentais
- Automação da execução da análise viabiliza sua utilização no apoio ao desenvolvimento
- Mudanças na tendência evolutiva da complexidade estrutural podem indicar mudanças na arquitetura do sistema

Estudos experimentais

Considerações gerais

Estudo 1

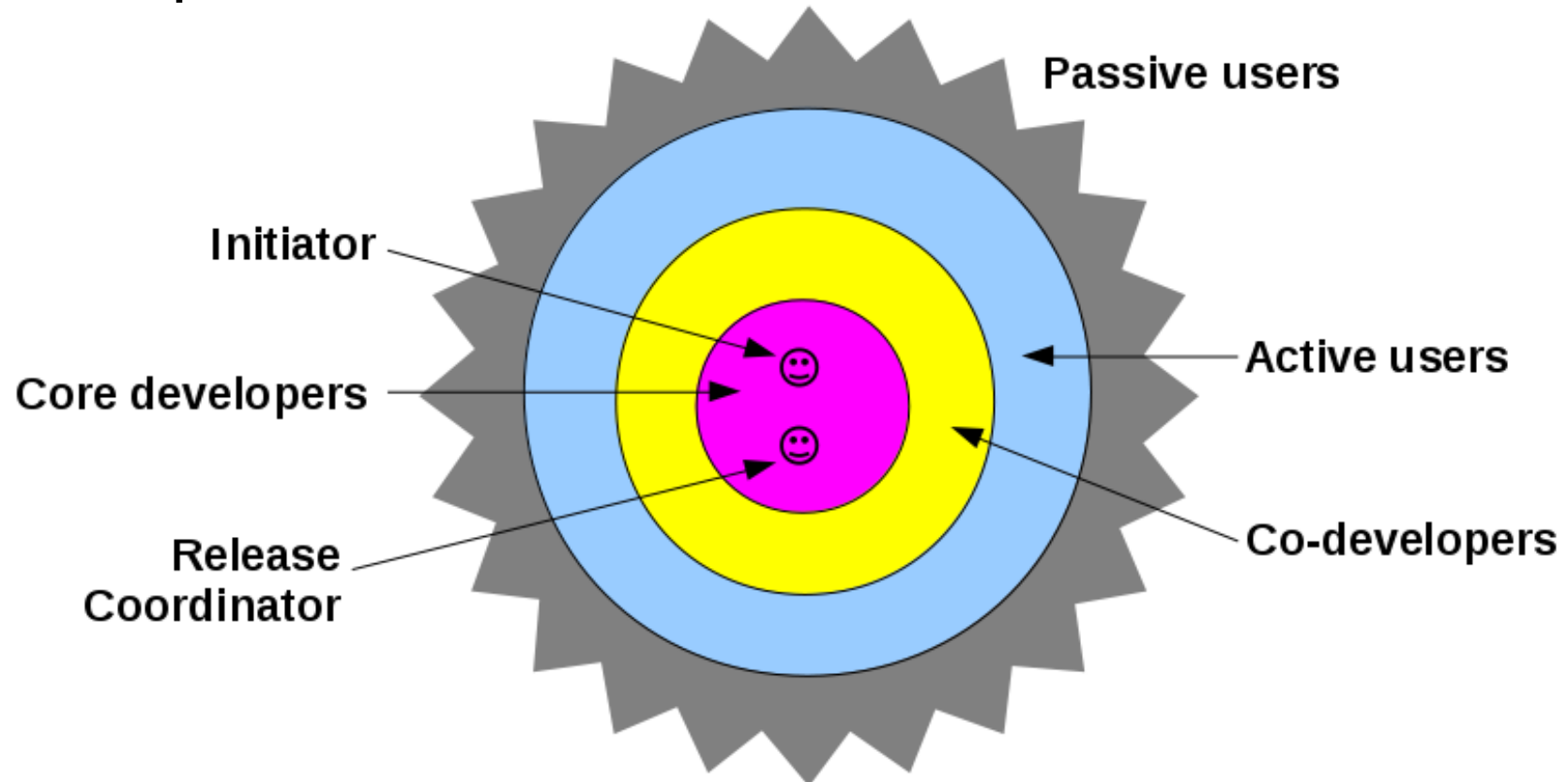
Estudo 2

Estudo 3

Estudo 4

Estudo 2: objetivo

Identificar a influência do nível de participação dos desenvolvedores sobre a evolução da complexidade estrutural



Estudo 2: hipóteses

H_1 : mudanças feitas por desenvolvedores centrais *introduzem menos complexidade estrutural* que as feitas por desenvolvedores periféricos

H_2 : mudanças feitas por desenvolvedores centrais *retiram mais complexidade estrutural* que as feitas por desenvolvedores periféricos

Estudo 2: projeto experimental

Amostra: 7 sistemas, servidores web (C)

Unidade de análise: commits

Variáveis:

- ΔSC : variação de complexidade
- L : nível de participação

Estudo 2: resultados

✓ H_1 : mudanças feitas por desenvolvedores centrais *introduzem menos complexidade estrutural* que as feitas por desenvolvedores periféricos

✓ H_2 : mudanças feitas por desenvolvedores centrais *retiram mais complexidade estrutural* que as feitas por desenvolvedores periféricos

Estudo 2: conclusões

- Diferença entre centro e periferia
- Importância da equipe central

Estudos experimentais

Considerações gerais

Estudo 1

Estudo 2

Estudo 3

Estudo 4

Estudo 3: objetivos

Estudar a influência dos seguintes fatores sobre a complexidade estrutural:

- Experiência no projeto
- Difusão da mudança
- Variação de tamanho

Estudo 3: hipóteses

H_1 : ↑ Experiência, ↓ aumento na complexidade

H_2 : ↑ Variação de tamanho, ↑ aumento na complexidade

H_3 : ↑ Difusão de mudança, ↑ aumento na complexidade

H_4 : ↑ Experiência, ↑ redução na complexidade

H_5 : ↑ Variação de tamanho, ↓ redução na complexidade

H_6 : ↑ Difusão de mudança, ↑ redução na complexidade

Estudo 3: projeto experimental

Amostra: 5 projetos de médio porte. C/C++/Java

Unidade de análise: commits

Variáveis:

- n – experiência. Número de mudanças anteriores
- ΔLOC – variação no tamanho
- CF – difusão de mudança (*changed files*)
- ΔSC – variação na complexidade estrutural
- ΔSC_i – aumento na complexidade estrutural
- ΔSC_d – redução na complexidade estrutural

Estudo 3: resultados

$$\Delta SC_i = \alpha_0 + \alpha_1 n + \alpha_2 \Delta LOC + \alpha_3 CF$$

$$\Delta SC_d = \alpha_0 + \alpha_1 n + \alpha_2 \Delta LOC + \alpha_3 CF$$

Estudo 3: resultados

TABLE III
REGRESSION MODELS FOR COMMITS THAT INCREASE STRUCTURAL COMPLEXITY

Project	#Commits	Intercept	$n (H_1)$	$\Delta LOC (H_2)$	CF (H_3)	Adj. R^2
clojure	287	0.031637	0.000108	0.000147 ***	0.014175 ***	0.20
node	133	0.340228 ***	-0.000479 **	0.000369	0.025465 *	0.10
redis	72	-1.848899	0.004697	0.00244	0.236758 ***	0.31
voldemort	217	0.026552	-5e-06	0.000195 **	0.002326 ***	0.22
zeromq2	71	0.224476 **	-0.000399	0.000746 **	0.004965	0.21

***: $p < 0.001$; **: $p < 0.01$; *: $p < 0.05$

TABLE IV
REGRESSION MODELS FOR COMMITS THAT DECREASE STRUCTURAL COMPLEXITY

Project	#Commits	Intercept	$n (H_4)$	$\Delta LOC (H_5)$	CF (H_6)	Adj. R^2
clojure	235	-0.064833	0.000422 **	0.000204	0.011493	0.04
node	107	0.313235 *	-0.00038	-0.000308	0.036385 **	0.10
redis	45	-0.359334	0.000952	0.000443	0.163235 **	0.18
voldemort	177	0.029518 *	2.3e-05	0.00013 **	0.001125	0.07
zeromq2	57	0.10271	-0.000372	-0.001258 ***	0.021416 ***	0.88

***: $p < 0.001$; **: $p < 0.01$; *: $p < 0.05$

Estudo 3: resultados

TABLE III
REGRESSION MODELS FOR COMMITS THAT INCREASE STRUCTURAL COMPLEXITY

Project	#Commits	Intercept	$n (H_1)$	$\Delta LOC (H_2)$	CF (H_3)	Adj. R^2
clojure	287	0.031637	0.000108	0.000147 ***	0.014175 ***	0.20
node	133	0.340228 ***	-0.000479 **	0.000369	0.025465 *	0.10
redis	72	-1.848899	0.004697	0.00244	0.236758 ***	0.31
voldemort	217	0.026552	-5e-06	0.000195 **	0.002326 ***	0.22
zeromq2	71	0.224476 **	-0.000399	0.000746 **	0.004965	0.21

***: $p < 0.001$; **: $p < 0.01$; *: $p < 0.05$

TABLE IV
REGRESSION MODELS FOR COMMITS THAT DECREASE STRUCTURAL COMPLEXITY

Project	#Commits	Intercept	$n (H_4)$	$\Delta LOC (H_5)$	CF (H_6)	Adj. R^2
clojure	235	-0.064833	0.000422 **	0.000204	0.011493	0.04
node	107	0.313235 *	-0.00038	-0.000308	0.036385 **	0.10
redis	45	-0.359334	0.000952	0.000443	0.163235 **	0.18
voldemort	177	0.029518 *	2.3e-05	0.00013 **	0.001125	0.07
zeromq2	57	0.10271	-0.000372	-0.001258 ***	0.021416 ***	0.88

***: $p < 0.001$; **: $p < 0.01$; *: $p < 0.05$

Estudo 3: resultados

TABLE III
REGRESSION MODELS FOR COMMITS THAT INCREASE STRUCTURAL COMPLEXITY

Project	#Commits	Intercept	$n (H_1)$	$\Delta LOC (H_2)$	CF (H_3)	Adj. R^2
clojure	287	0.031637	0.000108	0.000147 ***	0.014175 ***	0.20
node	133	0.340228 ***	-0.000479 **	0.000369	0.025465 *	0.10
redis	72	-1.848899	0.004697	0.00244	0.236758 ***	0.31
voldemort	217	0.026552	-5e-06	0.000195 **	0.002326 ***	0.22
zeromq2	71	0.224476 **	-0.000399	0.000746 **	0.004965	0.21

***: $p < 0.001$; **: $p < 0.01$; *: $p < 0.05$

TABLE IV
REGRESSION MODELS FOR COMMITS THAT DECREASE STRUCTURAL COMPLEXITY

Project	#Commits	Intercept	$n (H_4)$	$\Delta LOC (H_5)$	CF (H_6)	Adj. R^2
clojure	235	-0.064833	0.000422 **	0.000204	0.011493	0.04
node	107	0.313235 *	-0.00038	-0.000308	0.036385 **	0.10
redis	45	-0.359334	0.000952	0.000443	0.163235 **	0.18
voldemort	177	0.029518 *	2.3e-05	0.00013 **	0.001125	0.07
zeromq2	57	0.10271	-0.000372	-0.001258 ***	0.021416 ***	0.88

***: $p < 0.001$; **: $p < 0.01$; *: $p < 0.05$

Estudo 3: resultados

TABLE III
REGRESSION MODELS FOR COMMITS THAT INCREASE STRUCTURAL COMPLEXITY

Project	#Commits	Intercept	$n (H_1)$	$\Delta LOC (H_2)$	CF (H_3)	Adj. R^2
clojure	287	0.031637	0.000108	0.000147 ***	0.014175 ***	0.20
node	133	0.340228 ***	-0.000479 **	0.000369	0.025465 *	0.10
redis	72	-1.848899	0.004697	0.00244	0.236758 ***	0.31
voldemort	217	0.026552	-5e-06	0.000195 **	0.002326 ***	0.22
zeromq2	71	0.224476 **	-0.000399	0.000746 **	0.004965	0.21

***: $p < 0.001$; **: $p < 0.01$; *: $p < 0.05$

TABLE IV
REGRESSION MODELS FOR COMMITS THAT DECREASE STRUCTURAL COMPLEXITY

Project	#Commits	Intercept	$n (H_4)$	$\Delta LOC (H_5)$	CF (H_6)	Adj. R^2
clojure	235	-0.064833	0.000422 **	0.000204	0.011493	0.04
node	107	0.313235 *	-0.00038	-0.000308	0.036385 **	0.10
redis	45	-0.359334	0.000952	0.000443	0.163235 **	0.18
voldemort	177	0.029518 *	2.3e-05	0.00013 **	0.001125	0.07
zeromq2	57	0.10271	-0.000372	-0.001258 ***	0.021416 ***	0.88

***: $p < 0.001$; **: $p < 0.01$; *: $p < 0.05$

Estudo 3: conclusões

- Todos os fatores influenciam ao menos em 2 casos
- Difusão de Mudança é o fator mais influente (7/10)
- Variação de Tamanho influente em 5/10
 - Variação no tamanho nem sempre influencia variação na complexidade
- Projetos diferentes influenciados por fatores diferentes
- Aumento e redução da complexidade influenciados por fatores diferentes
- Existem outros fatores

Project	$\Delta SC > 0$	Adj. R^2	$\Delta SC < 0$	Adj. R^2
Clojure	$\Delta LOC, CF$	0.20	Experience	0.04
Node	Experience, CF	0.10	CF	0.10
Redis	CF	0.31	CF	0.18
Voldemort	$\Delta LOC, CF$	0.22	ΔLOC	0.07
Zeromq2	ΔLOC	0.21	$\Delta LOC, CF$	0.88

Estudos experimentais

Considerações gerais

Estudo 1

Estudo 2

Estudo 3

Estudo 4

Estudo 4: objetivos

Obter modelos com maior coeficiente de determinação em relação estudo 3.

Refinamentos realizados:

- Refinamento da difusão de mudança
(módulos alterados + módulos adicionados)
- Inclusão do grau de autoria

Estudo 4: hipóteses

H_1 : considerar módulos alterados e adicionados separadamente melhora o coeficiente de determinação

H_2 : introduzir grau de autoria melhora o coeficiente de determinação

H_3 : considerar módulos alterados e adicionados separadamente e introduzir o grau de autoria melhora o coeficiente de determinação

$H_4 - H_8$: §variável influencia o aumento da complexidade estrutural

$H_9 - H_{13}$: §variável influencia a redução da complexidade estrutural

Estudo 4: projeto experimental

Amostra: 5 projetos de médio porte (C/C++/Java)

Unidade de análise: commit

Variáveis independentes:

- *Exp* – experiência. Número de mudanças anteriores
- ΔLOC – variação no tamanho
- *CM* – módulos modificados
- *AM* – módulos adicionados
- *FA* – primeira autoria
- *DL* – modificações prévias
- *AC* – modificações prévias por outros

Estudo 4: resultados

- ✓ H_1 : considerar módulos alterados e adicionados separadamente melhora o coeficiente de determinação (3/10 casos)
- ✓ H_2 : considerar o grau de autoria melhora o coeficiente de determinação (3/10 casos)
- ✓ H_3 : considerar módulos alterados e adicionados separadamente e introduzir o grau de autoria melhora o coeficiente de determinação (8/10 casos)

Estudo 4: conclusões

- ΔLOC foi o fator mais influente
- Todos os componentes do grau de autoria (FA, DL, AC) influenciaram em pelo menos 3 casos
- Modelos apresentam grande variação nos seus coeficientes de determinação

Projeto	Fatores para $\Delta SC > 0$	R^2	Fatores para $\Delta SC < 0$	R^2
clojure	$AC, \Delta LOC$	0.11	FA, DL, AC, CM	0.69
node	$E, FA, \Delta LOC$	0.35	ΔLOC	0.04
redis	AM	0.93	FA, DL, AC	0.36
voldemort	$\Delta LOC, CM, AM$	0.65	CM	0.07
zeromq2	DL, AM	0.33	$DL, \Delta LOC$	0.48

Complexidade Estrutural

Fatores

Teoria

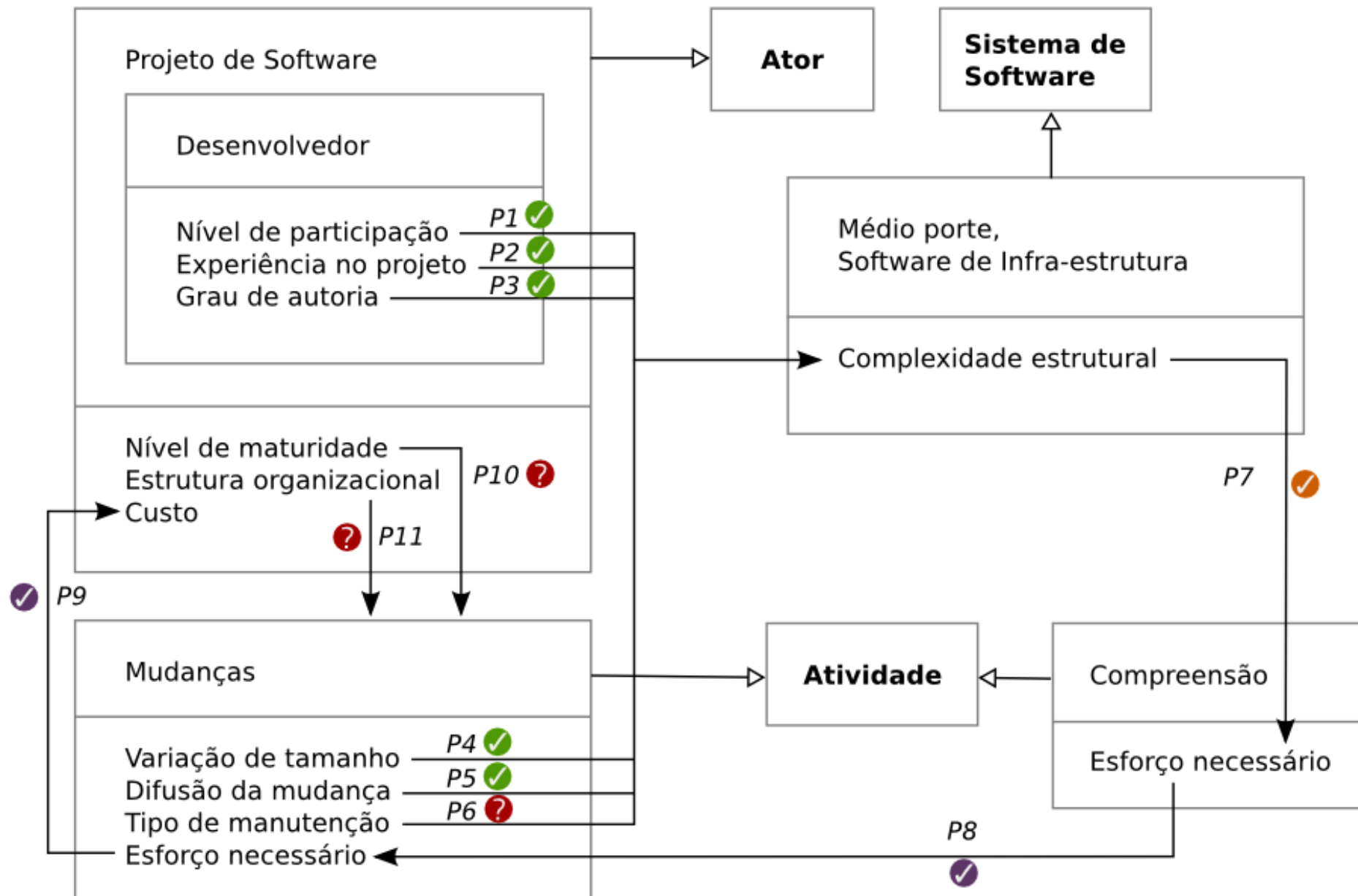
Estudos experimentais

Discussão

Conclusões

Avaliação da teoria

Suporte Experimental

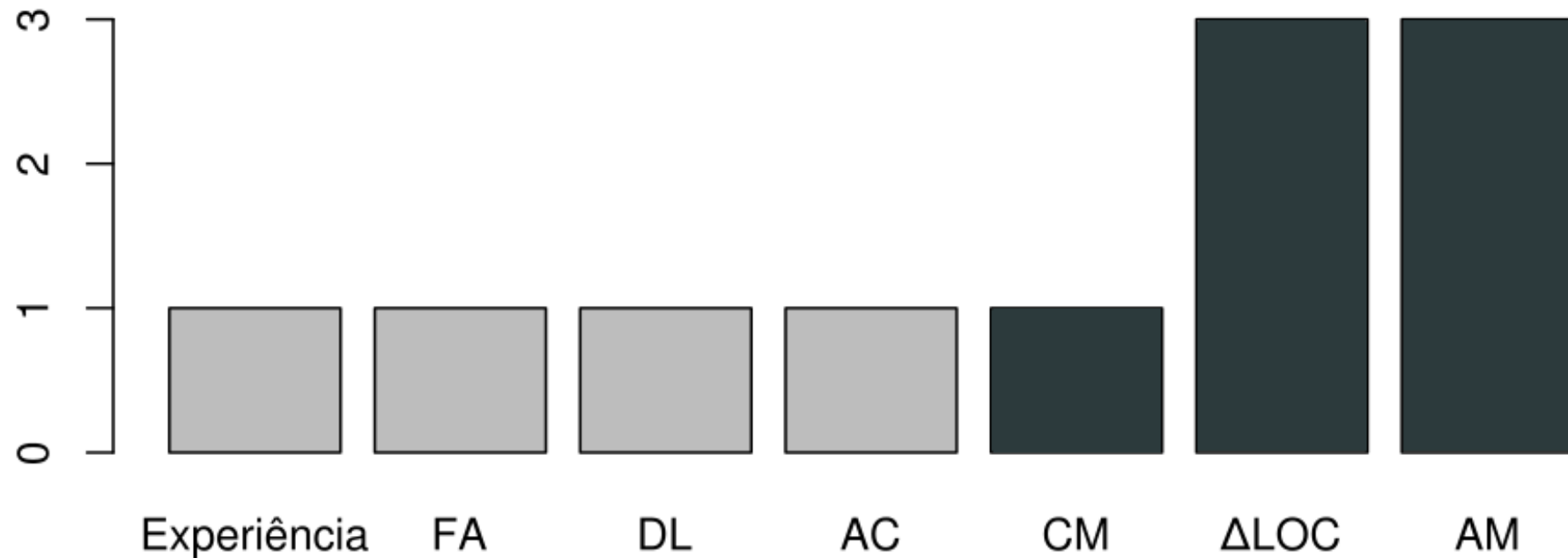


Avaliação da Teoria

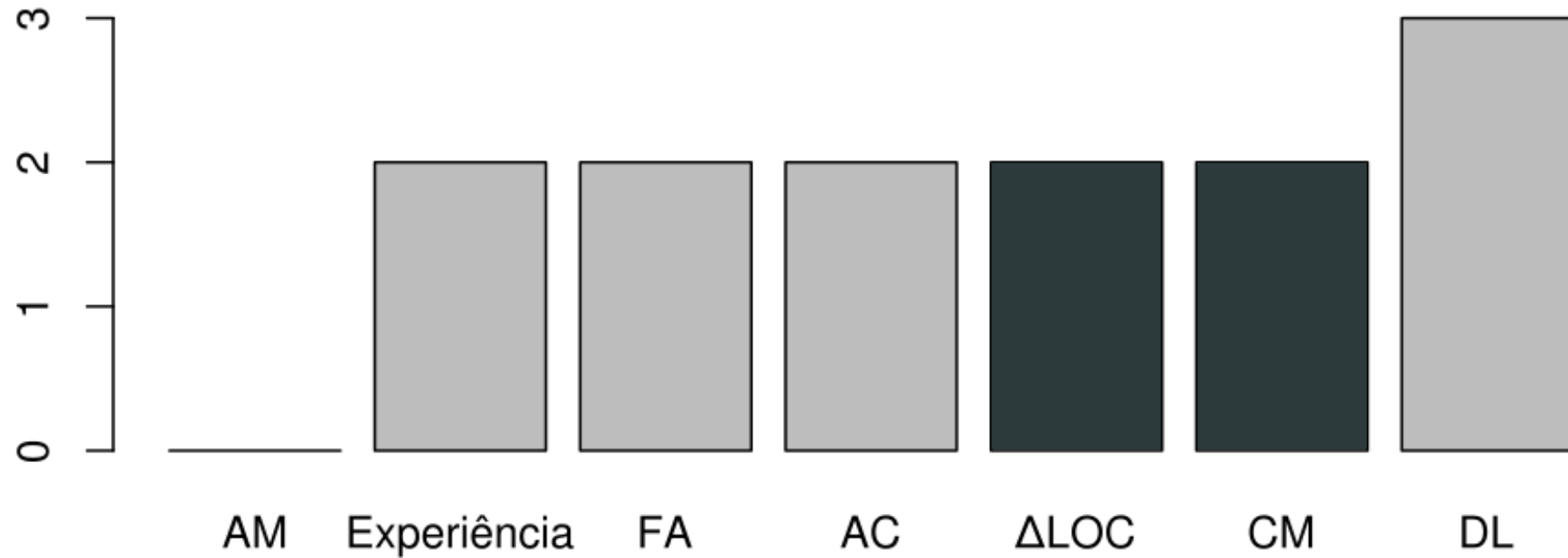
Critério	Avaliação
Testabilidade	Alta
Suporte experimental	Baixo
Poder de explicação	Baixo
Parcimônia	Média
Generalidade	Baixa
Utilidade	Alta

Fatores que influenciam a complexidade estrutural

Frequência dos fatores que influenciam o aumento da complexidade estrutural



Frequência dos fatores que influenciam a redução da complexidade estrutural



Complexidade Estrutural

Fatores

Teoria

Estudos experimentais

Discussão

Conclusões

Contribuições

Descontinuidades na complexidade
estrutural podem indicar mudanças
arquiteturais

Novas diferenças entre
desenvolvedores centrais
e periféricos

Diferentes fatores influenciam
diferentes projetos

Fatores diferentes influenciam
aumento e redução da complexidade
estrutural

Todos os fatores estudados possuem influência

Aumento da complexidade: maior frequência de fatores relacionados à manutenção

Redução da complexidade: frequência similar para fatores relacionados à manutenção e *fatores humanos*

Limitações

- Identificação dos fatores não foi sistemática
- Teoria construída de forma bottom-up
- Alguns modelos obtidos possuem baixo R^2
 - Estudos qualitativos são necessários para identificar outros fatores

Trabalhos Futuros

- Replicação dos estudos
- Explorar diferentes aspectos da complexidade estrutural além da média (e.g. variância)
- Estudo de caso com projeto existente
- utilizar complexidade estrutural como preditor para atributos de qualidade externa

