

## **Análise Descritiva dos Dados do Repositório do WordPress**

*Caio Ladeira Dornellas<sup>1</sup>, Cláudio Oliveira <sup>1</sup>  
Centro Universitário Academia, Juiz de Fora, MG*

*Tassio Ferenzini Martins Sirqueira<sup>2</sup>  
Centro Universitário Academia, Juiz de Fora, MG*

Linha de Pesquisa: Engenharia de Software

### **RESUMO**

Compreender a evolução das comunidades de desenvolvedores em torno de projetos de software de código aberto pode fornecer informações valiosas sobre o processo sociotécnico do desenvolvimento. Os repositórios de código-fonte são fontes ricas da história da evolução de projetos de software e podem ser acessados e estudados abertamente. Com a análise de dados históricos contidos em repositórios de código-fonte, buscou-se neste trabalho detectar e identificar padrões que podem explicar a evolução dos sistemas de software. Através de ferramentas de extração de logs específicas, foi coletado os dados para análise dos repositórios, em especial do WordPress, objeto deste estudo. Os dados extraídos foram agrupados para uma compor uma análise descritiva. Por meio dos resultados obtidos identificou-se um comportamento padrão ao longo do tempo nos contribuidores da comunidade do WordPress.

**Palavras-chave:** Código fonte. GitHub. Mineração de Repositórios.

### **ABSTRACT**

*Understanding the evolution of developer communities around open-source software projects can provide valuable information about the sociotechnical process of development. Source code repositories are rich sources of the*

---

<sup>1</sup> Discente do Curso de Engenharia de Software do Centro Universitário Academia – UniAcademia.

<sup>2</sup> Docente do Curso de Engenharia de Software do Centro Universitário Academia – UniAcademia. Orientador.

*evolutionary history of software projects and can be openly accessed and studied. By analyzing historical data contained in source code repositories, this work sought to detect and identify patterns that can explain the evolution of software systems. Through specific log extraction tools, data was collected to analyze the repositories, especially WordPress, the object of this study. The extracted data were grouped to compose a descriptive analysis. Through the results obtained a pattern of behavior was identified over time in the WordPress community contributors.*

**Keywords:** Source code. GitHub. Repository Mining.

## 1 INTRODUÇÃO

Nos últimos anos, houve um crescimento exponencial no número de artigos que mencionam “ecossistema” como tema (AMRIT *et al.*, 2022). O WordPress é um ecossistema de software que visa democratizar a publicação de sites, garantindo que qualquer pessoa não técnica possa criar um site, ao mesmo tempo que pode ser dimensionado até clientes corporativos com necessidades complexas.

A evolução das aplicações web, em especial dos ecossistemas de software, ganhou importância no processo de desenvolvimento e o fator evolução não pode ser ignorado. O processo de evolução do software desempenhou um papel importante durante o desenvolvimento do software e milhões de aplicações são desenvolvidos todos os anos em todo o mundo. Contudo, não há técnicas adequadas e sistemáticas disponíveis para a evolução de ecossistemas de software. Para Fernandez-Ramil *et al.* (2008), a evolução do software é o fenômeno da mudança de software ao longo dos anos e lançamentos, desde a concepção (formação do conceito) até o descomissionamento de um sistema de software. O termo “evolução” traz o foco para as mudanças gradativas implementadas no sistema, entretanto, algumas pessoas preferem descrever esse fenômeno como simplesmente manutenção de software, contudo, “manutenção”, a ênfase está em manter a satisfação das partes interessadas com o software durante a vida útil.

Esse trabalho é continuação do trabalho desenvolvido por Oliveira, Tostes e Sirqueira (2021). O objetivo deste estudo é realizar uma análise descritiva do WordPress baseando-se nas leis de evolução do software de Lehman (LEHMAN, 1996), com base nos dados coletados pela ferramenta GitLogAnalysis (SILVA e SIRQUEIRA, 2022) no repositório do WordPress. O WordPress, como um exemplo típico de ecossistemas de software baseados em plug-ins, é usado por milhões de sites e possui muitos plug-ins disponíveis (LIN *et al.*, 2022). Nossa suposição ao

estudar o WordPress é que tal ecossistema pode ser melhorado as descobertas de estudos empíricos de sistemas de software de longa duração.

Além dessa introdução, na seção 2 apresentaremos o referencial teórico do trabalho, já a seção 3 apresenta os trabalhos relacionados. A seção 4 explora os resultados da análise. Por fim, a seção 6 apresenta as limitações do trabalho, os próximos passos e as considerações finais.

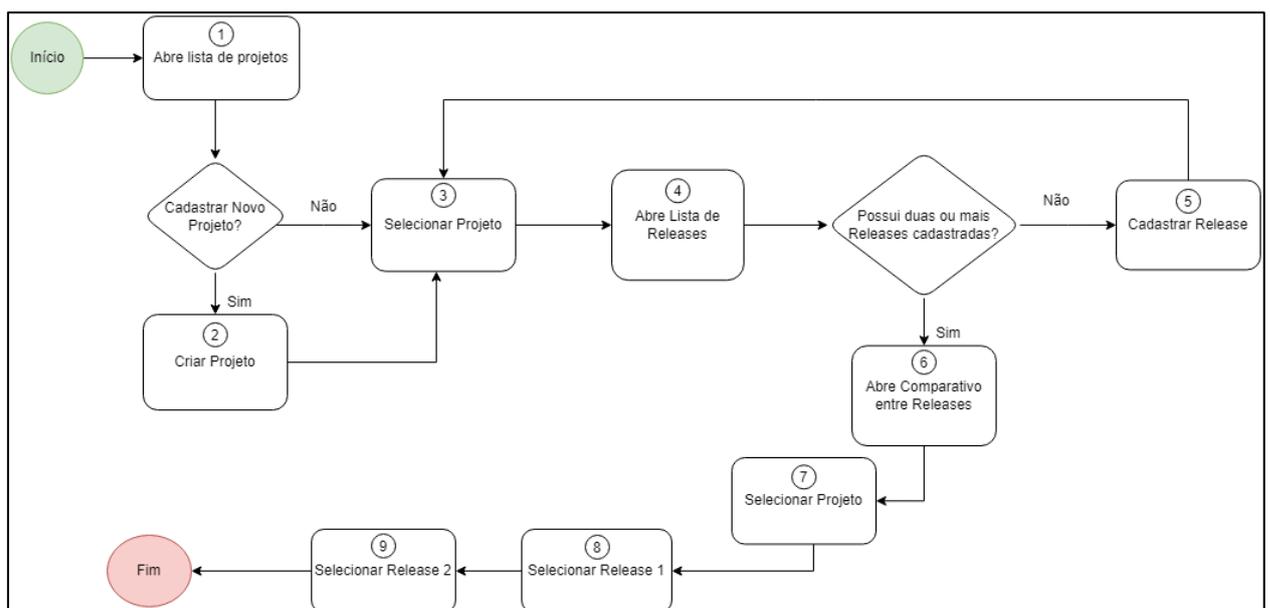
## 2 REFERENCIAL TEÓRICO

A mineração de repositórios de código fonte auxilia a validar as leis de evolução de software, além de auxiliar os gerentes de projetos a entenderem a dinâmica do sistema mantido, assim como a do grupo mantenedor. A seguir abordaremos sobre repositórios de código fonte, sobre o GitHub e mineração de repositórios.

### 2.1 GitLog Analysis

O GitLog Analysis, uma ferramenta que foi desenvolvida por Silva & Sirqueira (2022), que analisa e compara as métricas de software que tenha usado o Git como o versionador de código fonte. As informações geradas podem trazer uma melhor visão do projeto com base nos dados das *releases* geradas ao longo do tempo. Os dados presentes no arquivo de log do Git (*git log*) são extraídos e exibido em um dashboard da ferramenta. Um fluxograma da execução da ferramenta, que pode ser observado na Figura 1.

**Figura 1. Fluxograma de uso da ferramenta.**

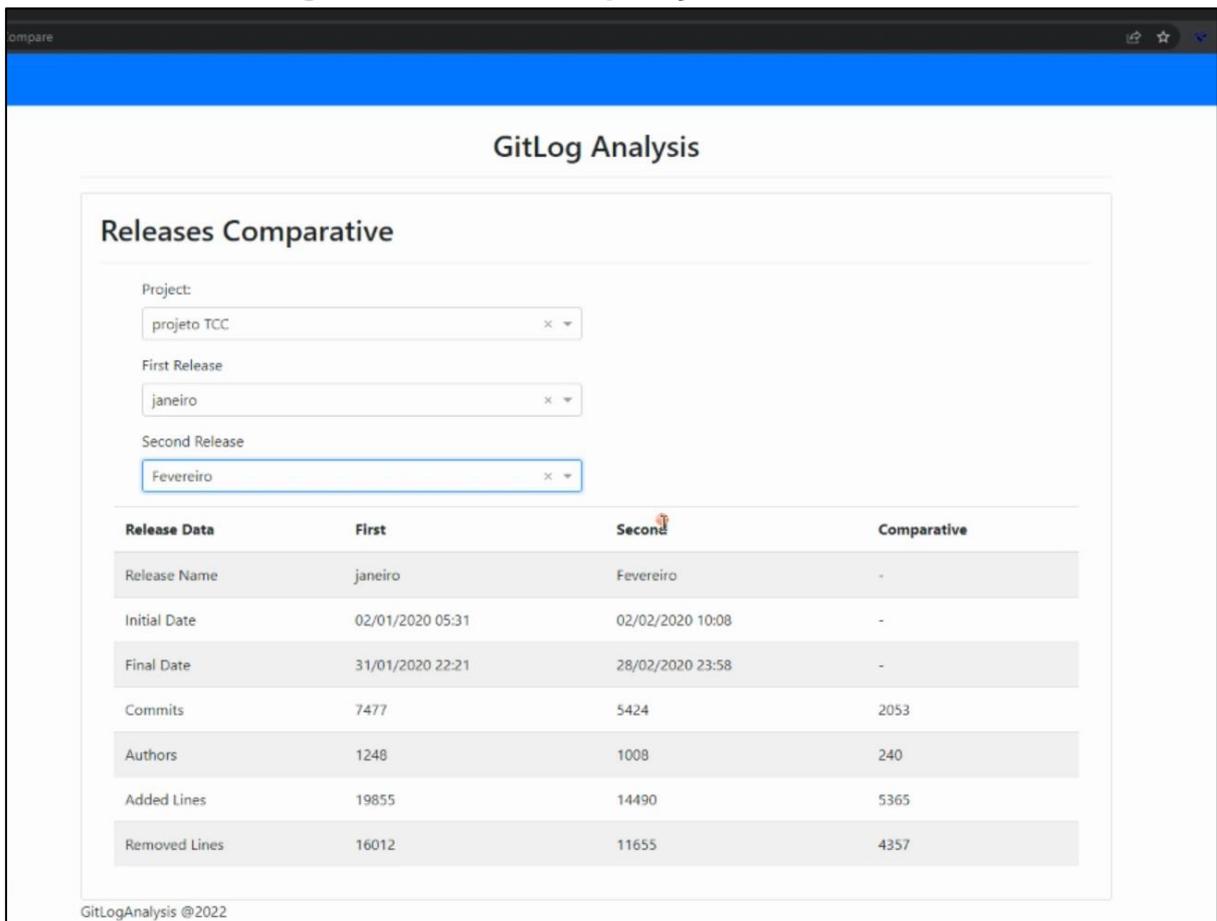


**Fonte:** Silva & Sirqueira (2022).

Para coleta e análise dos dados, é necessário escolher os períodos a serem analisados. Dessa forma, a ferramenta auxilia de forma visual na análise da evolução projeto com base nessas métricas que cada release possui.

A ferramenta GitLog Analysis pode apoiar os usuários nas atividades de análise do ciclo de vida de um software e os dados utilizados neste trabalho foram coletados pela ferramenta GitLog Analysis. Um exemplo da tela do software pode ser visto na Figura 2.

**Figura 2. Tela de Comparação das Releases.**



**Fonte:** Silva & Sirqueira (2022).

## 2.2 WordPress

O WordPress é oficialmente um Sistema de Gerenciamento de Conteúdo (CMS) que alimenta mais de 35% de todos os sites da internet. Um CMS é uma aplicação para a internet que permite que donos de sites, editores e autores

<sup>3</sup> O que é WordPress. Disponível em: < <https://www.hostinger.com.br/tutoriais/o-que-e-wordpress> >. Acessado em 30 de novembro de 2022.



gerenciem seus sites e publiquem conteúdo sem precisar de conhecimentos em programação.

O WordPress usa a linguagem de programação PHP4 e o banco de dados MySQL5. O WordPress é considerado um ecossistema de software por conta de sua possibilidade de integração com outros sistema e complementos que atualmente são mais de 55.0006.

O WordPress começou em 2003 criado a partir de um software chamado bs/cafeblog. A primeira versão lançada do WordPress (WordPress 1.0) foi em janeiro de 2004 e desde então o ecossistema recebendo atualizações frequentes e ampliando-se em termos de recursos.

Conforme dados do site oficial, desenvolvedores, criadores de conteúdo e donos de sites se encontram mensalmente em *meetups* em 817 cidades pelo mundo.

### 3 TRABALHOS RELACIONADOS

Em Cabot (2018), o autor abordou o crescimento do ecossistema nos últimos anos, o que o tornou o principal sistema gerenciador de conteúdo do mundo. O trabalho discute as métricas que o software, como a comunidade vem mantendo o software, em especial no desenvolvimento de plug-ins. Além disso, trabalho apresenta a fase atual do WordPress e as previsões para o futuro.

No trabalho apresentado em Patel *et al.* (2011), os autores apresentam uma comparação de desempenho entre os 3 principais sistemas gerenciadores de conteúdo, a saber, WordPress, Joomla e Drupal. O objetivo foi determinar quais as condições de trabalho ideal de cada plataforma. Além disso aborda as métricas de software e como as plataformas vêm evoluindo ao longo do tempo.

O trabalho de Shahzad *et al.* (2012) explora como as vulnerabilidades afetam o ciclo de vida de um software; incluindo discussões de como uma vulnerabilidade evolui entre versões e pode comprometer a segurança de um software. O estudo não se restringiu ao WordPress, mas apresentou resultados significativos sobre a implicação de uma vulnerabilidade no ciclo de vida de um software.

Por fim, o trabalho de UM *et al.* (2015) utiliza o WordPress como ecossistema para mensurar a taxa de inovação ao longo do tempo e como essa inovação afeta o ciclo de vida do software, englobando o núcleo do software e os plug-ins mais utilizados no mercado.

---

<sup>4</sup> PHP. Disponível em: < <https://www.php.net/> >. Acessado em 30 de novembro de 2022.

<sup>5</sup> MySQL. Disponível em: < <https://www.mysql.com/> >. Acessado em 30 de novembro de 2022.

<sup>6</sup> WordPress. Disponível em: < <https://br.wordpress.org/> >. Acessado em 30 de novembro de 2022.

## 4 RESULTADOS

Para a realização da análise do repositório do WordPress, em 01 de junho de 2022, via ferramenta GitLog Analysis, gerou-se um *dump* de todo o histórico da base desde a criação.

Uma vez que o objetivo deste estudo é analisar o ciclo de vida do software WordPress; definiu-se as seguintes questões de pesquisa para conduzir a análise, separadas em questão geral e específicas:

Questão Geral:

- QG – A comunidade do WordPress manteve-se ativa ao longo do tempo?

Questões específicas:

- QE1 – A quantidade de contribuidores cresceu ao longo do tempo?
- QE2 – Os contribuidores se mantêm ativos ao longo do tempo?
- QE3 – Existe uma estabilidade entre o número de *commits* entre cada versão?

A primeira análise foi levantar o nome e e-mail de todos os contribuidores do WordPress ao longo dos anos e, além disso, a quantidade de *commits* que cada usuário realizou, a fim de verificar quem são os principais contribuidores do projeto. O objetivo foi determinar quem são os contribuídos e o quanto eles contribuem com a comunidade do WordPress.

O resultado desta análise pode ser observado na Tabela 1.

**Tabela 1. Lista de contribuidores e número de *commits*.**

Lista de Contribuidores		
Nome	E-mail	Quantidade de Commits
ryan	ryan@1a063a9b-81f0-0310-95a4-ce76da25c4cd	7153
Sergey Biryukov	sergeybiryukov.ru@gmail.com	5366
nacin	nacin@1a063a9b-81f0-0310-95a4-ce76da25c4cd	2705
Andrew Nacin	wp@andrewnacin.com	2258
Scott Taylor	scott.c.taylor@mac.com	2047
matt	matt@1a063a9b-81f0-0310-95a4-ce76da25c4cd	714
azaozz	azaozz@1a063a9b-81f0-0310-95a4-ce76da25c4cd	1686



Drew Jaynes	info@drewapicture.com	1665
Andrew Ozz	admin@laptoptips.ca	1387
Dominik Schilling	dominikschilling+git@gmail.com	1212
westi	westi@1a063a9b-81f0-0310-95a4- ce76da25c4cd	1201
saxmatt	saxmatt@1a063a9b-81f0-0310-95a4- ce76da25c4cd	1200
John Blackburn	johnbillion@git.wordpress.org	1094
markjaquith	markjaquith@1a063a9b-81f0-0310-95a4- ce76da25c4cd	1057
Boone Gorges	boonebgorges@gmail.com	811
desrosj	desrosj@git.wordpress.org	760
Lance Willett	nanobar@gmail.com	758
Gary Pendergast	gary@pento.net	702
Weston Ruter	weston@xwp.co	623
dd32	dd32@1a063a9b-81f0-0310-95a4- ce76da25c4cd	603
Helen Hou-Sandí	helen.y.hou@gmail.com	575
Dion Hulse	wordpress@dd32.id.au	480
rboren	rboren@1a063a9b-81f0-0310-95a4- ce76da25c4cd	465
Andrea Fercia	a.fercia@gmail.com	454
scribu	scribu@1a063a9b-81f0-0310-95a4- ce76da25c4cd	408
Bot (Assets)	nobody@localhost	355
Jeremy Felt	jeremy.felt@gmail.com	353
Ryan Boren	ryan@boren.nu	343
Mark Jaquith	mark@wordpress.org	330
koopersmith	koopersmith@1a063a9b-81f0-0310-95a4- ce76da25c4cd	320
Daryl Koopersmith	dkoopersmith@gmail.com	317
Peter Wilson	wilson@peterwilson.cc	283
Aaron Jorbin	aaron@jorb.in	273
bumpbot	bumpbot@1a063a9b-81f0-0310-95a4- ce76da25c4cd	269
rob1n	rob1n@1a063a9b-81f0-0310-95a4- ce76da25c4cd	255
wpmuguru	wpmuguru@1a063a9b-81f0-0310-95a4- ce76da25c4cd	247
mikelittle	mikelittle@1a063a9b-81f0-0310-95a4- ce76da25c4cd	237

Pascal Birchler	pascal.birchler@gmail.com	229
audrasjb	audrasjb@git.wordpress.org	227
Konstantin Obenland	obenland@automattic.com	217
michelvaldrighi	michelvaldrighi@1a063a9b-81f0-0310-95a4- ce76da25c4cd	214
whyisjake	whyisjake@git.wordpress.org	205
hellofromTonya	hellofromTonya@git.wordpress.org	186
Felix Arntz	felix-arntz@leaves-and-love.net	173
duck_	duck_@1a063a9b-81f0-0310-95a4- ce76da25c4cd	164
Matt Thomas	iammatthomas@gmail.com	163
TimothyBlynJacob s	TimothyBlynJacobs@git.wordpress.org	132
lancewillett	lancewillett@1a063a9b-81f0-0310-95a4- ce76da25c4cd	129
Rachel Baker	rachel@rachelbaker.me	126
potbot	potbot@1a063a9b-81f0-0310-95a4- ce76da25c4cd	125
gziolo	gziolo@git.wordpress.org	122
Adam Silverstein	adamsilverstein@earthboundhosting.com	113
emc3	emc3@1a063a9b-81f0-0310-95a4- ce76da25c4cd	110
Ella Iseulde Van Dorpe	*@iseulde.com	109
Ian Stewart	ian@themeshaper.com	109
alex_t_king	alex_t_king@1a063a9b-81f0-0310-95a4- ce76da25c4cd	83
Bot (Translations)	nobody@localhost	80
ryelle	ryelle@git.wordpress.org	78
noisysocks	noisysocks@git.wordpress.org	77
Joe McGill	joemcgill@gmail.com	74
nbachiyski	nbachiyski@1a063a9b-81f0-0310-95a4- ce76da25c4cd	74
Ryan McCue	me@ryanmccue.info	72
youknowriad	youknowriad@git.wordpress.org	70
David A. Kennedy	me@davidakennedy.com	66
joedolson	joedolson@git.wordpress.org	65
K. Adam White	kadamwhite@gmail.com	62
iammatthomas	iammatthomas@1a063a9b-81f0-0310-95a4- ce76da25c4cd	57



ianbelanger	ianbelanger@git.wordpress.org	57
iandunn	iandunn@git.wordpress.org	57
jorgefilipecosta	jorgefilipecosta@git.wordpress.org	57
Mike Schroder	mschrode@umich.edu	57
atimmer	atimmer@git.wordpress.org	56
Peter Westwood	peter.westwood@ftwr.co.uk	56
davidbaumwald	davidbaumwald@git.wordpress.org	55
iandstewart	iandstewart@1a063a9b-81f0-0310-95a4- ce76da25c4cd	54
James Nylen	jnylen@gmail.com	52
Joe Hoyle	joehoyle@gmail.com	51
antpb	antpb@git.wordpress.org	43
Mel Choyce	melchoyce@gmail.com	39
Eric Lewis	eric.andrew.lewis@gmail.com	38
donncha	donncha@1a063a9b-81f0-0310-95a4- ce76da25c4cd	35
Tammie Lister	tammie@automattic.com	35
spacedmonkey	spacedmonkey@git.wordpress.org	32
PeteMall	PeteMall@1a063a9b-81f0-0310-95a4- ce76da25c4cd	31
Jon Cave	jon@lionsgoroar.co.uk	29
laurelfulford	laurelfulford@git.wordpress.org	28
jverber	jverber@1a063a9b-81f0-0310-95a4- ce76da25c4cd	23
johnjamesjacoby	johnjamesjacoby@git.wordpress.org	17
Aaron Campbell	aaron@xavisys.com	15
Clorith	Clorith@git.wordpress.org	14
Nikolay Bachiyski	nb@nikolay.bg	14
josephscott	josephscott@1a063a9b-81f0-0310-95a4- ce76da25c4cd	13
jffng	jffng@git.wordpress.org	10
ellatrix	ellatrix@git.wordpress.org	9
Konstantin Kovshenin	kovshenin@gmail.com	8
tellyworth	tellyworth@git.wordpress.org	8
Michael Adams	mda@blogwaffe.com	7
aduth	aduth@git.wordpress.org	6
Matt Mullenweg	m@mullenweg.com	5
jrf	jrf@git.wordpress.org	4
Michael Arestad	marestad@gmail.com	3
isabel_brison	isabel_brison@git.wordpress.org	2

Matias Ventura	mv@matiasventura.com	2
(no author)	(no author)@1a063a9b-81f0-0310-95a4- ce76da25c4cd	1
allancole	allancole@git.wordpress.org	1
dougal	dougal@1a063a9b-81f0-0310-95a4- ce76da25c4cd	1
omarreiss	omarreiss@git.wordpress.org	1
uid26904	uid26904@1a063a9b-81f0-0310-95a4- ce76da25c4cd	1

**Fonte: Elaboração própria.**

Na sequência, buscou-se determinar a média de *commits* por versão lançada do WordPress. O objetivo dessa análise é determinar se existe uma similaridade entre o número de *commits* de cada versão.

O resultado desta análise encontra-se na Tabela 2.

**Tabela 12. Média de *commits* por versão.**

Versão	Média de Versões por Dia
.70	3,12
1.0	4,64
1.2	3,79
1.5	3,14
2.0	2,9
2.1	4,68
2.2	4,45
2.3	7,1
2.5	4,9
2.6	11,23
2.7	6,87
2.8	3,94
2.9	15,32
3.0	8,04
3.1	6,27
3.2	7,28
3.3	7,83
3.4	11,06
3.5	6,9
3.6	9,69
3.7	18,77
3.8	8,72
3.9	10,59



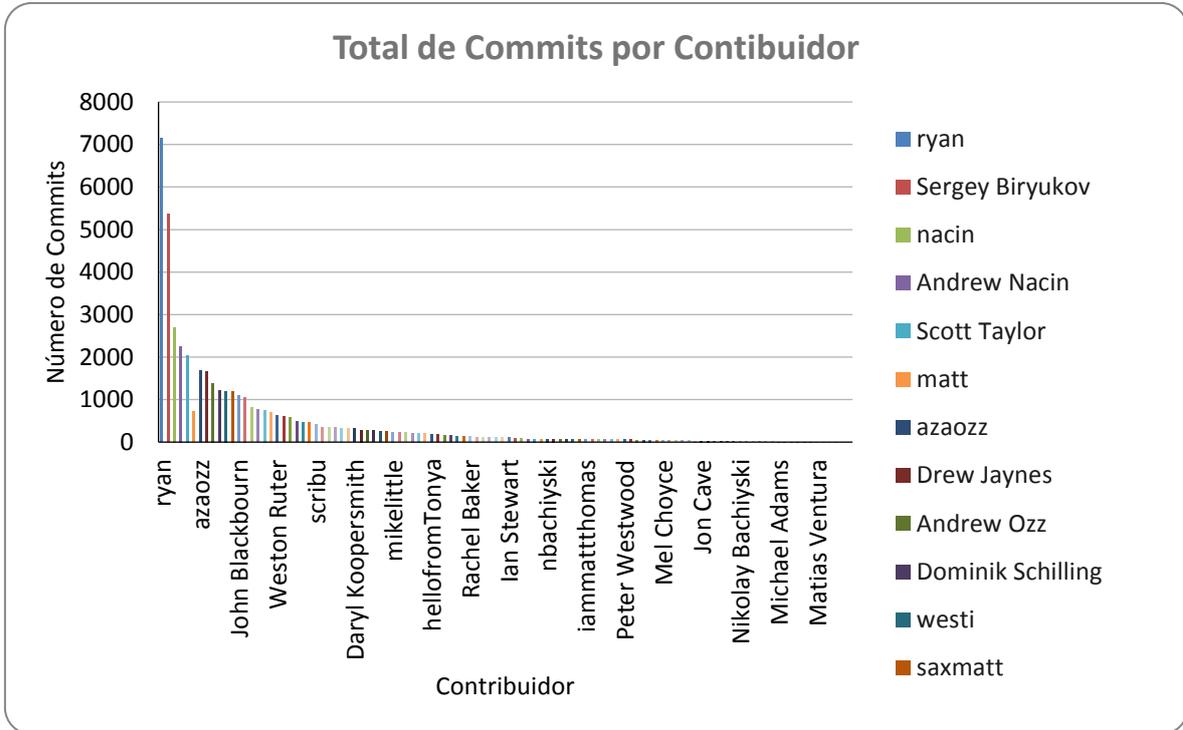
4.0	12,77
4.1	9,43
4.2	9,57
4.3	18,84
4.4	9,16
4.5	6,84
4.6	9,94
4.7	0,25
4.8	6,3
4.9	2,22
5.0	7,65
5.1	5,74
5.2	6,52
5.3	3,61
5.4	7,34
5.5	6,16
5.6	5,47
5.7	5,29
5.8	4,91
5.9	5,32
6.0	1,25

**Fonte: Elaboração própria.**

Para melhorar a visualização de contribuições, a Figura 3 apresenta um gráfico com os maiores contribuidores do projeto ao longo dos anos. Na Figura 3 é possível observar que existem mais de 10 (dez) membros da comunidade com mais de 1000 (mil) *commits*. Além disso, destaca-se que o maior contribuidor possui mais de 7000 (sete mil) *commits*,

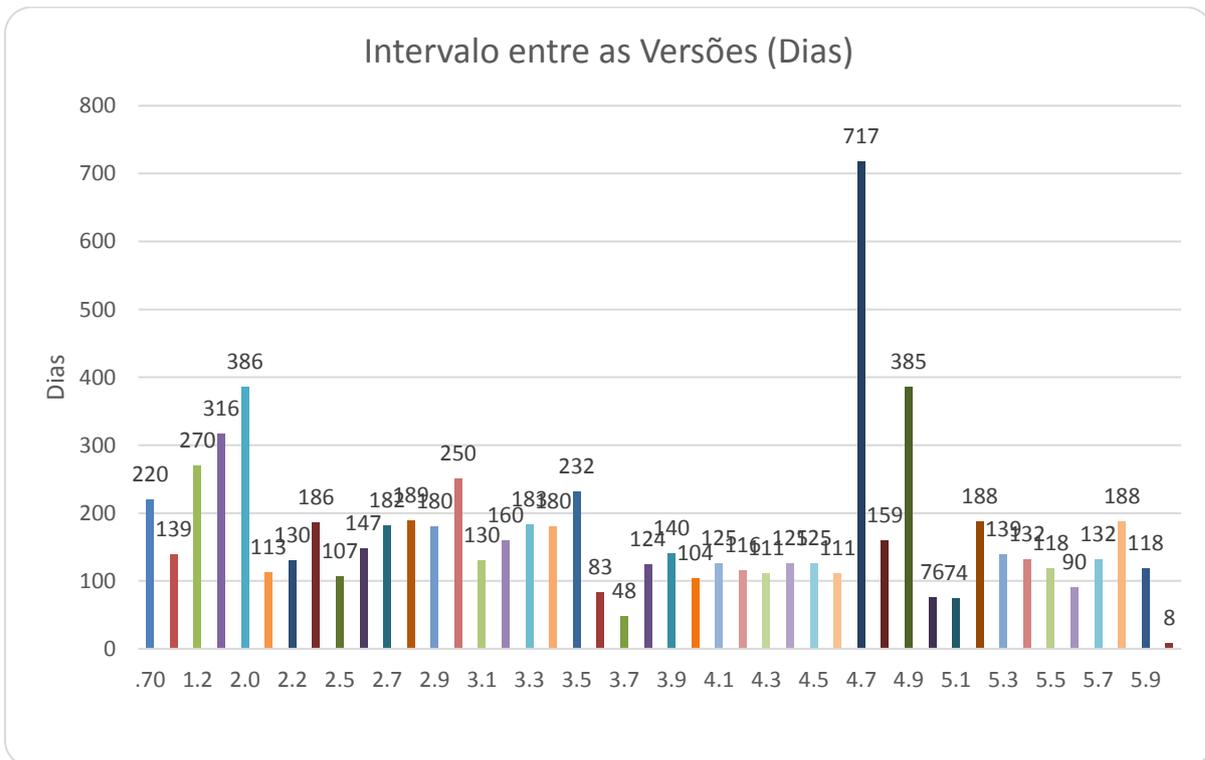
Já na Figura 4 temos a lista de versões lançadas e o intervalo entre dias do lançamento de cada versão. Algumas versões como a 2.0, 4.7 e 4.9 que diferem das demais por ultrapassar 1 (um) ano da versão predecessora.

**Figura 3. Maiores contribuidores do projeto.**



Fonte: Elaboração própria.

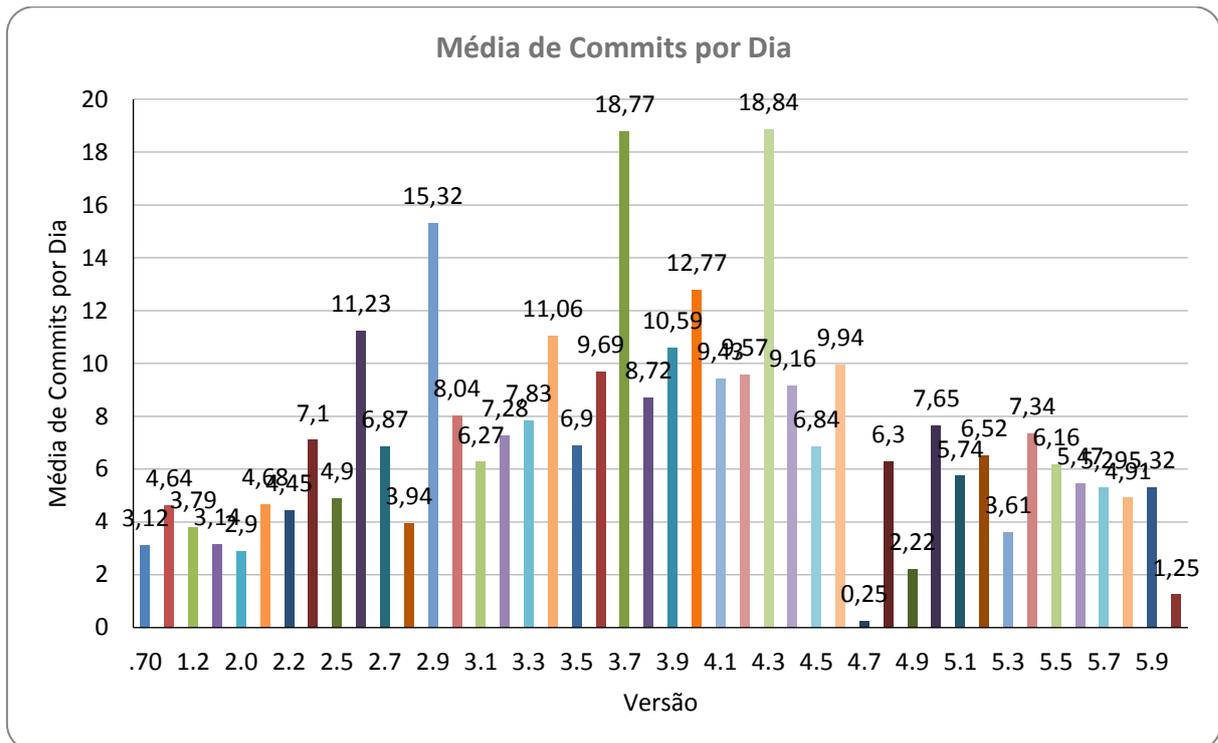
**Figura 4. Intervalo de dias entre as versões.**



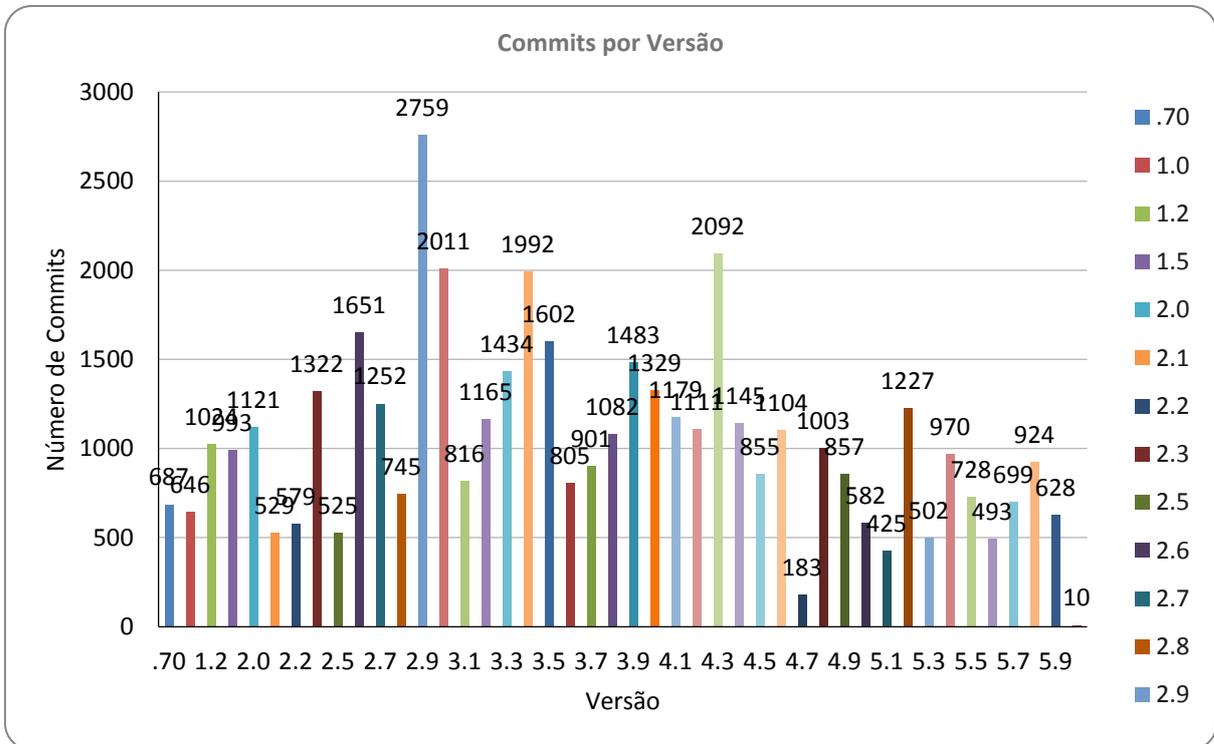
Fonte: Elaboração própria.

A Figura 5 representa a média de *commits* por dia que o repositório do WordPress recebe da comunidade por versão. Pretende-se no futuro explorar melhor quais os tipos de *commits*, para entendermos se são melhorias e/ou correções que a comunidade identificou ser necessária.

**Figura 5. Média de *commits* por dia.**



**Figura 6. Total de *commits* por versão.**



**Fonte: Elaboração própria.**

Conforme pode-se observar pelos dados, a comunidade se mantém ativa ao longo do tempo, é o que importante para u ecossistema de software manter sua longevidade, visto que o apoio da comunidade é peça chave disto.

## 5 CONSIDERAÇÕES FINAIS

Esse trabalho é resultado de uma iniciação científica do UniAcademia e buscou explorar a partir dos dados do repositório do WordPress como a comunidade vem mantendo o ecossistema ao longo do tempo.

De modo geral, conforme verificado ao longo de cada análise da seção 4, podemos concluir que a comunidade do ecossistema WordPress se manteve ativa ao longo de sucessivas evoluções, respondendo nossa questão geral. Isso se deve principalmente ao aumento de sites e sistema baseados no WordPress.

Nosso trabalho está sujeito a algumas ameaças à validade, a saber: (1) validade interna, relacionada às coletas e análise; e (2) validade externa, que discute a generalização de nossos resultados no contexto do WordPress.

Algumas limitações do trabalho são:



- A ferramenta utilizada para coleta dos dados limita-se ao versionador Git;
- Utilizou-se somente o repositório do WordPress durante o estudo;
- Carece de uma análise estatística validando as questões do estudo.

Como objetivos futuros, pretende-se melhorar os seguintes pontos:

- Novas análise, ampliando o número de repositórios;
- Realizar estudos estatísticos com os dados;
- Ampliar a quantidade de informações extraída dos repositórios.

Vale ressaltar que compreender a dinâmica do ciclo de vida de um software não é uma tarefa trivial, é um caminho que permite melhorarmos tal compreensão é a partir de dados históricos disponíveis em repositórios de código fonte.

## REFERÊNCIAS

- AMRIT, Chintan et al. **Guest Editorial: Collaboration and Innovation Dynamics in Software Ecosystems**. IEEE Transactions on Engineering Management, v. 69, n. 6, p. 3584-3588, 2022.
- CABOT, Jordi. **WordPress: A content management system to democratize publishing**. IEEE Software, v. 35, n. 3, p. 89-92, 2018.
- FERNANDEZ-RAMIL, Juan et al. **Empirical studies of open source evolution**. Software evolution, p. 263-288, 2008.
- LEHMAN, Manny M. **Laws of software evolution revisited**. In: European Workshop on Software Process Technology. Springer, Berlin, Heidelberg, 1996. p. 108-124.
- LIN, Jiahuei; SAYAGH, Mohammed; HASSAN, Ahmed E. **The Co-evolution of the WordPress Platform and its Plugins**. ACM Transactions on Software Engineering and Methodology, 2022.
- OLIVEIRA, Cláudio; TOSTES, Rogério; SIRQUEIRA, Tassio Ferenzini Martins. **GIT Viewer: uma plataforma para análise de dados do GitHub**. ANALECTA-Centro Universitário Academia, v. 6, n. 3, 2021.
- PATEL, Savan K.; RATHOD, V. R.; PARIKH, Satyen. **Joomla, Drupal and WordPress-a statistical comparison of open source CMS**. In: 3rd International Conference on Trendz in Information Sciences & Computing (TISC2011). IEEE, 2011. p. 182-187.



SHAHZAD, Muhammad; SHAFIQ, Muhammad Zubair; LIU, Alex X. **A large scale exploratory analysis of software vulnerability life cycles**. In: 2012 34th International Conference on Software Engineering (ICSE). IEEE, 2012. p. 771-781.

SILVA, Gabriel Rigolon; SIRQUEIRA, Tassio Ferenzini Martins. **GitLogAnalysis: Uma ferramenta para acompanhar a evolução de Software via Git**. Caderno de Estudos em Sistemas de Informação, v. 9, n. 1, 2022.

UM, SungYong; YOO, Youngjin; WATTAL, Sunil. **The evolution of digital ecosystems: A case of WordPress from 2004 to 2014**. 2015.