



**Processo para análise de colaboração em
ecossistema de software:
Um estudo na plataforma SYDLE ONE**

André Dias Nunes

JUIZ DE FORA
JULHO, 2023

**Processo para análise de colaboração em
ecossistema de software:
Um estudo na plataforma SYDLE ONE**

ANDRÉ DIAS NUNES

Universidade Federal de Juiz de Fora
Instituto de Ciências Exatas
Departamento de Ciência da Computação
Bacharelado em Ciência da Computação

Orientador: Victor Stroele De Andrade Menezes

JUIZ DE FORA
JULHO, 2023

PROCESSO PARA ANÁLISE DE COLABORAÇÃO EM
ECOSSISTEMA DE SOFTWARE:
Um estudo na plataforma SYDLE ONE

André Dias Nunes

MONOGRAFIA SUBMETIDA AO CORPO DOCENTE DO INSTITUTO DE CIÊNCIAS
EXATAS DA UNIVERSIDADE FEDERAL DE JUIZ DE FORA, COMO PARTE INTE-
GRANTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE
BACHAREL EM CIÊNCIA DA COMPUTAÇÃO.

Aprovado por:

Victor Stroele De Andrade Menezes
Doutor em Engenharia de Sistemas e Computação

José Maria Nazar David
Doutor em Engenharia de Sistemas e Computação

Regina Maria Maciel Braga Villela
Doutora em Engenharia de Sistemas e Computação

Mario Antônio Ribeiro Dantas
Doutor em Engenharia de Sistemas e Computação

JUIZ DE FORA
11 DE JULHO, 2023

Resumo

A colaboração entre desenvolvedores sempre desempenhou um papel fundamental no aprendizado, manutenção e evolução da engenharia de *software*. Através do compartilhamento de conhecimentos e do trabalho conjunto, esses profissionais impulsionam o desenvolvimento da área. Atualmente, a adesão às plataformas que fomentam essa interação, está em constante crescimento, impulsionada pelo surgimento de novas tecnologias, demandas do mercado e até mesmo como consequência do contexto global de isolamento resultante da pandemia de COVID-19.

As comunidades que se formam em torno de uma plataforma específica são responsáveis por criar um Ecossistema de *Software* (ECOS), constituídos por desenvolvedores, empresas, usuários e demais *stakeholders* que se envolvem ativamente na sua construção, aprimoramento e utilização. ECOS fornecem um ambiente rico para a colaboração, inovação e compartilhamento de recursos, tornando-se um componente essencial do progresso contínuo da engenharia de software.

Visando se adaptar a essa tendência e manter a competitividade no mercado, diversas empresas adotam plataformas e *frameworks* de desenvolvimento, criando assim ECOS internos. O presente estudo tem como propósito investigar técnicas de análise de dados relacionadas a Sistemas de Reputação, originalmente propostas para ECOS de plataformas abertas, a fim de propor um processo de análise de ambientes reais e privados de uma empresa específica.

Foi conduzida uma Prova de Conceito com o propósito de obter *insights* valiosos sobre os colaboradores mais influentes, suas especialidades e impacto no projeto, dentro do ECOS de uma empresa. Essa análise tem como finalidade auxiliar na gestão de recursos humanos, proporcionando informações relevantes para maximização da eficiência organizacional. Os resultados obtidos apontam para a viabilidade da proposta.

Palavras-chave: Ecossistema de software, sistema de reputação, gestão de recursos humanos, inteligência empresarial, SYDLE ONE

Abstract

The collaboration among developers has always played a fundamental role in the learning, maintenance, and evolution of software engineering. Through knowledge sharing and teamwork, these professionals drive the development of the field. Currently, the adoption of platforms that foster this interaction is constantly growing, driven by the emergence of new technologies, market demands, and even as a consequence of the global context of isolation resulting from the COVID-19 pandemic.

The communities that form around a specific platform are responsible for creating a Software Ecosystem (SECO), comprised of developers, companies, users, and other stakeholders who actively engage in its construction, improvement, and utilization. SECOs provide a rich environment for collaboration, innovation, and resource sharing, becoming an essential component of the ongoing progress of software engineering.

In order to adapt to this trend and maintain competitiveness in the market, many companies adopt development platforms and frameworks, thus creating internal SECOs. The purpose of this study is to investigate data analysis techniques related to Reputation Systems, originally proposed for SECOs of open platforms, in order to propose an analysis process for real and private environments within a specific company.

A Proof of Concept was conducted to obtain valuable insights into the most influential collaborators, their specialties, and impact on the project within a company's SECO. This analysis aims to assist in human resource management, providing relevant information for maximizing organizational efficiency. The results obtained indicate the feasibility of the proposal.

Keywords: Software ecosystem, reputation system, human resource management, business intelligence, SYDLE ONE

Agradecimentos

A Deus, pela minha vida, pelas portas abertas e por me ajudar a ultrapassar todos os obstáculos.

Aos meus pais e irmãos pelo constante encorajamento e suporte. Suas palavras de incentivo foram fundamentais para o meu desenvolvimento e sucesso. Agradeço por estarem sempre ao meu lado, compartilhando as alegrias e me incentivando nos momentos mais desafiadores.

Ao meu orientador, Professor Victor Stroele. Sua orientação e paciência foram inestimáveis durante todo o processo de elaboração deste trabalho. Seus conselhos foram fundamentais para a qualidade e direcionamento da pesquisa.

A todos os professores do Departamento de Ciência da Computação. Suas aulas e ensinamentos foram de grande importância para o meu crescimento pessoal e profissional. Agradeço por compartilharem seu conhecimento e experiência.

Aos amigos e colegas que estiveram ao meu lado durante essa jornada. Suas discussões, apoio e encorajamento foram essenciais para superar os desafios e manter a motivação. Agradeço por compartilharem ideias, experiências e por serem uma fonte de inspiração constante.

Mais uma vez, expresso minha profunda gratidão a todos os mencionados. Sou imensamente grato por todo o suporte, orientação e conhecimento que recebi.

“O homem não teria alcançado o possível se, repetidas vezes, não tivesse tentado o impossível.”

Max Weber

Conteúdo

Lista de Figuras	7
Lista de Tabelas	8
Lista de Abreviações	9
1 Introdução	10
1.1 Apresentação do Tema	10
1.2 Contextualização	11
1.3 Problema	12
1.4 Justificativa	13
1.5 Objetivos	14
1.5.1 Objetivos Gerais	14
1.5.2 Objetivos Específicos	15
1.6 Metodologia	16
1.6.1 Caracterização da pesquisa	16
1.6.2 Coleta de Dados	17
1.6.3 Análise dos Dados	17
1.6.4 Validação dos resultados	18
2 Fundamentação Teórica	19
2.1 Ecossistema de Software	19
2.1.1 Definição	20
2.1.2 Dimensões	23
2.2 Sistemas De Reputação	26
2.2.1 Desenvolvimento Distribuído de Software	27
2.2.2 Reputação	29
2.2.3 Colaborador Chave	30
2.3 Considerações Finais	31
3 Trabalhos Relacionados	33
3.1 Managing and Monitoring Software Ecosystem to Support Demand and Solution Analysis	33
3.2 Um Modelo Dinâmico de Reputação para Apoiar a Manutenção Colaborativa de Software	35
3.3 Topological Analysis in Scientific Social Networks to Identify Influential Researchers	36
3.4 Ecosystems in GitHub and a Method for Ecosystem Identification Using Reference Coupling	37
3.5 Complex Network Analysis in a Software Ecosystem: Studying the Eclipse Community	38
3.6 sSECO-Process: Avaliando a Dimensão Social em Ecossistemas de Software	39
3.7 Análise comparativa	39
3.8 Considerações Finais	40

4	Processo para Análise de Colaboração no Ecossistema	41
4.1	Definição do process	41
4.2	Atividade: Configuração de Análise	42
4.3	Subprocesso: Coleta de Dados do Ecossistema	43
4.4	Subprocesso: Análise de Reputação dos Colaboradores	45
4.4.1	Distância de Levenshtein	46
4.4.2	Decaimento temporal por período de referência	47
4.4.3	Decaimento temporal por exponencial	48
4.4.4	Período de referência X Fator de decaimento exponencial	49
4.4.5	Normalização	50
4.5	Atividade: Elaboração gráfica e armazenamento	50
5	Avaliação: Prova de Conceito	52
5.1	Amostra de dados	52
5.2	Perspectiva do colaborador	54
5.3	Perspectiva do Projeto	59
5.4	Perspectiva do setor	65
5.5	Validação dos resultados	71
5.6	Considerações Finais do Capítulo	72
6	Considerações Finais e Trabalhos Futuros	74
6.1	Contribuição	74
6.2	Limitações	75
6.3	Trabalhos futuros	76
	Referências	78

Lista de Figuras

2.1	Modelo de domínio das relações entre os termos atualmente existentes em ECOS (BERK; JANSEN; LUINENBURG, 2010)	22
2.2	Visão Tridimensional do ECOS. Adaptado de (CAMPBELL; AHMED, 2010)	24
4.1	Arquitetura do processo de análise de reputação.	42
4.2	Configurações da análise	43
4.3	Subprocesso de coleta de dados	43
4.4	Diagrama de classe estrutural da plataforma	44
4.5	Subprocesso de análise de dados	45
4.6	Comparação entre decaimento exponencial e período de referência	49
5.1	Hierarquia organizacional	53
5.2	Análise da reputação do colaborador em um projeto	54
5.3	Análise da reputação do colaborador em múltiplos projetos	55
5.4	Comparação entre os valores de contribuição por log do colaborador 54e78	57
5.5	Comparação entre os valores de contribuição por log do colaborador e130e	58
5.6	Comparação entre os valores de contribuição por log do colaborador efc78 .	58
5.7	Resultados da análise de reputação no Projeto	59
5.8	Análise de reputação no projeto por período de referência	63
5.9	Análise de reputação no Projeto por fator exponencial	64
5.10	Resultado da análise de reputação dos colaboradores por setor	65
5.11	Diagrama de Venn de colaboradores por setor	66
5.12	Diagrama de Venn de colaboradores atuais do projeto	67
5.13	Análise de reputação do setor CRM com decaimento exponencial	69
5.14	Análise de reputação do setor <i>Billing</i> com decaimento exponencial	70
5.15	Análise de reputação do setor Integrador com decaimento exponencial . . .	70

Lista de Tabelas

3.1	Tabela comparativa dos trabalhos relacionados.	40
5.1	Tabela de reputação do colaborador em um único projeto	55
5.2	Tabela de reputação do colaborador em múltiplos projetos	56
5.3	Tabela de reputação normalizada do colaborador em múltiplos projetos	56
5.4	Tabela de dados de contribuição no projeto	61
5.5	Tabela de dados normalizada da contribuição no projeto	62
5.6	Tabela de dados normalizada da contribuição no setor CRM	68
5.7	Tabela de dados normalizada da contribuição por colaboradores ativos no setor CRM	68

Lista de Abreviações

API	Application Programming Interface
BI	Business Intelligence
BPM	Business Process Management
CDB	Component-Based Development
CRM	Customer Relationship Management
DCC	Departamento de Ciência da Computação
DDS	Desenvolvimento Distribuído de Software
ECM	Enterprise Content Management
ECOS	Ecosystema de Software
GUI	Graphical User Interface
RPP	Research Platforms and Portals
SE	Software Engineering
SPL	Software Product Line
TCC	Trabalho de Conclusão de Curso
TI	Tecnologia da informação
TTL	Team Tech Lead
UFJF	Universidade Federal de Juiz de Fora

1 Introdução

1.1 Apresentação do Tema

Tanto no âmbito profissional quanto no acadêmico, os desenvolvedores desempenham um papel central nos Ecossistemas de Software (ECOS), utilizando-os como principal meio de interação para adquirir novos conhecimentos, buscar suporte e contribuir com suas habilidades na resolução de problemas, manutenção e aprimoramento de projetos existentes (GUERCIO et al., 2018). Além disso, a interação contínua nos ECOS promove um ambiente de inovação, onde ideias e soluções são compartilhadas, aprimoradas e implementadas para melhorar a qualidade e eficácia dos projetos de software. Os benefícios desses ecossistemas vão muito além das características técnicas, abrangendo uma variedade ampla de vantagens. No entanto, é importante ressaltar que, apesar dos avanços nessa área, ainda há muito a ser explorado em relação aos dados dos ECOS. A análise mais aprofundada desses dados pode proporcionar uma compreensão mais abrangente do comportamento dos participantes, suas necessidades e preferências, bem como identificar oportunidades de otimização e crescimento nos processos de desenvolvimento de software.

Neste contexto, adota-se a definição de ECOS como comunidades que se formam dentro e/ou fora de uma organização, interagindo por meio de uma plataforma tecnológica comum, visando contribuir e trocar informações para soluções e serviços (CAMPBELL; AHMED, 2010; MANIKAS; HANSEN, 2013; SANTOS; VIANA, 2016; GUERCIO, 2018). Essas comunidades são compostas por uma diversidade de participantes, incluindo desenvolvedores, *stakeholders*, *software* e muito mais.

Um exemplo comum de ECOS é a plataforma *GitHub*¹, uma plataforma amplamente utilizada pelos desenvolvedores para publicar e versionar seus projetos, permitindo um ambiente de trabalho colaborativo e público. Através do *GitHub*, os desenvolvedores podem compartilhar código-fonte, colaborar em projetos conjuntos e acompanhar as alterações realizadas ao longo do tempo. Essa plataforma promove a interação entre os

¹(www.github.com)

participantes, incentivando o compartilhamento de conhecimentos e o desenvolvimento conjunto.

Outro exemplo relevante de ECOS é a plataforma *Stack Overflow*², uma comunidade online onde desenvolvedores, usuários e organizações podem interagir e fornecer suporte em uma ampla variedade de tópicos relacionados a linguagens de programação, ferramentas e software. Por meio do *Stack Overflow*, os participantes podem fazer perguntas, obter respostas, compartilhar soluções e colaborar para resolver problemas técnicos específicos. Essa plataforma cria um ambiente dinâmico de aprendizado e troca de informações, impulsionando o desenvolvimento dos ECOS e dos desenvolvedores que deles fazem parte.

Além disso, em sua grande maioria, os projetos não são desenvolvidos isoladamente, mas sim utilizando tecnologias e componentes pré-existentes. Essas tecnologias, quando aplicadas, podem evoluir e ser aplicadas em novos projetos, criando uma rede de interdependências que impulsiona o desenvolvimento do ecossistema e dos próprios desenvolvedores (BLINCOE; HARRISON; DAMIAN, 2015). Essa interconexão entre projetos, tecnologias e desenvolvedores fortalece o ecossistema como um todo, permitindo a inovação contínua e o progresso da área de engenharia de software.

1.2 Contextualização

O Ecossistema de Software pode ser sistematicamente mapeado em uma perspectiva tridimensional, que compreende as dimensões “Técnica”, “Negócios” e “Social” (BARBOSA et al., 2013). A dimensão Técnica (*Architectural*) concentra-se na plataforma, diretamente relacionada à evolução e arquitetura de software. A dimensão de negócio (*Business*) concentra-se no gerenciamento de conhecimento, ou seja, artefatos, recursos, informações, inovação e planejamento estratégico (compreendendo como, quando, onde e quem realiza determinadas tarefas). A dimensão social se concentra em como os *stackholders* interagem, entendem e modificam o conhecimento em um ecossistema de software (CAMPBELL; AHMED, 2010).

Devido à crescente adoção de ECOS, como os exemplos já citados (GitHub e Stack

²(www.stackoverflow.com)

Overflow), a pesquisa tem se concentrado cada vez mais nesses ecossistemas, buscando extrair informações valiosas de seus vastos bancos de dados. Esses estudos visam explorar diversas áreas, como a detecção de especialistas, a análise dos relacionamentos entre projetos, a busca por melhores soluções, a evolução da arquitetura, as tecnologias aplicadas, os sistemas de reputação e as estatísticas das linguagens de programação (BADASHIAN et al., 2014; BLINCOE; HARRISON; DAMIAN, 2015; SILVA, 2018; ARAKAKI, 2016).

Essas pesquisas enfrentam desafios significativos, especialmente em relação à complexidade da coleta, tratamento e processamento de dados. Em plataformas abertas, a quantidade de dados disponíveis é imensa, demandando estratégias eficientes para lidar com essa vasta quantidade de dados. Já em ambientes profissionais e privados das empresas, a obtenção de amostras representativas e acessíveis é um desafio adicional, por se tratarem de dados sensíveis e sigilosos. A complexidade desses processos de coleta e tratamento de dados tem sido uma das principais dificuldades enfrentadas pelos pesquisadores (LÉLIS, 2017).

No entanto, os esforços empregados nesses estudos valem a pena, pois os ativos de conhecimento que podem ser derivados desses dados são de imenso valor. A análise e a validação desses ativos têm se mostrado uma área de pesquisa com grande potencial, oferecendo oportunidades de compreender e aproveitar os *insights* provenientes desses ecossistemas de software. Ao superar os desafios e avançar nessa área, é possível obter informações valiosas para aprimorar a prática profissional, impulsionar a inovação tecnológica e fortalecer a tomada de decisões no desenvolvimento de software.

1.3 Problema

As empresas que atuam no setor de tecnologia da informação e desenvolvimento de software estão constantemente em busca de métodos eficientes para gerenciar suas equipes. Para isso, é essencial aplicar estratégias que permitam mensurar o progresso dos projetos e o desempenho dos colaboradores. Atualmente, as metodologias ágeis têm sido amplamente adotadas visando aprimorar e estabelecer uma cadência eficiente no processo de desenvolvimento. No entanto, ainda há uma carência no que diz respeito ao aprimoramento do planejamento estratégico e gestão de pessoal.

À medida que as estratégias de coleta e processamento de dados no ecossistema de software continuam a ser exploradas em plataformas abertas, surgem questionamentos sobre a validação e eficácia dos processos e *frameworks* propostos por pesquisas anteriores, quando aplicados em ambientes internos e específicos das empresas, com foco na perspectiva da dimensão de negócios do ecossistema (BLINCOE; HARRISON; DAMIAN, 2015).

Encontrar um ambiente de teste adequado para implementar essas abordagens é uma tarefa complexa, uma vez que as informações armazenadas em bancos de dados podem acarretar em perdas financeiras significativas caso não sejam tratadas e protegidas de forma adequada. Isso dificulta a obtenção de uma amostra representativa de dados que as empresas possam disponibilizar para pesquisa. Além disso, cada ambiente empresarial apresenta fatores de adaptação únicos, o que aumenta a complexidade do processamento e análise dos dados.

Superar esses desafios exige um cuidadoso equilíbrio entre a busca por *insights* valiosos a partir dos dados disponíveis e a garantia da segurança e privacidade das informações sensíveis. A capacidade de adaptar os processos e *frameworks* existentes às particularidades de cada ambiente empresarial é fundamental para extrair resultados confiáveis e aplicáveis à realidade das organizações.

1.4 Justificativa

Ao superar as dificuldades citadas anteriormente, abre-se um caminho promissor para aprimorar o planejamento estratégico e a gestão de pessoal no contexto dos ecossistemas de software internos das empresas, impulsionando a eficiência e o sucesso dos projetos de desenvolvimento.

Uma das etapas fundamentais desse processo é uma metodologia capaz de identificar os colaboradores chave e avaliar sua reputação dentro do ECOS. Essa identificação é de extrema importância para o gerenciamento interno das empresas de desenvolvimento de software (BOEHM, 2006). Uma substituição mal planejada ou a perda de um colaborador estratégico pode impactar negativamente no andamento de um projeto, aumentando a curva de aprendizado da equipe e prolongando o prazo de entrega.

A partir dos dados obtidos da análise desse ambiente, busca-se estimar o desempenho, a eficiência e a *expertise* de cada colaborador, além de identificar áreas carentes de aprendizado. Essas informações são valiosas para auxiliar no processo de remuneração, no planejamento estratégico para formação e manutenção de equipes, na estimativa do ritmo de trabalho de uma equipe para determinar a necessidade de reforços, entre outros aspectos relevantes (GUERCIO, 2018).

A abordagem mencionada apresenta uma ampla gama de benefícios, em grande parte, ainda inexplorados. Seu desenvolvimento e aplicação podem fornecer uma base sólida para a solução de problemas desconhecidos no contexto dos ecossistemas internos das empresas de desenvolvimento de software. Para isso, é essencial realizar validações e pesquisas específicas no ambiente interno de cada empresa, considerando suas características e particularidades. Essa abordagem pode abrir novas perspectivas e trazer vantagens significativas para a gestão de recursos humanos e o sucesso dos projetos de desenvolvimento de software.

1.5 Objetivos

1.5.1 Objetivos Gerais

O objetivo deste trabalho é propor um processo para a coleta e análise das amostras obtidas em um ambiente interno de uma empresa privada de desenvolvimento de software. O propósito deste processo é possibilitar um estudo detalhado dos colaboradores envolvidos em projetos específicos, a fim de identificar os principais contribuidores em cada área e compreender o impacto de sua ausência no projeto. Além disso, busca-se identificar as especialidades de cada colaborador, avaliando suas competências técnicas e áreas de conhecimento específicas. Informações valiosas para o planejamento estratégico das equipes, a alocação eficiente de recursos e a identificação de oportunidades de desenvolvimento profissional.

Ao utilizar dados reais de um ambiente profissional específico, este trabalho contribuirá para a validação e aplicação prática das metodologias propostas, fornecendo informações valiosas para o gerenciamento de equipes de desenvolvimento de software.

Além disso, ao envolver os colaboradores da pesquisa por meio da coleta de *feedback*, será possível obter *insights* adicionais para a identificação de melhorias e aperfeiçoamentos nas metodologias propostas. Essa abordagem colaborativa e participativa contribuirá para a validação mais robusta dos resultados obtidos, uma vez que os próprios profissionais envolvidos no ambiente de desenvolvimento poderão oferecer perspectivas valiosas e experiências práticas para avaliar a efetividade das metodologias propostas.

1.5.2 Objetivos Específicos

As pesquisas atuais sobre a análise de dados de Ecossistemas de Software em plataformas abertas têm apresentado avanços promissores. Um objetivo específico deste trabalho é propor e validar metodologias relacionadas às propostas presentes em estudos anteriores, em um ambiente interno de uma empresa privada de desenvolvimento de software. Através da análise das informações obtidas nesse ambiente, busca-se obter *insights* relevantes e promover a discussão sobre a relação dos resultados alcançados com os *feedbacks* dos colaboradores da pesquisa, visando contribuir na relação entre a dimensão social e de negócio do ecossistema da empresa. Essa abordagem visa identificar semelhanças com a realidade do ambiente de desenvolvimento profissional, bem como diferenças, falhas e oportunidades de melhoria que possam ser identificadas e validadas. Ao promover essa discussão e incorporar os *feedbacks* dos colaboradores, é possível enriquecer os resultados da pesquisa e fornecer uma visão mais abrangente e contextualizada das metodologias propostas.

A amostra de dados para testes foi obtida de um ambiente real de desenvolvimento de software, mais especificamente das equipes de desenvolvedores que atuam na plataforma *SYDLE ONE*³. A escolha da plataforma *SYDLE ONE* como fonte de dados para este estudo é fundamentada em seu amplo escopo de funcionalidades, incluindo automação de processos (BPM), gerenciamento de documentos (ECM), gestão e relacionamento (CRM), *e-commerce*, ciência de dados, *billing* entre outros, que possibilita o desenvolvimento de projetos abrangentes e diversificados. Uma empresa que utiliza essa plataforma concordou em disponibilizar parte de seus dados para a análise proposta. A

³(www.sydle.com)

obtenção de informações das equipes que trabalham na plataforma *SYDLE ONE* permite a coleta de dados representativos e relevantes, essenciais para a realização da análise proposta neste estudo. A colaboração da empresa em ceder esses dados contribui para a obtenção de uma amostra significativa, enriquecendo assim a validade e a robustez dos resultados obtidos.

1.6 Metodologia

1.6.1 Caracterização da pesquisa

Este trabalho adota uma vertente metodológica de natureza quantitativa, que busca coletar dados e estatísticas reais para a validação de modelos relacionados à gestão de colaboradores em ecossistemas de software. A escolha por uma abordagem quantitativa permite a obtenção de resultados mensuráveis e objetivos, fornecendo uma base sólida para a análise e avaliação das estratégias propostas.

Durante a fase inicial, foram empreendidos estudos aprofundados acerca dos fundamentos de ECOS e da sua literatura correspondente. Posteriormente, foram empregados alguns dos modelos identificados em trabalhos relacionados para extrair questões de relevância no contexto da gestão de pessoas. Nesse sentido, foi elaborada uma especificação de um processo embasado nessas pesquisas, composto por atividades principais, sendo: filtragem e coleta de dados, utilização de modelos semânticos, análise dos resultados obtidos e validação dos mesmos. Cada uma dessas atividades desempenha um papel crucial na estruturação e aplicação eficiente do processo em questão.

A combinação de uma abordagem quantitativa, a utilização de um processo consistente e a aplicação em projetos reais centralizados em uma plataforma específica aumentam a robustez e a aplicabilidade dos resultados obtidos. Isso contribui não apenas para o avanço do conhecimento acadêmico nessa área, mas também para a implementação prática de estratégias eficazes de gerenciamento de colaboradores em Ecossistemas de Software.

1.6.2 Coleta de Dados

A etapa inicial do processo compreende a obtenção dos dados, no qual foi realizado um estudo no ambiente de uma empresa real de desenvolvimento de software, denominada aqui como “empresa parceira” para garantir a preservação de seu anonimato. A amostra de dados é abrangente e inclui *logs* (registros) de desenvolvimento de diferentes colaboradores na mesma plataforma. Esses *logs*, semelhantes aos *commits* da plataforma GitHub, contêm informações sobre o conteúdo modificado, o autor, a data e o setor de atuação (como *billing*, CRM, BPM, *E-commerce*, etc.).

Para coletar os dados, foi realizada uma busca abrangente utilizando o mecanismo de busca *ElasticSearch* no banco de dados da plataforma. Em seguida, a amostra foi disponibilizada por meio de um *endpoint API*.

Durante esse processo, foi adotado um tratamento adequado para transformar ou remover informações confidenciais e pessoais relacionadas à empresa e aos seus colaboradores, a fim de preservar a privacidade e a segurança dos envolvidos.

Além disso, foi realizada uma filtragem dos dados com base nas necessidades e características específicas do projeto em questão. Essa filtragem permitiu selecionar os dados relevantes e adequados, conforme os critérios estabelecidos, a fim de obter uma amostra mais precisa e direcionada para a análise.

1.6.3 Análise dos Dados

Após a fase de coleta de dados, foi conduzida uma análise detalhada utilizando métodos de identificação para o Sistema de Reputação, baseando-se nos estudos de Guercio (2018), Guercio et al. (2017), Lélis (2017), Hendrikx, Bubendorfer e Chard (2015). O objetivo principal nessa etapa é identificar os colaboradores-chave em cada projeto, suas áreas de especialização e o impacto da sua ausência no desenvolvimento do projeto.

A partir dos resultados obtidos, busca-se fomentar uma discussão abrangente sobre a aplicabilidade do processo e das metodologias em diferentes contextos, a fim de aprimorar o conhecimento sobre a gestão de ECOS em ambientes diversos. Essa análise comparativa possibilita identificar possíveis adaptações necessárias para utilizar essas abordagens em ambientes restritos, como empresas privadas, e validar sua eficácia

na identificação da reputação dos colaboradores.

Essa avaliação e validação das abordagens utilizadas contribuem para a consolidação do conhecimento na área, fornecendo diretrizes práticas e embasamento teórico para a gestão de colaboradores chave em ambientes profissionais e acadêmicos, impulsionando o desenvolvimento e a evolução contínua da engenharia de software.

1.6.4 Validação dos resultados

Os resultados obtidos foram apresentados aos colaboradores e gestores envolvidos na pesquisa, com o objetivo de avaliar e validar a aplicabilidade do processo proposto, por meio de uma comparação com o *feedback* das características individuais dos colaboradores e do projeto em questão. Além de identificar possíveis melhorias e ajustes necessários para atender às demandas e particularidades dos diferentes contextos organizacionais.

Essa avaliação das abordagens utilizadas desempenham um papel fundamental na consolidação do conhecimento na área, proporcionando diretrizes práticas e embasamento teórico para a gestão de colaboradores-chave em ambientes profissionais e acadêmicos. Essas informações contribuem para impulsionar o desenvolvimento e a contínua evolução da engenharia de software, permitindo avanços significativos na compreensão e na aplicação efetiva do ECOS na gestão de pessoas em projetos de desenvolvimento de software.

2 Fundamentação Teórica

Neste capítulo, serão apresentados os tópicos que constituem a base teórica deste trabalho. Iniciaremos com a exploração do conceito de ecossistema de software, fornecendo sua definição, conceitos básicos, origens e dimensões relevantes. Em seguida, será abordado o sistema de reputação, juntamente com a definição do Desenvolvimento Distribuído de Software, onde serão discutidos os colaboradores-chave, suas especialidades e reputação dentro do ecossistema, visando o gerenciamento eficiente da empresa. Por fim, serão apresentadas as considerações finais deste capítulo, resumindo os principais pontos abordados e estabelecendo uma transição adequada para os próximos capítulos do trabalho.

2.1 Ecossistema de Software

No contexto atual, as empresas e fornecedores de *software* estão cada vez mais abrindo seus negócios para interações com outras empresas de *software*, resultando em sua integração em um ecossistema (ECOS) junto a desenvolvedores e parceiros. No entanto, esses atores, especialmente os fornecedores de *software* que ocupam uma posição central no ecossistema, enfrentam dificuldades em compreender plenamente o ECOS e como suas ações afetam o desempenho e a evolução desse ambiente (BERK; JANSEN; LUINENBURG, 2010).

Para melhor compreender e modelar o ECOS, é necessário determinar suas características formais, como tamanho, meios de subsistência, presença de padrões ou organizações de padrões, além dos diferentes papéis que os fornecedores de software podem desempenhar nesse contexto. No entanto, o desafio de modelagem reside no fato de que atualmente não existem formalismos de modelagem bem estabelecidos para ECOS (BRINKKEMPER; SOEST; JANSEN, 2009). Diante dessa realidade, as empresas de software precisam considerar seu papel estratégico no ecossistema para garantir sua sobrevivência, enfrentando diversos desafios de pesquisa tanto em níveis técnicos quanto de negócios (JANSEN; FINKELSTEIN; BRINKKEMPER, 2009).

Esta seção pretende introduzir conceitos básicos sobre o ECOS, suas origens,

as propostas de modelagem existentes e suas dimensões fundamentais. Por meio dessa exploração, será possível estabelecer uma base sólida para o entendimento do ECOS e para abordar os desafios associados a esse ambiente dinâmico e complexo.

2.1.1 Definição

O estudo do ECOS na comunidade de Engenharia de Software foi inicialmente motivado pela evolução de uma linha de Produtos de Software (*Software Product Line* - SPL) e Reuso de Software, visando permitir que desenvolvedores externos contribuíssem para plataformas que, até então, eram fechadas em uma indústria global de software impulsionada pelo paradigma de Desenvolvimento Baseado em Componentes (*Component-Based Development* - CBD) (BOSCH, 2009; SANTOS, 2016).

Um ecossistema de software (ECOS) é um conjunto de atores e artefatos, tanto dentro como fora de uma organização ou comunidade, que trocam recursos e informações em torno de uma plataforma de tecnologia comum (JANSEN; FINKELSTEIN; BRINK-KEMPER, 2009). Esse contexto influencia as decisões de gerenciamento e desenvolvimento dessas plataformas, especialmente no que diz respeito a modelos de arquitetura, governança e colaboração em diversos domínios de aplicação. É essencial integrar mecanismos e ferramentas que apoiem a troca de informações, recursos e artefatos, garantindo a comunicação e interação efetiva entre desenvolvedores e usuários (SANTOS; VIANA, 2016). Em essência, os atores podem englobar qualquer entidade envolvida ou com interesse no ecossistema.

No ECOS, uma plataforma é o espaço onde os produtos e serviços fornecidos podem ser incluídos, modificados ou estendidos como artefatos de software (SANTOS, 2016). Dentro de um ecossistema de software, cada ator é impulsionado por um conjunto específico de interesses e modelos de negócio, estabelecendo conexões e relacionamentos com outros atores e com o próprio ECOS (MANIKAS; HANSEN, 2013). Os atores podem assumir diferentes formas, como empresas, organizações, setores de uma empresa, usuários finais de produtos de software, fornecedores, clientes e outras partes interessadas (LIMA, 2015).

Essa diversidade de atores no ecossistema reflete a complexidade e a variedade de

interações que ocorrem entre os diversos participantes. Cada ator possui seus próprios objetivos, motivações e estratégias, o que influencia suas decisões e a forma como interagem com os demais atores. Essas interações podem ser de natureza colaborativa, competitiva ou complementar, dependendo das relações estabelecidas e dos interesses em jogo. Segundo Iansiti e Levien (IANSITI; LEVIEN, 2004), existem diferentes papéis dentro de um Ecossistema de Software, tais como *keystones*, *dominators* e *niche players*.

Keystone é a organização que lidera uma plataforma ECOS cria e compartilha valor com o resto do ecossistema, desempenhando um papel fundamental na sua estrutura e funcionamento. Responsáveis por fornecer os principais componentes e/ou serviços que constituem a base da plataforma. Geralmente, exercem a maior influência sobre o ecossistema, sendo capazes de estabelecer diretrizes e padrões que orientam as interações e desenvolvimento dentro do ecossistema (IANSITI; LEVIEN, 2004).

Dominator extrai o máximo valor do ecossistema ao fornecer uma plataforma alternativa, colocando em risco a sua saúde e sustentabilidade.

Niche Players são os *stackholders* que agem para desenvolver ou aprimorar capacidades especializadas que os diferenciam de outros *stackholders*, alavancando os recursos da rede enquanto ocupam apenas uma pequena parte da própria rede. Individualmente, podem influenciar, comprometer-se, contribuir, promover ou estender a plataforma e juntos compõem a maioria do ecossistema em massa total e variedade. Em (HAGEL; BROWN; DAVISON, 2008), são subdivididos entre:

- *Influencer*: comprometem-se cedo e proeminentemente desenvolvendo para o ECOS e contribui para sua saúde ao se comprometer com uma estratégia de modelagem, complementando o keystone.
- *Hedger*: desenvolvem seus produtos ou serviços para suportar múltiplas plataformas de modelagem.
- *Disciple*: se comprometem exclusivamente com a plataforma de um ECOS.

Para o ECOS, analogamente ao ecossistema animal, *health* são todos os meios que um *hub* tem para melhorar a saúde de seu ecossistema, incluindo níveis de produtividade, robustez e criação de nichos (RAPPORT; COSTANZA; MCMICHAEL, 1998). Essas

características do ecossistema de negócios e seus relacionamentos podem ser observados no modelo da Figura 2.1.

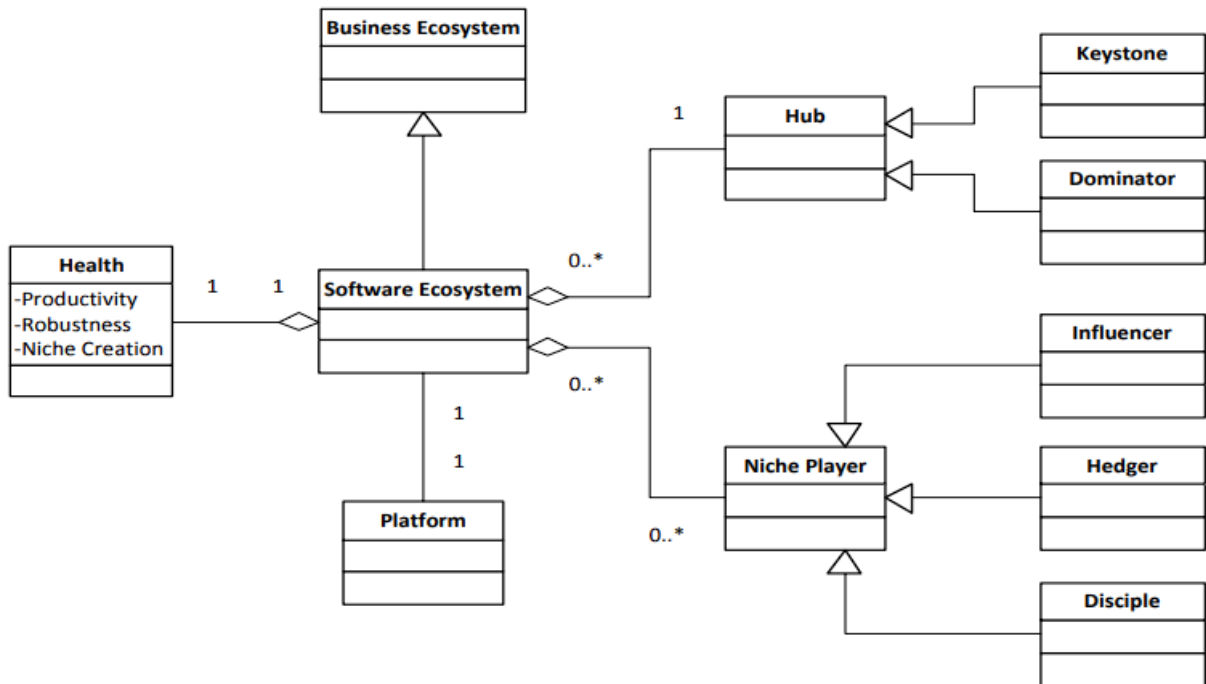


Figura 2.1: Modelo de domínio das relações entre os termos atualmente existentes em ECOS (BERK; JANSEN; LUINENBURG, 2010)

Em relação às definições mencionadas, um termo de importância fundamental é a plataforma do ecossistema. Uma plataforma de ecossistema refere-se a um conjunto de soluções disponibilizadas aos membros do ecossistema por meio de pontos de acesso ou interfaces específicas (IANSITI; LEVIEN, 2004). No contexto de um ECOS, a plataforma é frequentemente representada por um produto de software, mas também pode envolver bibliotecas de software que os “*niche players*” podem utilizar para extrair e criar valor (BERK; JANSEN; LUINENBURG, 2010).

A plataforma do ecossistema desempenha um papel central na facilitação da colaboração, inovação e interconexão entre os atores do ECOS. Ela fornece um ambiente comum no qual os participantes podem desenvolver e integrar suas soluções, compartilhar recursos, trocar informações e explorar oportunidades de negócio. Através da plataforma, os “*niche players*” podem aproveitar os recursos e funcionalidades disponibilizados pelos “*keystones*” e “*dominators*”, alavancando-os para criar valor adicional e atender às necessidades específicas de seus segmentos de mercado. A plataforma também desempenha um

papel importante na criação de um mercado sustentável e dinâmico, incentivando a participação ativa dos atores e promovendo a evolução. Ao fornecer um conjunto de interfaces padronizadas e ferramentas de desenvolvimento, a plataforma facilita a interoperabilidade e a integração de soluções, reduzindo as barreiras para a entrada de novos participantes e incentivando a colaboração entre os atores existentes.

2.1.2 Dimensões

As empresas de software estão constantemente aprimorando suas operações de negócios em várias áreas, como tecnologia, gerenciamento e processos de desenvolvimento de produtos, visando capturar uma maior participação de mercado e aumentar sua lucratividade. No ambiente competitivo atual, a chave para o sucesso empresarial está na inovação, e as organizações estão continuamente buscando a adoção de práticas inovadoras em áreas-chave de suas operações.

O conceito ECOS surge como uma abordagem transitória, evolucionária e inovadora, derivada da arquitetura geral e da teoria do desenvolvimento de redes sociais, evoluindo o conceito de Linha de Produtos de Software. Para gerenciar a diversidade de organizações e relacionamentos criados em torno do ECOS por programadores, fornecedores, parceiros e clientes/usuários, foi desenvolvido uma visão tridimensional para representar essa complexidade do ECOS (BARBOSA et al., 2013; CAMPBELL; AHMED, 2010) representado pela Figura 2.2.



Figura 2.2: Visão Tridimensional do ECOS. Adaptado de (CAMPBELL; AHMED, 2010)

Essencialmente, todas as três dimensões são fortemente integradas por meio do processo de engenharia de software, amplamente aceitas pela literatura e adotadas em trabalhos subsequentes (GUERCIO, 2018), permitindo uma compreensão mais abrangente e aprofundada de sua complexidade, possibilitando o desenvolvimento de estratégias eficazes de gestão e inovação. Ao considerar as interações entre as dimensões, as organizações podem potencializar as oportunidades de negócio, promover a colaboração e a sinergia entre os atores do ecossistema, e impulsionar a inovação contínua no desenvolvimento de software.

Dimensão Técnica

Também definida como dimensão da arquitetura, possui foco na plataforma ECOS, ou seja, concentra-se no mercado, tecnologia, infraestrutura ou organização que compõem o ecossistema. Ela é abordada por meio do processo de engenharia de domínio da plataforma, que estabelece o ciclo de vida da plataforma, gerencia a similaridade e a variabilidade (definindo recursos) e desenvolve uma arquitetura de Linha de Produtos de Software (SPL) para a plataforma (SANTOS; WERNER, 2011).

Essa dimensão está mais relacionada aos níveis de escopo organizacional e Redes de Fornecimento de Software (Software Supply Networks - SSN) em uma perspectiva interna, em vez de abordar o escopo do ECOS em uma visão externa. Seu foco principal é compreender como a Engenharia de Software (SE) é aplicada no design, desenvolvimento e manutenção da plataforma, considerando a natureza específica da plataforma como uma SPL (BERK; JANSEN; LUINENBURG, 2010; SANTOS; WERNER, 2011).

É essencial estabelecer práticas eficazes de engenharia de domínio, identificando as características comuns e as variações dentro do ecossistema (KIPPER, 2010). A engenharia de domínio é uma proposta surgida no âmbito da Engenharia de Software que visa o aumento do reuso no desenvolvimento de sistemas, buscando uma redução dos custos e do tempo de desenvolvimento (KIPPER, 2010). Isso permite a definição de uma arquitetura flexível e modular, capaz de atender às demandas específicas dos atores do ecossistema. A adoção de abordagens de SPL permite a reutilização de componentes e a rápida adaptação da plataforma às mudanças tecnológicas e de mercado, fornecendo uma base sólida para o desenvolvimento e a evolução contínua do ecossistema de software.

Dimensão Social

A dimensão social concentra-se no equilíbrio entre a proposição e a realização da utilidade à medida que os stakeholders se integram, ampliam e modificam o conhecimento e interagem no ecossistema. Seu objetivo é avaliar a promoção do ECOS, reconhecendo de forma implícita e explícita as capacidades e o engajamento dos stakeholders. Além disso, busca o conhecimento sobre quais colaborações, desenvolvimento de código e outras oportunidades de redes sociais contribuem para as partes interessadas (SANTOS; WERNER, 2011).

Essa dimensão reconhece a importância dos stakeholders no sucesso do ECOS e destaca a necessidade de estabelecer relações sólidas e colaborativas entre eles. Ela envolve a identificação dos diferentes atores do ecossistema, como desenvolvedores, fornecedores, parceiros e clientes, e a compreensão de seus interesses, necessidades e expectativas. Ao promover a interação e a troca de conhecimentos entre os stakeholders, é possível fortalecer a colaboração, impulsionar a inovação e maximizar o valor gerado pelo ecossistema na

totalidade.

Dimensão de negócios

A dimensão de negócios do ECOS concentra-se no fluxo de conhecimento dentro da organização, abrangendo artefatos, recursos e informações. Seu objetivo é estabelecer metas e planos de ação por meio de programas e projetos, impulsionando a inovação e conectando o ECOS ao mercado. Essa dimensão busca entender como, quando, onde e por quem as metas serão cumpridas para o planejamento estratégico efetivo (SANTOS; WERNER, 2011).

Nessa dimensão, é essencial gerenciar o fluxo contínuo de conhecimento, tanto dentro da organização como em sua interação com o ecossistema. Isso envolve o desenvolvimento e compartilhamento de artefatos e informações relevantes para a criação de valor e o avanço das atividades do ECOS. Além disso, é necessário estabelecer metas claras e definir planos de ação alinhados com a visão estratégica da organização, visando impulsionar a inovação e promover a competitividade no mercado. Isso envolve a análise de tendências, demandas do mercado e necessidades dos clientes, para serem estabelecidas estratégias adequadas de entrada e crescimento. É importante entender como as inovações podem ser integradas ao ecossistema, identificar os recursos necessários para sua implementação e determinar os responsáveis por executar as metas e planos estabelecidos.

2.2 Sistemas De Reputação

As organizações fazem investimentos significativos em seus sistemas de software, que se tornaram ativos estratégicos para os negócios. No entanto, para manter o valor desses ativos, é necessário realizar atualizações e melhorias constantes. Nesse contexto, e visando manter a produtividade e a qualidade, as empresas têm adotado práticas de distribuição de suas atividades (LÉLIS, 2017).

Nesta seção, abordaremos o conceito de Desenvolvimento Distribuído de Software, que é uma abordagem adotada atualmente pelas empresas, permitindo uma maior agilidade e eficiência na entrega de produtos e serviços de software. Também será introduzido o conceito de colaborador chave, que desempenha um papel crucial no sucesso

dos projetos de desenvolvimento de software. Um colaborador chave é aquele que possui conhecimento especializado em determinada área, desempenha um papel estratégico na equipe, com um impacto significativo no resultado do projeto. A identificação dos colaboradores chave pode ser facilitada por meio de um sistema de reputação, que avalia e classifica a contribuição e o desempenho dos indivíduos com base em critérios estabelecidos.

2.2.1 Desenvolvimento Distribuído de Software

O ambiente empresarial atual, impulsionado pelo processo de globalização, tem apresentado desafios significativos para o desenvolvimento de software. Esse processo de globalização tem levado a uma crescente fragmentação e globalização do desenvolvimento de software, o que aumenta a complexidade das atividades de construção de sistemas de informação. Como resultado, o campo do desenvolvimento de software tem se tornado cada vez mais interdisciplinar, exigindo o uso de conceitos e práticas provenientes de diversas áreas. Nesse contexto, surge a abordagem de Desenvolvimento Distribuído de Software (DDS), na qual recursos humanos geograficamente distribuídos, como programadores, designers e clientes, colaboram em suas atividades durante o processo de desenvolvimento (AUDY, 2007).

Existem três fatores principais que criam um ambiente propício para o desenvolvimento distribuído de software, todos relacionados à evolução do ambiente de negócios. Em primeiro lugar, a globalização impulsiona a necessidade de colaboração em equipes distribuídas geograficamente, aproveitando talentos e recursos em diferentes localidades. Em segundo lugar, a crescente importância dos sistemas de informação nas empresas demanda o desenvolvimento de soluções cada vez mais complexas, o que requer a colaboração de especialistas de diversas áreas. Por fim, a terceirização de processos, conhecida como *outsourcing*, tem sido adotada por muitas empresas como uma estratégia para aumentar a eficiência e reduzir os custos, levando à distribuição geográfica das atividades de desenvolvimento (AUDY, 2007).

O desenvolvimento de software é uma atividade complexa que enfrenta vários desafios, os quais podem ser abordados a partir de diferentes perspectivas. Esses desafios

podem ser agrupados em três categorias principais:

- **Pessoas:** relacionados às características que afetam diretamente os recursos humanos envolvidos no desenvolvimento de software. Isso inclui aspectos como a capacitação e qualificação dos profissionais, a coordenação e colaboração efetiva entre as equipes, a motivação e engajamento dos indivíduos, a produtividade e o trabalho em equipe (AUDY, 2007). Superar esses desafios requer a implementação de práticas de gestão de pessoas eficientes e o estabelecimento de uma cultura colaborativa e de alto desempenho.
- **Processos:** relacionados à forma como o projeto de software é desenvolvido. Incluem aspectos como a análise de custo-benefício, a especificação adequada dos requisitos, a aplicação de abordagens de Desenvolvimento Distribuído de Software (DDS), a gestão eficaz do projeto, a identificação e gerenciamento de riscos, a gestão do conhecimento, a manutenção de software, o planejamento organizacional, a estimativa de prazos e recursos, a reutilização de componentes e o controle da qualidade e teste do software (AUDY, 2007). Superar esses desafios requer a adoção de metodologias de desenvolvimento eficientes, o uso de ferramentas adequadas e a implementação de processos bem definidos.
- **Tecnologia:** Esses desafios envolvem o uso de diferentes ferramentas e equipamentos técnicos que podem servir como suporte ou base para o desenvolvimento de software. Incluem aspectos como a seleção e utilização de ferramentas de apoio ao desenvolvimento, a infraestrutura de comunicação e colaboração, e a adoção de tecnologias emergentes que possam melhorar a eficiência e qualidade do processo de desenvolvimento (AUDY, 2007). Superar esses desafios requer o conhecimento e a aplicação adequada das tecnologias disponíveis, bem como o acompanhamento constante das inovações tecnológicas relevantes para a área.

A compreensão e superação desses desafios são essenciais para o sucesso do desenvolvimento distribuído de software. Ao enfrentar esses desafios, as organizações podem aproveitar os benefícios da colaboração global, como a disponibilidade de talentos especializados em diferentes partes do mundo, a redução de custos e a aceleração do tempo

de desenvolvimento. No entanto, é importante adotar abordagens e práticas adequadas para lidar com as complexidades inerentes ao desenvolvimento distribuído de software e garantir a eficácia e a qualidade dos resultados alcançados.

A colaboração é essencial para o bom funcionamento das atividades distribuídas das equipes de manutenção que desejam colaborar e devem se organizar e focar em elementos como cooperação, comunicação e coordenação (FUKS; RAPOSO; GEROSA; LUCENA, 2003; LÉLIS, 2017). Para colaborar, os indivíduos devem trocar informações (comunicação), organizar (coordenação) e operar juntos em um espaço compartilhado (cooperação) (FUKS; RAPOSO; GEROSA; LUCENA, 2003). A reputação se torna um elemento chave, afetando estes elementos de colaboração (LÉLIS, 2017).

2.2.2 Reputação

No contexto do ECOS, é de grande importância explorar a disponibilidade de informações sobre a reputação dos desenvolvedores, uma vez que esse é um aspecto crucial para aprimorar a colaboração em equipes distribuídas e colaborativas de desenvolvimento (LÉLIS, 2017). Disponibilizar recursos e mecanismos de visualização para apoiar a análise dessas informações tem o potencial de melhorar a descoberta e identificação de dados relevantes para os gestores tomarem decisões ao alocar tarefas e responsabilidades entre os membros da equipe (LÉLIS, 2017; HENDRIKX; BUBENDORFER; CHARD, 2015).

A reputação pode ser entendida como uma medida coletiva de confiabilidade, baseada em *feedback* ou classificações fornecidas pelos membros de uma comunidade. No caso de um desenvolvedor desempenhando tarefas de manutenção, sua reputação seria construída com base nas percepções geradas por suas atividades anteriores. Os dados passados fornecem a reputação atual do desenvolvedor, e é importante ressaltar que essa reputação não é estática e pode evoluir ao longo do tempo, refletindo a evolução e aprimoramento das habilidades e desempenho do profissional (LÉLIS, 2017), assim como também pode ser afetada negativamente. Nesse sentido, os sistemas de reputação desempenham um papel importante ao coletar, agregar e distribuir informações relevantes sobre a reputação dos desenvolvedores.

Os sistemas de reputação fornecem uma estrutura organizada, permitindo que os

gestores tenham uma visão completa do desempenho e da confiabilidade de cada desenvolvedor. A reputação é normalmente representada por uma pontuação ou classificação, que reflete a percepção coletiva da comunidade em relação a essa entidade, quanto mais positivos, maior será a reputação da entidade. Uma das principais vantagens dos sistemas de reputação é a transparência que eles proporcionam. Um Sistema de Reputação pode ser avaliado com base em dois aspectos técnicos principais: a arquitetura de rede e o modelo do motor de reputação (HENDRIKX; BUBENDORFER; CHARD, 2015).

A arquitetura de rede de um sistema de reputação estabelece a estrutura e o fluxo de comunicação entre os participantes envolvidos no processo de reputação. Ela define como as interações e avaliações ocorrem, como os *feedbacks* são coletados e como os dados são armazenados e compartilhados. Essa arquitetura pode ser centralizada, onde um servidor central é responsável por coletar e distribuir as informações de reputação, ou pode ser descentralizada, onde a coleta e o compartilhamento de informações são distribuídos entre os participantes (JØSANG; ISMAIL; BOYD, 2007).

O motor de reputação, por sua vez, desempenha um papel fundamental no processamento e cálculo da reputação com base nos dados coletados. Ele implementa algoritmos e métricas específicas para atribuir pesos aos *feedbacks* e calcular a reputação de cada indivíduo. O motor de reputação pode considerar diferentes fatores, como a confiabilidade dos avaliadores, o tempo e a relevância das interações, a fim de gerar uma medida confiável e representativa da reputação de um indivíduo (JØSANG; ISMAIL; BOYD, 2007).

Outro aspecto a ser considerado é o envelhecimento das qualificações de um indivíduo. Pode ser apropriado atribuir menos peso a qualificações mais antigas no cálculo da reputação. Isso evita, por exemplo, que os indivíduos sejam penalizados com índices baixos mesmo após terem modificado sua conduta (JØSANG; QUATTROCIOCCI, 2009).

2.2.3 Colaborador Chave

Neste estudo, será adotado o termo "Colaborador Chave" para se referir aos profissionais que possuem especialização e influência em um setor específico de atuação. Esses colaboradores desempenham um papel crucial no funcionamento de uma equipe ou organização,

trazendo conhecimentos técnicos aprofundados, experiência relevante e habilidades específicas.

A identificação e análise desses Colaboradores Chave proporcionam uma compreensão mais precisa das competências e expertise presentes na equipe. Esses dados valiosos podem ser aplicados de diversas maneiras, como no desenvolvimento de metodologias para gerenciamento e manutenção de equipes. Ao reconhecer os Colaboradores Chave, é possível direcionar adequadamente os recursos, atribuir responsabilidades eficientemente e promover uma distribuição equilibrada do conhecimento dentro da equipe.

Essas informações também podem ser utilizadas para aprimorar os processos de tomada de decisão, permitindo que os gestores identifiquem os colaboradores relevantes em determinados domínios e os envolvam ativamente em projetos ou tarefas específicas. Além disso, a identificação dos Colaboradores Chave pode facilitar a identificação de oportunidades de aprendizado e desenvolvimento profissional, uma vez que esses indivíduos podem atuar como mentores ou facilitadores para outros membros da equipe que desejam adquirir conhecimentos em áreas específicas.

Ao compreender e utilizar de maneira estratégica os dados obtidos a partir da identificação dos Colaboradores Chave, as organizações podem otimizar o desempenho das equipes, promover uma cultura de colaboração e fortalecer suas capacidades no âmbito profissional. Essa abordagem contribui para o crescimento e sucesso contínuo das equipes, além de criar um ambiente propício para a inovação e o avanço da organização como um todo.

2.3 Considerações Finais

O objetivo deste capítulo foi estabelecer os fundamentos conceituais abordados neste trabalho. Inicialmente, foi introduzido o conceito de Ecossistema de Software e sua relevância no contexto atual do desenvolvimento de software, especialmente no ambiente empresarial.

A análise da inter-relação das três dimensões do ECOS será o principal foco da pesquisa, visando a geração de conhecimentos que auxiliem no gerenciamento interno das empresas. Nesse sentido, destacou-se a importância do Sistema de Reputação, pre-

sente no Desenvolvimento Distribuído de Software, e sua relevância na identificação de colaboradores-chave. Embora esses conceitos já tenham sido testados em ambientes abertos relacionados ao ECOS, ainda carecem de validação completa em ambientes internos e privados de empresas de desenvolvimento de software.

A adoção de um sistema de reputação possibilita a identificação e o reconhecimento dos colaboradores mais habilidosos e confiáveis, promovendo uma melhor alocação de recursos e aumentando a eficiência das equipes distribuídas. Essa abordagem contribui para o sucesso dos projetos de desenvolvimento de software, assegurando que as pessoas certas estejam envolvidas nas tarefas adequadas, resultando em entregas de alta qualidade e maior satisfação dos clientes.

3 Trabalhos Relacionados

Neste capítulo, são apresentados seis artigos relacionados a este trabalho, que abordam diferentes aspectos do estudo de ECOS. Esses artigos contribuem para a literatura sobre ECOS, incluindo métodos de identificação de reputação para auxiliar a gestão de TI, métricas de redes complexas e outros aspectos relevantes para o estudo.

Uma tabela é apresentada, correlacionando os aspectos dos testes realizados nos artigos relacionados e neste trabalho. Essa tabela permite uma análise comparativa dos resultados e contribui para a compreensão das similaridades e diferenças entre as abordagens adotadas.

Os artigos relacionados fornecem uma base sólida de conhecimento e fundamentam o desenvolvimento deste trabalho, enriquecendo a discussão e a validação das metodologias propostas. A análise e comparação dos resultados obtidos nesses artigos contribuem para a construção de um corpo de conhecimento mais completo sobre ECOS e fortalecem a relevância deste estudo.

3.1 Managing and Monitoring Software Ecosystem to Support Demand and Solution Analysis

Em SANTOS (2016), a perspectiva de ECOS é examinada para propor e avaliar uma abordagem de gerenciamento e monitoramento para apoiar as atividades de gestão de TI, mais especificamente a análise de demanda e solução. A pesquisa iniciou uma investigação de um *framework* para modelagem e análise de ECOS para os seguintes objetivos:

1. Desenvolver um *framework* com base em uma revisão da literatura para ajudar os pesquisadores a compreender melhor as dimensões e principais conceitos de ECOS e analisar as plataformas das organizações.
2. Identificar mecanismos de gestão de TI em governança e socialização no contexto de ECOS com base na opinião de especialistas.

3. Identificar os principais indicadores de monitoramento para a gestão de TI quanto à sustentabilidade de uma plataforma em um ambiente de ECOS, com base em estudos observacionais em cenários reais.
4. Definir métodos de gestão e monitoramento de ECOS para suporte às atividades de gestão de TI, mais especificamente análise de demanda e solução.
5. Avaliação de alguns módulos da abordagem com praticantes (gerentes e arquitetos de TI) realizando análise de demanda e solução em um cenário real.

Um estudo de viabilidade foi conduzido com profissionais em um cenário real para avaliar a abordagem e contribuir para a pesquisa e prática na comunidade ECOS. O piloto foi realizado inicialmente com três participantes e, após refinamento, o estudo foi realizado com oito participantes. No entanto, o uso da abordagem não se mostrou tão eficiente, uma vez que os praticantes precisaram de algum tempo para aprender a utilizá-la em seu benefício.

Após analisar as respostas dos participantes às tarefas propostas e as avaliações do estudo, foram encontradas indicações de que a abordagem é aplicável para o gerenciamento e monitoramento de ECOS, especialmente para a análise de demanda e soluções no contexto de gestão de TI. Foram identificadas diversas oportunidades de melhoria, principalmente relacionadas à interface gráfica do usuário (GUI). Os participantes reconheceram a utilidade da abordagem e compreenderam o impacto da perspectiva ECOS em suas atividades diárias. (SANTOS, 2016). O estudo também destaca a importância de manter uma plataforma sustentável.

A principal dificuldade encontrada na análise está relacionada ao tamanho da amostra, uma vez que o número de participantes foi pequeno, o que não é ideal do ponto de vista estatístico (SANTOS, 2016). Infelizmente, essa é uma dificuldade recorrente em estudos empíricos na área de Engenharia de Software, especialmente para métodos que requerem avaliação industrial, como é o caso deste estudo. Portanto, o trabalho apresenta uma limitação em relação aos resultados, que são considerados indicações e não evidências. O presente trabalho contribui nesse sentido, fornecendo uma amostra significativa de dados que permite gerar evidências e validar o processo modelado.

Ao abordar a perspectiva ECOS e oferecer uma metodologia para o gerenciamento de TI em ambientes de plataforma, a pesquisa de SANTOS (2016) e o presente trabalho enriquecem a compreensão e a aplicação prática desse conceito emergente na comunidade de Engenharia de Software.

3.2 Um Modelo Dinâmico de Reputação para Apoiar a Manutenção Colaborativa de Software

No trabalho de LÉLIS (2017), é apresentado um modelo para o cálculo da reputação de desenvolvedores de software, utilizando técnicas de Dinâmica de Sistemas. Esse modelo permite simular a evolução da reputação ao longo do tempo e serve como base para a construção de uma infraestrutura de informações de reputação dinâmica. O objetivo dessa infraestrutura é gerenciar e monitorar as informações de reputação de desenvolvedores distribuídos geograficamente, a fim de apoiar a atribuição de tarefas de manutenção.

O trabalho também aborda elementos de visualização e colaboração, integrados às atividades de manutenção de software, dentro dessa infraestrutura. Foram realizadas demonstrações de provas de conceito e experimentos utilizando dados reais de uma empresa, a fim de determinar a viabilidade e conformidade do modelo proposto, bem como dos recursos fornecidos pela infraestrutura.

Os experimentos envolveram a avaliação da viabilidade do modelo de reputação, comparando as escolhas dos gerentes ao atribuir desenvolvedores com os maiores índices de reputação, por meio da coleta de dados simulados. As provas de conceito demonstraram a viabilidade de utilização e adesão à visualização, assim como os aspectos colaborativos proporcionados pela infraestrutura. Os dados foram utilizados para construir as equações necessárias para simular o processo e observar o comportamento da reputação dos desenvolvedores. Além disso, as informações simuladas foram associadas às escolhas dos gerentes ao atribuir os desenvolvedores às solicitações.

A teoria da Dinâmica de Sistemas foi aplicada nesse contexto, permitindo entender o passado, analisar o presente e projetar o comportamento futuro da reputação dos desenvolvedores. Os modelos desenvolvidos podem ser considerados preditivos, desde que

sejam baseados em informações estáveis e em um banco de dados confiável.

A infraestrutura proposta por LÉLIS (2017) serve como base para o módulo de sistema de reputação do processo proposto neste trabalho, devido às características similares do cenário de teste e dos dados utilizados. A utilização desse modelo e infraestrutura contribui para a compreensão e aplicação prática do cálculo da reputação dos desenvolvedores.

3.3 Topological Analysis in Scientific Social Networks to Identify Influential Researchers

GUÉRCIO et al. (2017) apresentam uma análise da colaboração, modelada a partir da estrutura de redes complexas, onde os nós da rede representam agentes que colaboram para atingir objetivos comuns. A partir daí, busca-se encontrar pesquisadores da rede que possuem maior importância na produção científica da rede modelada. Os nós mais importantes têm maior potencial de disseminação de informações e podem atingir diversos agentes, reduzindo o esforço e a perda de comunicação. Dessa forma, as informações produzidas e suportadas pelos usuários mais importantes da rede têm maior visibilidade. Além disso, agentes com maior influência na rede também possuem maior potencial de colaboração, pois estão fortemente conectados a um grande número de nós.

O objetivo é descrever como a topologia de rede pode ser usada para identificar pesquisadores que possam ter o maior potencial de colaboração na rede modelada. Após identificar os agentes com maior potencial de colaboração, foi realizada a remoção de nós para verificar o impacto dessa perda de pesquisador. O impacto é avaliado pela distância média de colaboração na rede, que indica a colaboração entre os nós que a compõem (GUÉRCIO et al., 2017).

Foi utilizada uma medida de centralidade para detectar a força da colaboração na rede social científica brasileira. Os resultados permitem a análise de redes complexas, observando o nível de colaboração entre pesquisadores, seus relacionamentos fortes e fracos e tenta identificar pesquisadores que desempenham um papel central na rede de coautoria. Além disso, a rede complexa foi classificada como uma rede livre de escala.

Os fundamentos da análise são aplicados no processo deste trabalho para auxiliar na identificação dos relacionamentos dos colaboradores entre as equipes, bem como sua influência e especialidade.

3.4 Ecosystems in GitHub and a Method for Ecosystem Identification Using Reference Coupling

Neste artigo, BLINCOE; HARRISON; DAMIAN (2015) propõem um novo método chamado acoplamento de referência é para detectar dependências técnicas entre projetos. Esse método estabelece dependências por meio de referências cruzadas entre projetos especificados pelo usuário.

O método foi utilizado para analisar o ECOS em projetos hospedados no GitHub e identificou diversas características. Constatou-se que a maioria dos ECOS está centrada em torno de um projeto e está interconectada com outros ECOS. O tipo predominante de ecossistemas são aqueles que desenvolvem ferramentas de apoio ao desenvolvimento de software (BLINCOE; HARRISON; DAMIAN, 2015).

Verificou-se também que o comportamento social dos proprietários do projeto se alinha bem com as dependências técnicas no ecossistema, mas o comportamento social dos colaboradores do projeto era inconsistente com essas dependências (BLINCOE; HARRISON; DAMIAN, 2015).

O estudo demonstra a importância de ECOS para a manutenção e desenvolvimento colaborativo de projetos, da mesma forma que empresas com várias equipes trabalhando em projetos diferentes podem colaborar para melhorar a produtividade, qualidade e velocidade de seus respectivos projetos. Cenário que se enquadra à amostra de dados coletados neste trabalho.

3.5 Complex Network Analysis in a Software Ecosystem: Studying the Eclipse Community

O objetivo de GUERCIO et al. (2018) neste artigo, é modelar a dimensão social dos ECOS por meio de dados de contribuição. A avaliação das informações da dimensão social do SECO visa esclarecer como os desenvolvedores colaboram nos projetos e entre si. Este artigo visa identificar aspectos que destacam as interações entre os diferentes grupos de pesquisa. Essa interação reflete as ações de alguns colaboradores para ajudar outros em dificuldade em determinados aspectos dos projetos em que estão envolvidos.

Diferentes redes complexas foram modeladas usando dados extraídos do sistema de controle de versão do ecossistema Eclipse. As redes representam interações entre usuários e projetos. As análises de rede realizadas indicam a viabilidade de usar essa abordagem para identificar contribuintes significativos no ecossistema, melhorar a saúde e a eficiência de ECOS. Ao aumentar a força e o número de relacionamentos entre os membros ativos do ecossistema e os recém-chegados, é possível fortalecer o ecossistema e ajudar os novos membros a desenvolver novas capacidades que contribuirão para o ecossistema (GUERCIO et al., 2018).

Três redes diferentes foram modeladas. A primeira envolve as contribuições dos usuários do ecossistema para o projeto. A segunda rede avalia os colaboradores que fizeram comentários. Na terceira rede, foi usado uma medida R que usa informações das contribuições de cada usuário para traçar um relacionamento entre os usuários com base em suas contribuições (GUERCIO et al., 2018).

Após a modelagem dos dados, foi possível verificar se as redes estavam fortemente conectadas e como seus membros estavam distribuídos. Dentre as dificuldades encontradas, pode-se destacar o longo tempo gasto para extrair e limpar os dados, e a dificuldade em atribuir corretamente um valor de importância às contribuições. Neste trabalho serão implementados e aperfeiçoados novos métodos para atribuir valor no sistema de reputação e validar os modelos de redes na nova amostra de dados.

3.6 sSECO-Process: Avaliando a Dimensão Social em Ecossistemas de Software

GUERCIO (2018) apresenta o sSECO-Process, um processo de análise da dimensão social de ECOS, apoiado em técnicas de Redes Complexas, que permite apresentar as relações existentes em ECOS através da visualização e utilização de métricas relevantes da rede. Esse processo serve como base para analisar diferentes ECOS, e deu origem a tecnologias que dão suporte à coleta, tratamento, armazenamento e processamento de dados, que auxiliarão cientistas a realizar novas pesquisas relacionadas ao desenvolvimento de software. O trabalho também propõe medidas cujo objetivo é medir o nível de colaboração entre os usuários em um projeto, e como os usuários podem ser associados sem um registro das interações entre eles.

Foi realizada uma avaliação inicial com dados reais e suporte especializado para avaliar o processo desenvolvido, e após a avaliação inicial o processo foi ampliado para aprimorar e complementar as atividades realizadas durante a análise da dimensão social de ECOS (GUERCIO, 2018).

A fonte de dados escolhida para a extração foi o GitHub, por possibilitar que estudos com outros escopos sejam possíveis. A forma de extração, assim como o do presente trabalho, é via API, disponibilizada pela própria plataforma. Outra semelhança está nos próprios dados, que seguem padrões de informações similares, facilitando a replicação do processo de coleta, tratamento e análise desses dados.

3.7 Análise comparativa

Os trabalhos relacionados são comparados a um conjunto de características relacionadas aos seus dados, metodologia e contribuição na Tabela 3.1, organizada nas seguintes colunas:

- Trabalho: Referência ao trabalho relacionado.
- Dados: Representa a origem dos dados coletados.
- ECOS: Representa o ecossistema e/ou a plataforma de análise.

- Contribuição: Principal contribuição para o processo proposto neste trabalho.

Tabela 3.1: Tabela comparativa dos trabalhos relacionados.

Trabalho	Dados	ECOS	Contribuição
(SANTOS, 2016)	Planilhas e documentos	Organização Bancária	Framework para análise de ECOS
(LÉLIS, 2017)	Solicitações de mudanças e histórico de alterações	MantisBT	Infraestrutura de reputação dinâmica
(GUÉRCIO et al., 2017)	Relação entre autores de artigos científicos	Lattes Platform	Identificação de influência na rede
(BLINCOE; HARRISON; DAMIAN, 2015)	Código fonte	GitHub	Identificação de dependência entre projetos e comportamento social
(GUERCIO et al., 2018)	Código fonte e comunicação entre colaboradores	Eclipse	Identificação da interação entre diferentes grupos
(GUERCIO, 2018)	Código fonte	GitHub	Processo para análise de ECOS e métricas de redes complexas

3.8 Considerações Finais

O objetivo deste capítulo é apresentar trabalhos relacionados e precursores, seus objetivos, resultados, contribuições e deficiências. Cada trabalho elabora e detalha processos e/ou *frameworks* que contribuiram para a literatura e desenvolvimento deste TCC. Além disso, os testes e resultados apresentados promovem uma discussão comparativa com os resultados deste trabalho, avaliando as semelhanças e diferenças para auxiliar a análise de viabilidade desses modelos em ambientes de desenvolvimento de software profissional e privado.

Por fim, é apresentada uma tabela de análise comparativa entre os trabalhos com relação às suas características, ambientes, testes aplicados e área de contribuição.

4 Processo para Análise de Colaboração no Ecosistema

Neste capítulo, é apresentada a definição de um processo cujas atividades utilizam a plataforma *SYDLE ONE*⁴ e a biblioteca *seaborn* do *python*, com foco na coleta de dados do ecossistema e na análise da reputação dos colaboradores. O objetivo principal é estabelecer um sistema robusto e abrangente que permita explorar a complexidade e as particularidades do ecossistema de amostra.

Ao adotar a plataforma *SYDLE ONE* como base, aproveitamos seu potencial para viabilizar a coleta de dados em grande escala. Por meio de uma abordagem cuidadosamente elaborada, com o desenvolvimento de métodos eficientes para extrair informações cruciais do ecossistema, permitindo uma visão aprofundada da contribuição de cada colaborador. Com base nos dados coletados, busca-se compreender a confiabilidade, a experiência e o histórico de desempenho de cada membro do ecossistema. Essa análise permite identificar e reconhecer os contribuidores mais valiosos, assim como sua especialidade com base na sua atuação.

4.1 Definição do process

A amostra de dados utilizada nesta pesquisa foi coletada diretamente da plataforma *SYDLE ONE*, com o intuito de melhor se adequar às características particulares da mesma e de seus colaboradores. O processo foi organizado em quatro atividades principais, visando proporcionar uma análise detalhada do ecossistema. A Figura 4.1 apresenta o ciclo de funcionamento do processo, sendo que os subprocessos de coleta e análise de dados serão apresentados nas seções a seguir.

⁴(www.sydle.com)

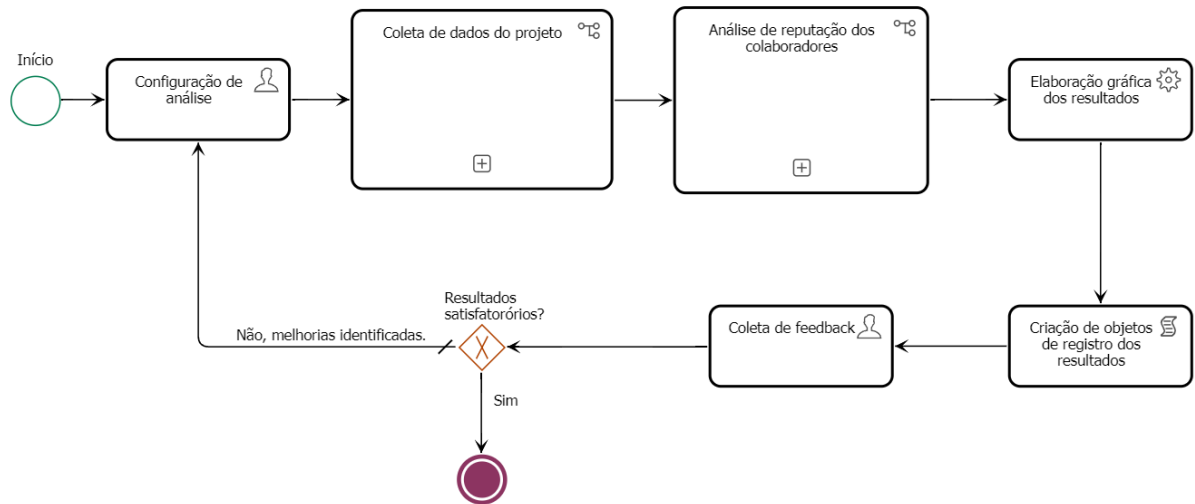


Figura 4.1: Arquitetura do processo de análise de reputação.

4.2 Atividade: Configuração de Análise

Na etapa inicial, procede-se a configuração dos dados a serem coletados, uma atividade humana que requer uma atuação manual. O desenvolvimento na plataforma ocorre em pacotes segmentados, que por sua vez são segmentados em classes, as quais são relacionadas apenas abstratamente ao setor de atuação. Portanto, é imperativo estabelecer e agrupar a correspondência entre cada classe e sua categoria correspondente, além de excluir classes originadas de sistemas nos quais os colaboradores não estão diretamente envolvidos.

A contribuição do desenvolvedor é permanentemente registrada no histórico do sistema, garantindo que suas ações sejam rastreáveis ao longo do tempo. Portanto, as configurações do sistema também permitem a exclusão de colaboradores que já não fazem parte da empresa, assegurando a atualização e a integridade dos dados. Essa funcionalidade é fundamental para manter o sistema alinhado com a estrutura organizacional atualizada.

A Figura 4.2 exibe a tela de configurações da análise. Cada setor é relacionado com os pacotes e, se necessário, a classes específicas. A filtragem das classes que se enquadram nos padrões de análise é realizado internamente, então o número total de

classes analisáveis é calculado e exibido. Os campos referentes ao “Período de referência” e “Fator de decaimento” serão explicados nas seções 4.4.2 e 4.4.3.

Período de referência	Fator de decaimento		
180	0,000087		
Configuração de setor			
Setor	Pacotes IDs	Classes IDs	Total de classes analisáveis
CRM	criptografado	criptografado	34
Billing	criptografado	criptografado	30
Integrador	criptografado	criptografado	15
Exclusão de usuários			
criptografado			

Figura 4.2: Configurações da análise

4.3 Subprocesso: Coleta de Dados do Ecossistema

A Figura 4.3 ilustra a estrutura do subprocesso de coleta e filtragem de dados do projeto.

Esse subprocesso é composto em quatro etapas principais:

- Coletada dos dados a partir das configurações predefinidas da relação de classes por setor.
- A aplicação de filtros baseados nas características do projeto e remover classes que não se enquadram no formato esperado de análise, como classes originadas do sistema ou que não são relevantes para o objetivo do estudo.
- Disponibilização dos dados por meio de uma *API* para as etapas subsequentes do processo.

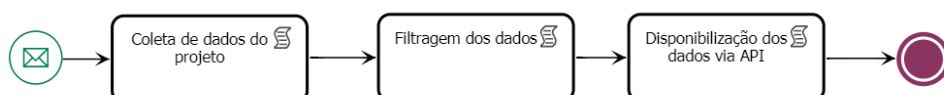


Figura 4.3: Subprocesso de coleta de dados

A plataforma de desenvolvimento do ECOS possui uma característica de *low code*, o que implica em um controle de versão distinto do padrão utilizado em outros ecossistemas

como GitHub. Nesse contexto, as classes são criadas e definidas como objetos que contêm propriedades e scripts. Para efetuar alterações, é necessário acessar visualmente o objeto em questão e realizar as edições desejadas. A cada salvamento, um registro (*log*) é criado no histórico, contendo a versão anterior da classe, juntamente com informações sobre a data/hora da versão e seu autor. A Figura 4.4 exemplifica um diagrama de classe resumido da plataforma.

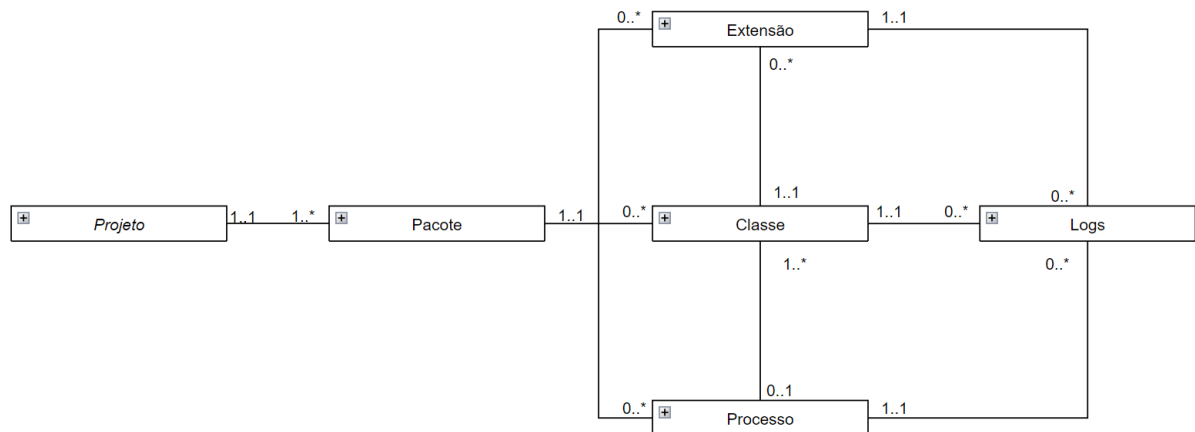


Figura 4.4: Diagrama de classe estrutural da plataforma

Na coleta de dados, cada *log* das classes pertencentes aos pacotes em avaliação é analisado. Entretanto, durante o desenvolvimento, é comum a criação de vários *logs* consecutivos pelo mesmo desenvolvedor. Para preservar a integridade do cálculo da contribuição na fase de análise, foi estabelecida a seguinte regra: caso hajam *logs* consecutivos do mesmo autor no mesmo dia, apenas o mais recente é considerado na coleta dos dados. Essa regra assegura que a contribuição do desenvolvedor seja representada de forma precisa, evitando a duplicação indevida de registros e mantendo a consistência do processo de análise.

4.4 Subprocesso: Análise de Reputação dos Colaboradores

A Figura 4.5 apresenta a estrutura do subprocesso de análise dos dados coletados. Esse subprocesso ocorre externamente à plataforma, utilizando *Python*⁵ e a biblioteca *difflib*⁶ para calcular o valor da contribuição com base na Distância de Levenshtein, que será apresentada na seção 4.4.1. O subprocesso é composto em quatro etapas principais:

- Obter os dados coletados por meio da *API* disponibilizada, garantindo o acesso às informações necessárias para a análise.
- Aplicação da metodologia de cálculo da contribuição, realizando a comparação dos *logs* e aplicando a Distância de Levenshtein.
- Aplicação das métricas ponderadas, que consideram a data da contribuição em relação ao valor atribuído.
- Normalização dos resultados, para auxiliar na análise comparativa, permitindo uma análise mais precisa e equilibrada dos resultados.

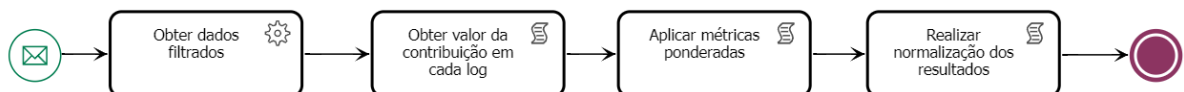


Figura 4.5: Subprocesso de análise de dados

Com base nos dados coletados dos *logs*, é realizado um processo de agrupamento de acordo com suas classes de origem e ordenados por data de criação. Em seguida, cada registro é avaliado por meio de comparações entre versões, da mais antiga até a mais recente.

⁵<https://www.python.org/>

⁶<https://docs.python.org/3/library/difflib.html>

4.4.1 Distância de Levenshtein

Como mencionado anteriormente, o versionamento na plataforma de análise difere do comumente utilizado em plataformas como o *GitHub*, onde cada *commit* carrega os dados das alterações realizadas. Nesta plataforma, cada *log* criado armazena apenas a versão completa da classe anterior à alteração. Portanto, tornou-se necessário um método para avaliar e calcular o valor da diferença em cada *log*. O algoritmo escolhido foi a Distância de Levenshtein, pois, para a análise realizada, o valor da contribuição será calculado com base nos *scripts* da classe, facilmente tratados como uma *string*.

A Distância de Levenshtein, também conhecida como distância de edição, é uma medida que quantifica a diferença entre duas sequências. Essa medida é definida como o número mínimo de operações necessárias para transformar uma sequência em outra (YUJIAN; BO, 2007). Dada uma sequência **s1** de comprimento **m** e uma sequência **s2** de comprimento **n**, a distância de *Levenshtein* entre **s1** e **s2** é calculada utilizando a fórmula recursiva 4.1:

$$\text{dist}(i, j) = \begin{cases} j & \text{se } i = 0 \\ i & \text{se } j = 0 \\ \text{dist}(i - 1, j - 1) & \text{se } s1[i] = s2[j] \\ \min(\text{dist}(i - 1, j) + 1, \\ \text{dist}(i, j - 1) + 1, \\ \text{dist}(i - 1, j - 1) + 1) & \text{caso contrário} \end{cases} \quad (4.1)$$

Onde:

- **i** e **j** são índices que variam de **0** a **m** e de **0** a **n**, respectivamente.
- **dist(i, j)** representa a distância de *Levenshtein* entre os primeiros **i** elementos da sequência **s1** e os primeiros **j** elementos da sequência **s2**.
- **s1[i]** e **s2[j]** são os elementos correspondentes das sequências **s1** e **s2**.
- **dist(i-1, j) + 1** representa uma operação de remoção.

- $\text{dist}(\mathbf{i}, \mathbf{j}-1) + 1$ representa uma operação de inserção.
- $\text{dist}(\mathbf{i}-1, \mathbf{j}-1) + 1$ representa uma operação de substituição.

O algoritmo preenche uma matriz $(\mathbf{m}+1) \times (\mathbf{n}+1)$ utilizando a fórmula acima, onde cada célula representa a distância entre os prefixos correspondentes de $\mathbf{s1}$ e $\mathbf{s2}$. O valor final na célula (\mathbf{m}, \mathbf{n}) é a distância total entre $\mathbf{s1}$ e $\mathbf{s2}$ (YUJIAN; BO, 2007).

Após o cálculo da distância entre um colaborador e um projeto é necessário avaliar o fator tempo para medir o impacto atual da contribuição no projeto e na experiência do colaborador, refletindo assim em sua reputação atual. Com esse objetivo, são implementadas duas métricas: decaimento temporal por período de referência, e decaimento temporal por exponencial, que estão descritas nas próximas subseções.

4.4.2 Decaimento temporal por período de referência

Essa métrica leva em consideração o tempo decorrido desde a contribuição do colaborador. A medida que o tempo avança, a contribuição passada recebe um peso menor em relação às contribuições mais recentes. Esse decaimento é aplicado de forma linear, onde cada contribuição é ponderada pelo tempo em dias que se passou desde sua realização. Adicionalmente, é incorporado o conceito de um “período de referência em dias”, que define a janela de tempo na qual as contribuições são consideradas relevantes para o cálculo da reputação atual do colaborador. Esse período de referência exerce influência sobre o nível de decaimento aplicado às contribuições.

A Equação 4.2 expressa o cálculo da contribuição ponderada por período de referência.

$$\text{Contribuição} = \text{Distância de Levenshtein} \times \left(1 - \frac{\text{Diferença de Tempo}}{\text{Período de Referência}} \right) \quad (4.2)$$

Na equação utilizada, a “Diferença de Tempo” refere-se à quantidade de dias entre a data do evento em questão e a data de referência estabelecida. Por sua vez, o “Período de Referência” é um valor determinante para a taxa de decaimento temporal aplicada. Um valor comumente utilizado para o “Período de Referência” é de 90 dias e

180 dias, mas esses valores podem variar dependendo do contexto, e são definidos durante a fase de configuração do sistema.

No entanto, é importante mencionar que uma desvantagem dessa abordagem ocorre quando a diferença de tempo entre a data do evento e a data de referência é maior do que o período estabelecido. Nesses casos, o valor da contribuição passa a ser negativo, o que pode não ser desejável dependendo do contexto e das regras de avaliação adotadas. É fundamental considerar esse aspecto e realizar uma análise cuidadosa ao interpretar as contribuições em relação à diferença de tempo e ao período de referência escolhido.

4.4.3 Decaimento temporal por exponencial

A abordagem adotada neste caso emprega um decaimento exponencial para avaliar as contribuições dos colaboradores ao longo do tempo. Esse decaimento é projetado de forma a atribuir um peso significativamente maior às contribuições mais recentes, enquanto as contribuições passadas têm um impacto progressivamente menor. Essa estratégia visa refletir a importância das atividades recentes no projeto e na reputação do colaborador.

Ao ajustar os parâmetros da função exponencial, é possível alcançar um decaimento menos acentuado no início, permitindo que as contribuições iniciais mantenham um certo grau de influência. Ao mesmo tempo, à medida que a diferença de tempo aumenta, o decaimento se torna mais pronunciado, resultando em uma queda mais acentuada na pontuação. No entanto, um aspecto importante dessa abordagem é o controle do decaimento para que a pontuação nunca atinja zero, garantindo assim que as contribuições mais antigas continuem a ser consideradas na avaliação da reputação do colaborador.

A Equação 4.3 é utilizada para calcular a pontuação levando em consideração este fator de decaimento, representado por *Decay*. Esse fator controla a taxa na qual a pontuação diminui à medida que o tempo aumenta. Quanto maior o valor do fator *Decay*, mais rapidamente a pontuação diminuirá ao longo do tempo.

$$\text{Contribuição} = \text{Distância de Levenshtein} \times e^{-\text{Decay} \times \text{Diferença de Tempo}} \quad (4.3)$$

É importante considerar que o fator de decaimento (*Decay*) é um parâmetro que pode variar dependendo do contexto e das necessidades do sistema. É comum que diferentes aplicações ajustem esse parâmetro para alcançar um equilíbrio adequado entre a relevância das contribuições recentes e passadas na avaliação da pontuação.

4.4.4 Período de referência X Fator de decaimento exponencial

Na Figura 4.6 pode ser observado um gráfico com a representação das equações 4.3 e 4.2. O eixo horizontal é utilizado para denotar o tempo decorrido, medido em unidades de tempo (dias, meses, anos etc.), desde a contribuição inicial. Por sua vez, o eixo vertical é utilizado para representar a pontuação atribuída à contribuição. No gráfico, são considerados dois períodos de referência: 90 dias e 180 dias para serem comparados com dois fatores de decaimento de 0,0035 e 0,000087, respectivamente. Esses valores foram escolhidos visando serem proporcionais aos períodos de referência correspondentes, onde se aproximam na metade dos períodos de referência respectivos.

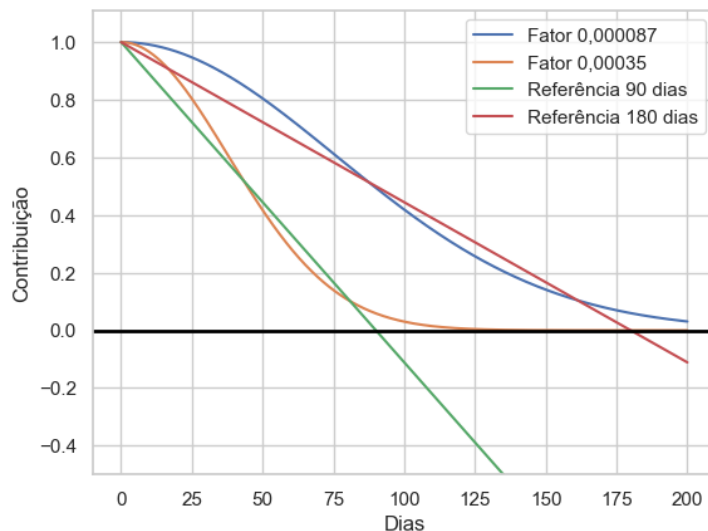


Figura 4.6: Comparação entre decaimento exponencial e período de referência

Essa representação gráfica permite visualizar a progressão do decaimento ao longo do tempo e comparar as influências proporcionais dos períodos de referência de 90 e 180 dias. Observa-se que os fatores têm impactos significativos nas pontuações, porém com padrões ligeiramente distintos. O gráfico exemplifica a natureza do decaimento e dos fatores proporcionais aos períodos de referência, mas os valores numéricos exatos e a

forma específica das curvas podem variar dependendo dos parâmetros e configurações adotados.

O valor resultante dessa análise é então categorizado e somado para compor a reputação do colaborador no setor referente ao pacote do qual a classe pertence. Isso significa que a reputação do colaborador é calculada com base em sua contribuição individual em cada *log*.

4.4.5 Normalização

No passo seguinte, para auxiliar na comparação entre os setores e colaboradores, é realizada uma normalização da pontuação dos colaboradores. A normalização é um processo que ajusta a pontuação para uma escala específica, geralmente entre 0 e 1, facilitando a comparação e interpretação dos resultados, especialmente útil quando diferentes contribuições têm magnitudes diferentes e é necessário comparar a relevância relativa entre elas. A normalização aplicada dimensiona a pontuação para o intervalo entre 0 e 1, considerando o valor máximo observado.

Dessa forma, ao avaliar a contribuição do colaborador é possível obter uma visão mais abrangente e precisa de seu desempenho em relação a cada setor e ao projeto. Essas métricas auxiliam na identificação de padrões de comportamento e na atribuição de uma reputação adequada, o que pode ser utilizado para tomar decisões informadas sobre a participação e o envolvimento do colaborador nas atividades relacionadas ao setor.

4.5 Atividade: Elaboração gráfica e armazenamento

Durante a fase de avaliação dos resultados obtidos, a biblioteca *Seaborn*⁷ para *Python* foi utilizada como ferramenta fundamental. Com o auxílio de gráficos comparativos, foram analisadas as perspectivas de projeto, setor e colaborador com base nos valores normalizados de contribuição total em relação à contribuição por fator de decaimento, bem como em relação à contribuição total em relação ao período de referência.

A biblioteca *Seaborn* é amplamente reconhecida por sua capacidade de visu-

⁷<https://seaborn.pydata.org/>

alização de dados, oferecendo uma variedade de gráficos estatísticos e ferramentas de análise. Sua utilização proporcionou uma representação visual clara e intuitiva das relações entre as diferentes variáveis em estudo, facilitando a obtenção de *insights* significativos e descobertas aos stakeholders e equipes envolvidas no processo de avaliação, fundamentais para embasar a tomada de decisões no contexto avaliado.

Após a obtenção dos resultados e gráficos, são criados objetos na plataforma para armazenar as informações individuais dos colaboradores e dos setores. Esses objetos são criados como parte do processo de armazenamento e organização dos dados, com o objetivo de facilitar o acesso e a visualização das informações relevantes.

Para os colaboradores, são criados objetos que contêm informações como usuário, histórico de contribuições, reputação individual e especialidade. Esses objetos permitem armazenar e atualizar os dados específicos de cada colaborador, facilitando o acompanhamento de suas contribuições ao longo do tempo.

Da mesma forma, para os setores, são criados objetos que contêm informações como nome, reputação dos colaboradores e colaborador-chave. Esses objetos permitem armazenar e organizar os dados específicos de cada setor, permitindo uma análise mais precisa da contribuição e desempenho de cada área do projeto.

Esses objetos criados na plataforma são essenciais para a organização e gerenciamento dos dados, possibilitando consultas rápidas e eficientes, além de facilitar a geração de relatórios e análises posteriores. Com eles, é possível armazenar de forma estruturada as informações individuais dos colaboradores e dos setores, contribuindo para uma melhor compreensão e avaliação do projeto como um todo.

5 Avaliação: Prova de Conceito

Neste capítulo é conduzida uma Prova de Conceito (PoC) com o intuito de verificar a viabilidade do processo, validando-o em baixa escala. A PoC normalmente caracteriza uma implementação resumida e simples, que antecede o protótipo do projeto e assim busca validar o software ainda na sua etapa conceitual.

Os resultados da PoC provenientes de uma análise realizada em um ecossistema real são apresentados. São detalhadas as características dos projetos, bem como dos colaboradores envolvidos, acompanhados de exemplos gráficos ilustrativos de suas respectivas reputações.

Por fim, uma entrevista também foi conduzida entre os colaboradores, com o intuito de obter seus *feedbacks* acerca dos resultados alcançados. Essa entrevista visou capturar a percepção dos colaboradores sobre a precisão e confiabilidade dos dados gerados por meio da análise do ecossistema.

5.1 Amostra de dados

Para a realização deste estudo analítico, optou-se por selecionar um projeto que abrange três setores fundamentais da empresa: CRM, *Billing* e Integrador. Cada setor é composto por uma equipe de colaboradores atualmente constituída por 5, 7 e 3 membros, respectivamente. Devido à natureza prolongada do projeto, uma variedade de colaboradores já fez parte dele ao longo do tempo, buscando adquirir experiência com a plataforma e as metodologias adotadas pela empresa.

A Figura 5.1 exemplifica a estrutura organizacional da empresa. Onde cada projeto pode ou não possuir mais de um time, e cada time possui um líder (*Team Tech Lead* (TTL)).

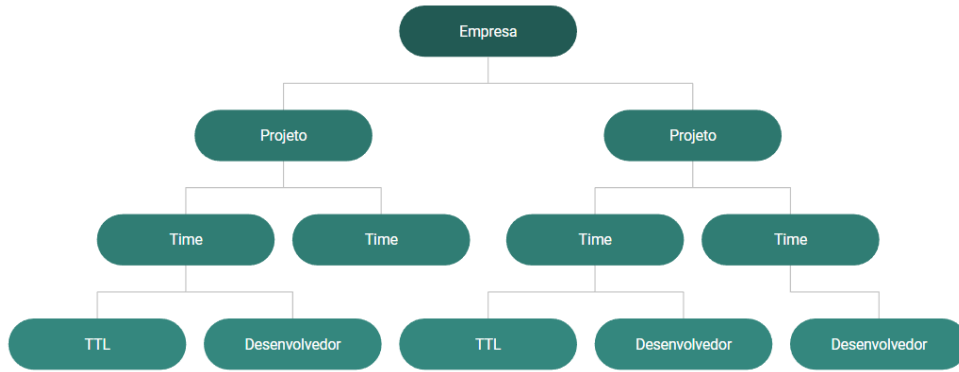


Figura 5.1: Hierarquia organizacional

A fim de facilitar a análise e evitar um processo de filtragem e adaptação complexo e oneroso, optou-se por categorizar apenas as principais classes de atuação nos setores mencionados. Essa abordagem permitiu concentrar a avaliação nas áreas mais relevantes, enquanto reconhecendo a possibilidade de existirem extensões e processos secundários que não foram considerados em profundidade neste estudo.

É importante destacar que a seleção dos critérios e a organização do processo de avaliação contaram com a participação ativa e consultoria dos próprios colaboradores envolvidos junto a empresa. Essa abordagem colaborativa contribuiu para garantir a validade e a eficácia da análise, ao mesmo tempo, em que promoveu um ambiente de aprendizado mútuo entre os profissionais que participaram do projeto em análise neste estudo.

O estudo realizado proporcionou resultados significativos em três perspectivas distintas: a do projeto, a do setor e a dos colaboradores envolvidos. Ao considerar as opiniões e as percepções dos colaboradores, foi possível obter uma visão mais abrangente das oportunidades de aprendizado, da satisfação no trabalho e do impacto do projeto em suas trajetórias profissionais.

Ao estender a análise para outros projetos, uma perspectiva adicional seria incluída: a da empresa como um todo. Essa perspectiva ampliada proporcionaria informações valiosas para aprimorar a gestão de projetos e a tomada de decisões estratégicas ao nível organizacional.

A fim de preservar a segurança das informações dos usuários nas análises conduzidas, os dados pessoais foram criptografados e apenas um identificador representativo é

exibido publicamente. Dessa forma, as informações sensíveis são protegidas e a privacidade dos usuários é mantida.

5.2 Perspectiva do colaborador

Após a implementação do *processo*, os resultados obtidos foram cuidadosamente organizados e armazenados no banco de dados da plataforma. Ao avaliarmos a perspectiva do colaborador, fica evidente que sua reputação individual não se restringe apenas ao projeto atual em que está envolvido. Sua reputação está intrinsecamente ligada ao seu tempo de contratação na empresa e à sua participação em múltiplos projetos ao longo do tempo. Essa abordagem leva em consideração não apenas o desempenho do colaborador em um único contexto, mas também sua experiência acumulada e sua especialidade em diferentes áreas.

Para ilustrar a atuação do colaborador em múltiplos projetos, selecionamos um colaborador como exemplo e aplicou-se a mesma metodologia em outro projeto, enfocando sua participação nesse novo contexto. Algumas limitações burocráticas impediram a coleta completa de dados de todos os colaboradores nesse segundo projeto, mas a análise específica desse colaborador em particular nos forneceu informações suficientes para avaliar a metodologia proposta.

A análise gera resultados individuais para cada colaborador, como ilustrado na Figura 5.2.

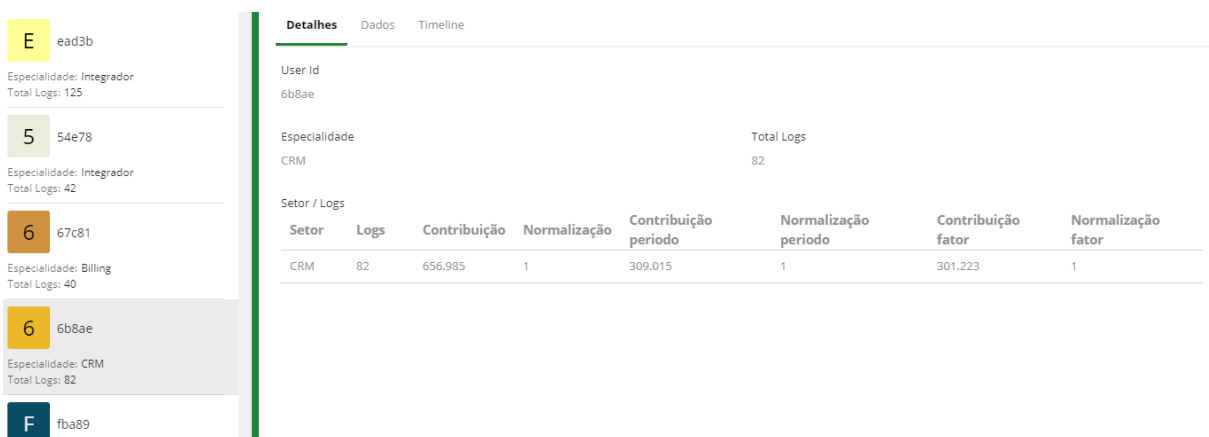


Figura 5.2: Análise da reputação do colaborador em um projeto

A Tabela 5.1 exibe os resultados dos valores de contribuição e da reputação individual do colaborador em um projeto específico. Nessa tabela, é possível observar os diferentes valores de contribuição atribuídos para o setor CRM e o cálculo da reputação individual com base nessas contribuições.

Além disso, a especialidade do colaborador é calculada com base no maior valor de contribuição ponderado pelo fator de decaimento exponencial. No exemplo apresentado na tabela, como há apenas um setor considerado, ele é automaticamente selecionado como a especialidade do colaborador.

Setor	Logs	Contribuição	C. Período	C. Exponencial
CRM	82	656985	309015	301223

Tabela 5.1: Tabela de reputação do colaborador em um único projeto

A tabela apresentada oferece uma visão clara dos resultados da análise, permitindo uma avaliação precisa da contribuição individual do colaborador em relação a cada setor do projeto. Além disso, por meio da identificação da especialidade, é possível destacar a área em que o colaborador demonstrou a maior contribuição e, conforme o critério adotado neste estudo, sua especialização.

Porém, ao considerar um segundo projeto e combinar as contribuições de ambos, obtém-se o resultado total da contribuição do mesmo colaborador. Essa visão mais abrangente é representada na Figura 5.3, onde são apresentados os valores de contribuição de cada projeto, bem como a reputação total do colaborador considerando ambos os projetos.

User Id							
6b8ae							
Especialidade	Total Logs						
CRM	132						
Setor / Logs							
Setor	Logs	Contribuição	Normalização	Contribuição período	Normalização período	Contribuição fator	Normalização fator
CRM	98	763.443	1	290.090	1	303.356	1
ServiceDesk	34	262.455	0.3437781209	-62.697	-0.2161282919	3.468	0.0114321127

Figura 5.3: Análise da reputação do colaborador em múltiplos projetos

As Tabelas 5.2 e 5.3 apresentam os resultados obtidos a partir da somatória das

análises dos dois projetos. Essa abordagem de considerar a reputação do colaborador em múltiplos projetos permite uma visão mais abrangente e precisa de suas habilidades e especialidades, levando em conta a soma total de suas contribuições ao longo do tempo. Dessa forma, podemos identificar padrões de desempenho, áreas de maior experiência e o impacto geral do colaborador em diversos contextos de trabalho.

Setor	Logs	Contribuição	C. Período	C. Exponencial
CRM	98	763443	290090	303356
ServiceDesk	34	262455	-62697	3468

Tabela 5.2: Tabela de reputação do colaborador em múltiplos projetos

Colaborador	Normalização	N. Período	N. Exponencial
CRM	1	1	1
ServiceDesk	0,34378	-0,21612	0,01143

Tabela 5.3: Tabela de reputação normalizada do colaborador em múltiplos projetos

É importante ressaltar que, embora a análise tenha sido aplicada apenas a um colaborador específico neste segundo projeto, os resultados obtidos fornecem *insights* valiosos sobre a viabilidade e eficácia da metodologia em avaliar a reputação de colaboradores envolvidos em múltiplos projetos. Isso sugere que a metodologia pode ser replicada e aplicada de forma semelhante a outros colaboradores em diferentes contextos.

Ao analisar as Tabelas 5.2 e 5.3, fica evidente o impacto negativo do cálculo da contribuição quando o período de referência é superior ao definido. Nesses casos, o valor da contribuição se torna negativo, afetando diretamente o valor total da contribuição do colaborador. Isso demonstra a efetividade do cálculo por fator de decaimento exponencial, pois mesmo quando o valor é mínimo, ele ainda é contabilizado positivamente e não afeta os resultados anteriores.

Essa observação ressalta a importância do fator de decaimento exponencial na avaliação da contribuição ao longo do tempo. Ele permite que contribuições antigas sejam gradualmente reduzidas, mantendo o foco nas contribuições mais recentes e relevantes. Dessa forma, o cálculo da contribuição se adapta às mudanças na participação e

envolvimento do colaborador ao longo do tempo, garantindo resultados mais precisos e atualizados.

Além disso, a influência negativa do cálculo da contribuição quando o período de referência é excedido destaca a necessidade de definir corretamente o período de análise, levando em consideração a relevância e o tempo de atuação do colaborador em relação ao projeto. Isso garante que a avaliação da contribuição seja adequada e não distorcida por informações desatualizadas ou irrelevantes.

As Figuras 5.4, 5.5 e 5.6 fornecem uma análise comparativa entre o valor da total da contribuição e o valor relativo obtido das métricas por meio do fator de decaimento exponencial e por período de referência (utilizando referências de 180 e 90 dias). Cada ponto do gráfico representa o valor médio de todas as contribuições registradas nos *logs* realizados no dia. O eixo vertical representa o valor da contribuição, enquanto o eixo horizontal representa o tempo em dias de diferença com relação à data da análise. Essas visualizações permitem uma avaliação do comportamento dos cálculos mencionados anteriormente e destacam os pontos discutidos anteriormente na seção 4.4.4.

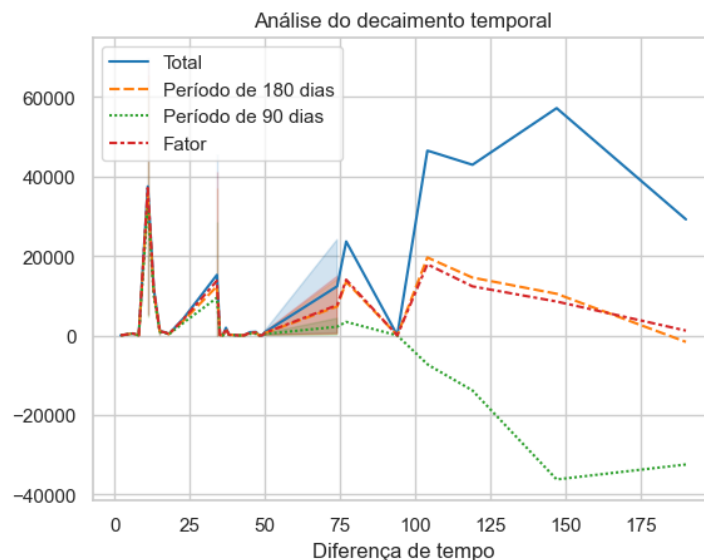


Figura 5.4: Comparação entre os valores de contribuição por log do colaborador 54e78

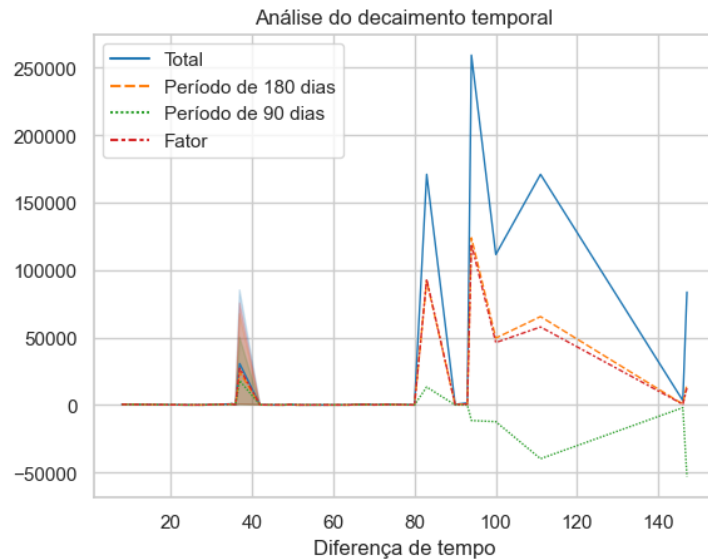


Figura 5.5: Comparação entre os valores de contribuição por log do colaborador e130e

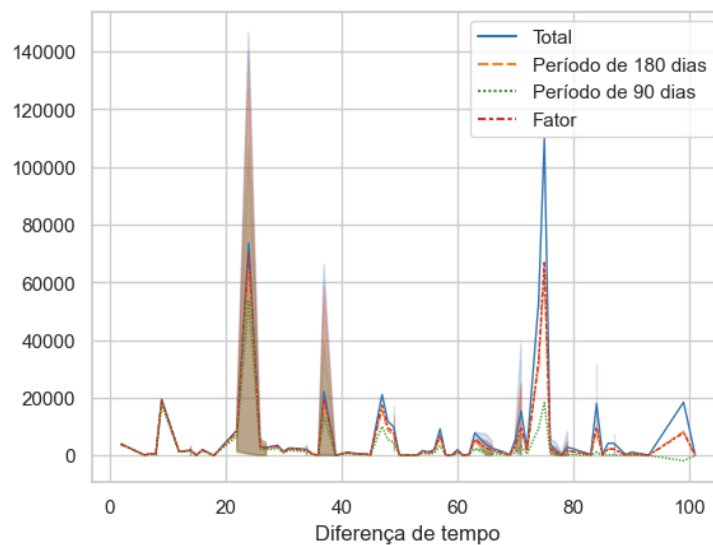


Figura 5.6: Comparação entre os valores de contribuição por log do colaborador efc78

Como pode ser observado nas figuras, quando o período de referência é atingido, o valor da contribuição por período se torna inversamente proporcional, resultando em uma redução significativa na contribuição total. Esse comportamento destaca a importância de utilizar um fator de decaimento exponencial para evitar a queda brusca na contribuição quando o período é excedido. Por outro lado, a contribuição por fator de decaimento exponencial se mantém em um meio termo, considerando um decaimento que se assemelha ao período de 180 dias. Isso proporciona uma avaliação mais suave e equilibrada da contribuição ao longo do tempo, mantendo uma proporção adequada entre as contribuições

mais recentes e as anteriores.

Esses resultados reforçam a eficácia do cálculo por fator de decaimento exponencial na avaliação da contribuição dos colaboradores ao longo do tempo. Ele permite uma transição gradual e mais precisa entre as contribuições, evitando flutuações bruscas e distorções nos resultados. Além disso, a comparação com as contribuições por período de referência destaca a importância de definir adequadamente o período de análise para garantir resultados coerentes e relevantes.

Portanto, as figuras fornecem uma representação visual clara do comportamento dos diferentes cálculos de contribuição ao longo do tempo, demonstrando a importância do uso do fator de decaimento exponencial para uma avaliação mais consistente e equilibrada da contribuição dos colaboradores.

5.3 Perspectiva do Projeto

A análise de cada colaborador dentro do projeto é organizada e catalogada de forma a fornecer uma visão estruturada e organizada das contribuições dos colaboradores, bem como a identificação do colaborador-chave responsável pelo projeto atualmente. Os resultados foram ordenados com base no valor obtido da contribuição por fator exponencial. Em seguida, é identificado o atual colaborador chave do projeto, representado pelo campo “Influencer Id”. A Figura 5.7 ilustra como os resultados são apresentados na plataforma.

User / Logs							
User Id	Logs	Contribuição	Normalização	Contribuição período	Normalização período	Contribuição fator	Normalização fator
7a40c	110	987.166	0,8701768180507967	741.264	1	792.038	1
58078	9	1.134.443	1	691.331	0,9326380344924345	738.903	0,9329135723286004
1bae3	49	1.083.552	0,9551400995907242	678.196	0,9149183017116709	699.987	0,883779566131928
e130e	41	953.625	0,8406107666934346	437.408	0,5900839646873448	428.782	0,5413654395369919
6b8ae	82	656.985	0,5791256149493628	309.015	0,4168757689568089	301.223	0,3803138233266586
3ebcf	42	562.714	0,4960266844610086	261.834	0,35322638088454317	253.907	0,3205742653761562
67ff5	1	300.201	0,26462413713161437	218.479	0,29473844676120786	243.609	0,30757236395223464
54e78	42	419.090	0,36942358496636674	212.972	0,287309244749509	221.421	0,27955855653390366

Figura 5.7: Resultados da análise de reputação no Projeto

Os resultados obtidos foram cuidadosamente organizados e apresentados na Tabela 5.4, que exibem as relações dos valores de contribuição por colaborador, e na Tabela 5.5 de valores normalizados. Ambas as tabelas foram ordenadas em função do valor da contribuição relativa por fator de decaimento exponencial.

Colaborador	Logs	Contribuição	C. Período	C. Exponencial
7a40c	110	987166	741264	792038
58078	9	1134443	691331	738903
1bae3	49	1083552	678196	699987
e130e	41	953625	437408	428782
6b8ae	82	656985	309015	301223
3ebcf	42	562714	261834	253907
67ff5	1	300201	218479	243609
54e78	42	419090	212972	221421
12f86	8	225435	177963	179663
194fc	71	205643	143279	156489
9095c	48	126791	100665	110054
f42e9	42	118379	103763	107500
5354b	1	80619	79275	80555
63283	4	105721	73115	73932
818fa	2	88174	56574	66964
ead3b	125	74654	50275	54185
67c81	40	59205	39414	42611
00000	5	364070	42747	40829
921b6	13	81919	31392	38164
124ec	3	168483	39935	36155
fba89	62	37990	26985	29952
a5c07	11	90465	10500	27684
5f2f5	1	32964	17580	17841
916ec	51	20499	15738	17557
efc78	2	1178	689	726
ee868	1	303	297	302
d083f	1	436	-234	0

Tabela 5.4: Tabela de dados de contribuição no projeto

Colaborador	Normalização	N. Período	N. Exponencial
7a40c	0.8701	1.0	1.0
58078	1.0	0.9326	0.9329
1bae3	0.9551	0.9149	0.8837
e130e	0.8406	0.5900	0.5413
6b8ae	0.5791	0.4168	0.3803
3ebcf	0.4960	0.3532	0.3205
67ff5	0.2646	0.2947	0.3075
54e78	0.3694	0.2873	0.2795
12f86	0.1987	0.2400	0.2268
194fc	0.1812	0.1932	0.1975
9095c	0.1117	0.1358	0.1389
f42e9	0.1043	0.1399	0.1357
5354b	0.0710	0.1069	0.1017
63283	0.0931	0.0986	0.0933
818fa	0.0777	0.0763	0.0845
ead3b	0.0658	0.0678	0.0684
67c81	0.0521	0.0531	0.0537
00000	0.3209	0.0576	0.0515
921b6	0.0722	0.0423	0.0481
124ec	0.1485	0.0538	0.0456
fba89	0.0334	0.0364	0.0378
a5c07	0.0797	0.0141	0.0349
5f2f5	0.0290	0.0237	0.0225
916ec	0.0180	0.0212	0.0221
efc78	0.0010	0.0009	0.0009
ee868	0.0002	0.0004	0.0003
d083f	0.0003	-0.000	0.0

Tabela 5.5: Tabela de dados normalizada da contribuição no projeto

A análise dos gráficos das Figuras 5.8 e 5.9, gerados a partir dos resultados normalizados, proporcionou uma visão abrangente das relações entre a contribuição total e a contribuição por período de referência, assim como sua relação com a contribuição por fator de decaimento. O eixo horizontal representa o valor normalizado da contribuição total do colaborador, enquanto o eixo vertical representa o valor normalizado da contribuição relativa, por fim o tamanho do ponto representa a quantidade total de *logs* do colaborador. Essas representações visuais permitiram uma comparação detalhada entre essas métricas.

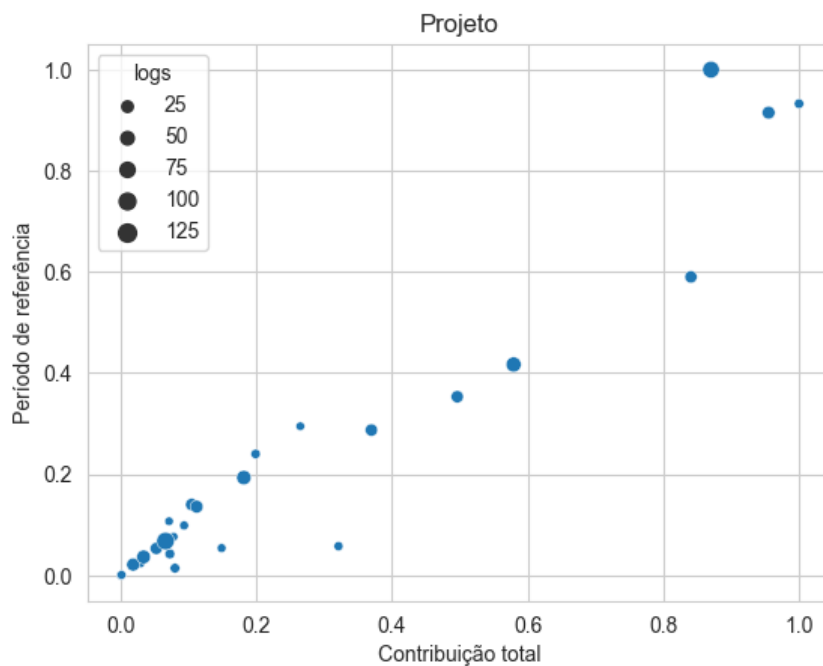


Figura 5.8: Análise de reputação no projeto por período de referência

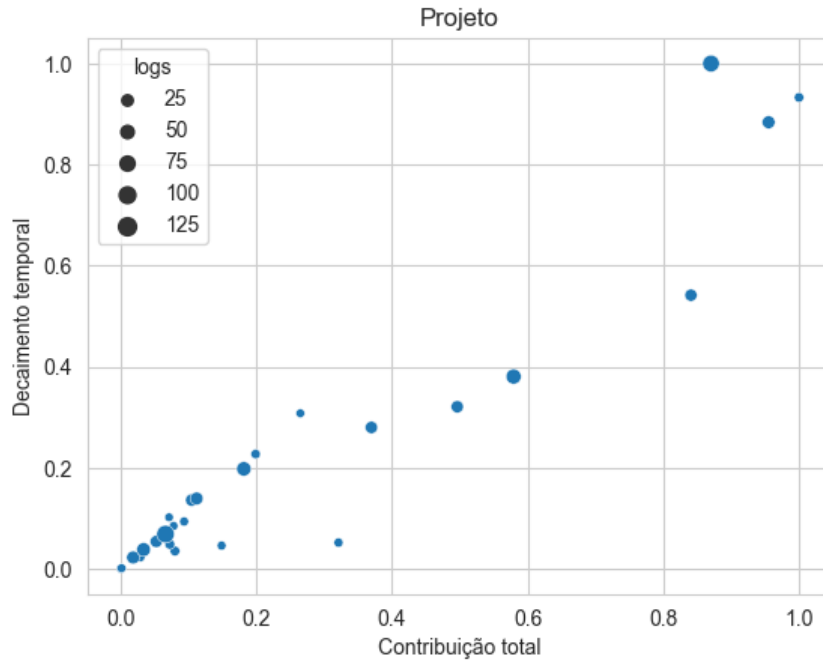


Figura 5.9: Análise de reputação no Projeto por fator exponencial

Ao examinar os gráficos em conjunto, podemos destacar que quando o resultado para o gráfico da Figura 5.8 não é negativo, os resultados obtidos se assemelham consideravelmente. Isso demonstra claramente a eficácia da aplicação da equação baseada no fator de decaimento para determinar a contribuição efetiva.

O principal fator analisado é que embora um colaborador possa apresentar a maior contribuição total para o sistema, isso não garante, necessariamente, que ele seja reconhecido como o colaborador-chave atual. Essa constatação ganha ainda mais relevância quando consideramos o período em que cada colaborador contribuiu. Em outras palavras, a longevidade e a consistência da contribuição de um colaborador podem desempenhar um papel fundamental na determinação do colaborador-chave.

Essas observações fornecem informações valiosas sobre a dinâmica da contribuição dentro do sistema em análise possibilitando a tomada de decisões embasadas e justas na identificação dos colaboradores mais influentes e estratégicos para o sucesso contínuo do projeto ou organização.

Uma característica notável e distinta do projeto em análise, é o fluxo de colaboradores que desempenharam um curto período de atuação no projeto. Isso se reflete na distribuição populacional representados no intervalo de 0.0 a 0.4, que tende a ser mais

numeroso circunstancialmente. Enquanto os líderes (*Team Tech Leads*) de cada setor de atuação, devido à sua contribuição contínua ao longo do tempo, assumem uma posição de maior destaque e influência, concentrando-se no canto superior direito dos gráficos.

5.4 Perspectiva do setor

De maneira análoga ao processo adotado na perspectiva do projeto, os setores individuais são reconhecidos e analisados de forma separada. Cada colaborador contribui de maneira específica em relação a um setor específico, permitindo a identificação da atuação de cada colaborador em determinado setor e sua respectiva contribuição. Cada setor possui sua própria relação de contribuição por colaborador, como demonstrado na Figura 5.10.



Setor		Influencer Id						
CRM		1bae3						
User / Logs								
User Id	Logs	Contribuição	Normalização	Contribuição período	Normalização período	Contribuição fator	Normalização fator	
1bae3	49	1.083.552	1	678.196	1	699.987	1	
e130e	41	953.625	0.8800915876672278	437.408	0.6449580947100838	428.782	0.6125570903459636	
6b8ae	82	656.985	0.6063253078763179	309.015	0.45564261658871474	301.223	0.43032656320760243	
67ff5	1	300.201	0.27705269336404714	218.479	0.32214728485570543	243.609	0.34801932035880667	
194fc	71	205.643	0.18978600011813	143.279	0.21126488507747024	156.489	0.22355986611179923	
f42e9	42	118.379	0.10925087120876524	103.763	0.1529985431940029	107.500	0.15357428066521236	

Figura 5.10: Resultado da análise de reputação dos colaboradores por setor

Com a análise completa da contribuição de cada colaborador, é mapeado a distribuição dos colaboradores, exibido pela Figura 5.11 que apresenta o diagrama que ilustra o resultado dessa relação.

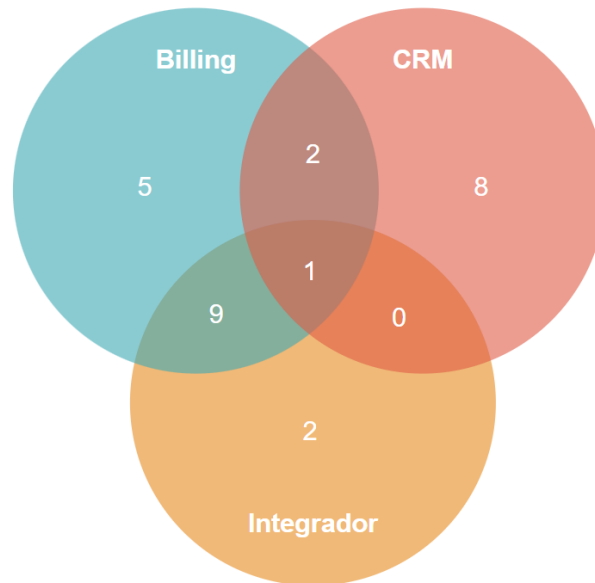


Figura 5.11: Diagrama de Venn de colaboradores por setor

É importante ressaltar que, devido a fatores como aumento das demandas nos *backlogs*, priorização de tarefas ou interdependências, é comum que um colaborador atue em mais de um setor ao longo do projeto. No caso específico do projeto em análise, o setor de integrador desempenha um papel fundamental na integração da plataforma com outros sistemas, conforme as necessidades do cliente. Em particular, esses colaboradores concentram suas atividades principalmente em integrações relacionadas ao *Billing*.

Ao realizar a aplicação de um filtro para considerar apenas os colaboradores atualmente envolvidos no projeto, é possível obter uma visão mais precisa da distribuição desses colaboradores nos diferentes setores de atuação. A Figura 5.12 representa essa distribuição ilustrativa. Essa filtragem permite focar na análise dos colaboradores que estão efetivamente contribuindo para o projeto no momento, possibilitando observar como os atuais colaboradores estão distribuídos entre os diversos setores, proporcionando uma compreensão mais atualizada da equipe e de suas atribuições.

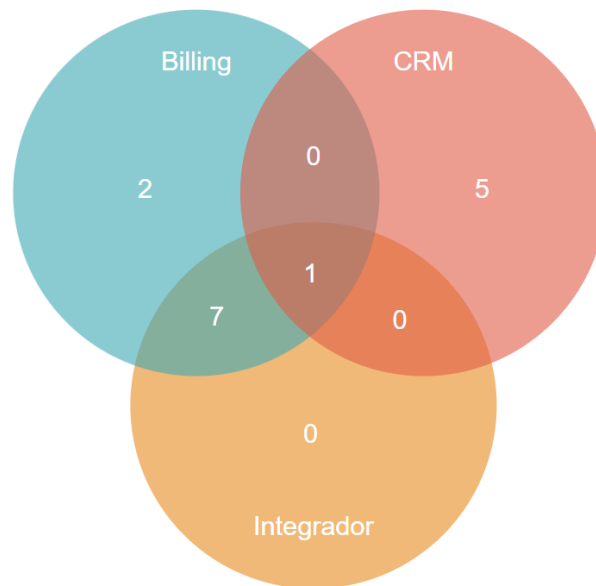


Figura 5.12: Diagrama de Venn de colaboradores atuais do projeto

Ao levar em consideração a distribuição dos atuais colaboradores pelos setores de atuação, os gestores e líderes de equipe podem tomar decisões mais embasadas, como identificar necessidades de treinamento, realocação de recursos ou contratação de novos colaboradores para suprir demandas específicas em cada setor.

A contribuição por fator de decaimento exponencial também é levada em consideração para determinar o colaborador-chave. Na Tabela 5.6, são apresentados os resultados obtidos pela análise do setor de CRM, enquanto na Tabela 5.7, são aplicados filtros para exibir apenas os colaboradores atuais do projeto em análise.

Colaborador	Normalização	N. Período	N. Exponencial
1bae3	1.0	1.0	1.0
e130e	0.8800	0.6449	0.6125
6b8ae	0.6063	0.4556	0.4303
67ff5	0.2770	0.3221	0.3480
194fc	0.1897	0.2112	0.2235
f42e9	0.1092	0.1529	0.1535
818fa	0.0755	0.0884	0.0956
124ec	0.1554	0.0588	0.0516
a5c07	0.0057	0.0065	0.0070
00000	0.0007	0.0007	0.0008
d083f	0.0004	-0.000	0.0

Tabela 5.6: Tabela de dados normalizada da contribuição no setor CRM

Colaborador	Normalização	N. Período	N. Exponencial
1bae3	1.0	1.0	1.0
e130e	0.8800	0.6449	0.6125
6b8ae	0.6063	0.4556	0.4303
194fc	0.1897	0.2112	0.2235
f42e9	0.1092	0.1529	0.1535
a5c07	0.0057	0.0065	0.0070

Tabela 5.7: Tabela de dados normalizada da contribuição por colaboradores ativos no setor CRM

Ao analisar os resultados, é possível observar que, caso não haja mudança no colaborador-chave em relação ao tempo total do projeto e considerando a filtragem dos colaboradores ativos, os valores normalizados não serão alterados, uma vez que o ponto de referência permanece o mesmo indivíduo. No entanto, se ocorrer uma troca do colaborador-chave para a mesma situação, os valores de comparação normalizados de todos os colaboradores serão recalculados, garantindo assim a atualização e confiabilidade da análise.

Novamente nota-se a valorização dos colaboradores ativos que estão envolvidos no projeto há mais tempo, ilustrado pelos gráficos nas Figuras 5.13, 5.14 e 5.15, que mostram a relação entre a contribuição total e o valor da contribuição com decaimento exponencial para cada setor avaliado. O eixo horizontal representa o valor normalizado da contribuição total do colaborador, enquanto o eixo vertical representa o valor normalizado da contribuição relativa por decaimento exponencial, por fim o tamanho do ponto representa a quantidade total de *logs* do colaborador.

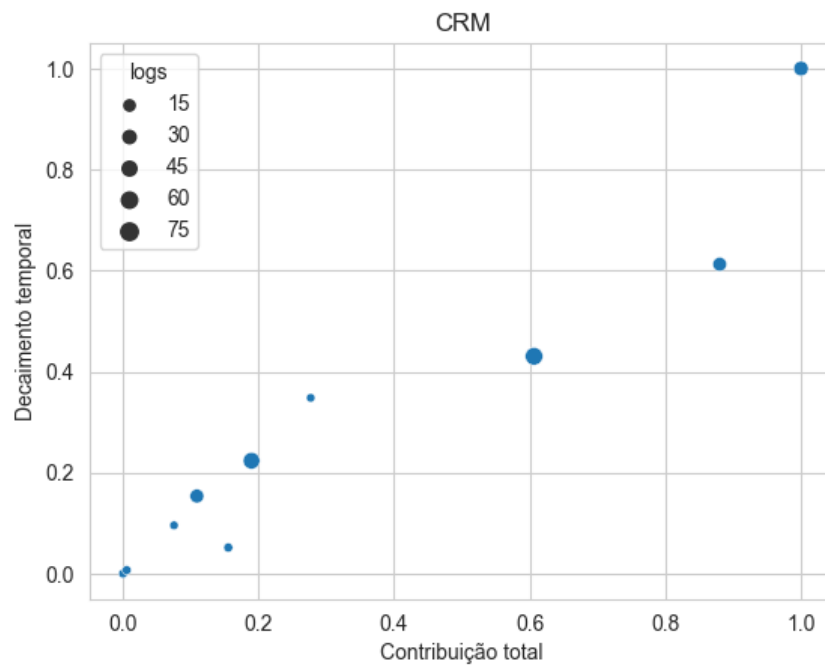


Figura 5.13: Análise de reputação do setor CRM com decaimento exponencial

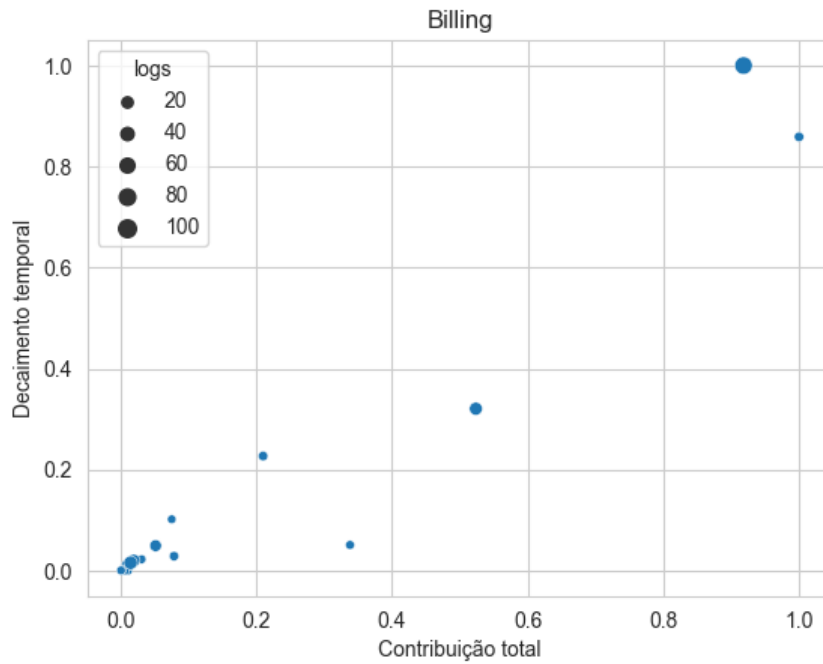


Figura 5.14: Análise de reputação do setor *Billing* com decaimento exponencial

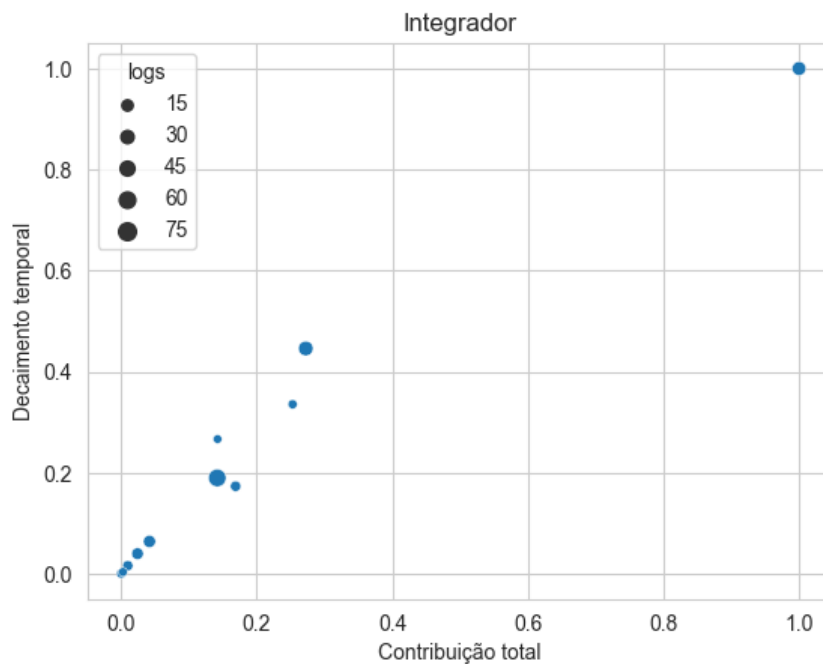


Figura 5.15: Análise de reputação do setor Integrador com decaimento exponencial

Essas figuras evidenciam como a contribuição dos colaboradores é ponderada ao longo do tempo, reconhecendo a importância da experiência e da continuidade no projeto. Colaboradores que têm participação consistente e duradoura são valorizados, demonstrando a influência positiva que seu conhecimento acumulado e envolvimento contínuo

podem ter no sucesso do projeto.

5.5 Validação dos resultados

Foi realizado um estudo por meio de uma pesquisa direcionada aos colaboradores, visando obter seus *feedbacks* referentes aos resultados alcançados através da análise do ecossistema. Essa pesquisa teve como propósito capturar a percepção dos colaboradores em relação à precisão e confiabilidade dos dados gerados por meio desse processo de análise.

Os *feedbacks* desempenham um papel essencial na validação e aprimoramento contínuo do sistema de análise, fornecendo *insights* valiosos que auxiliam na interpretação e aplicação adequada dos resultados obtidos, visto que cada projeto dentro do ecossistema possui suas próprias peculiaridades e, conseqüentemente, demandam ajustes e adaptações específicas. Essa compreensão aprofundada é essencial para aprimorar o sistema de análise, garantindo que as soluções e as decisões tomadas sejam adequadas e efetivas em relação às características específicas de cada contexto de trabalho.

A pesquisa levanta 3 questões respondidas por 9 colaboradores:

Q1: Os resultados analisados em relação à sua contribuição correspondem à sua realidade de atuação ao longo do tempo? Q2: Você concorda com os resultados obtidos em relação à seleção dos colaboradores-chave do projeto e do seu setor? Q3: Com base nas respostas anteriores, você considera o método utilizado de atribuição de valor da contribuição efetivo?

Ao analisar as respostas de cada questão individualmente, observamos o seguinte:

R1: a maioria dos colaboradores concordou com os resultados, afirmando que eles condizem com sua contribuição ao longo do tempo em cada setor. No entanto, um colaborador expressou discordância, mencionando que o resultado de seu setor foi definido erroneamente devido à limitação da amostra de dados analisados. Esse colaborador ressaltou que sua contribuição estava mais relacionada aos processos na plataforma, que não foram incluídos na análise. Essa observação levanta a necessidade de considerar uma ampliação da abordagem de análise para incluir esses aspectos relevantes.

R2: todos os colaboradores responderam positivamente, concordando com os resultados obtidos em relação à seleção dos colaboradores-chave do projeto e do setor. Eles

reconheceram que os colaboradores-chave eram aqueles que estavam ativamente envolvidos no projeto por um período mais longo e exerciam posições de liderança nos times. No entanto, foi mencionado que a presença de colaboradores inativos poderia ter menos influência no projeto, sugerindo assim uma possível adaptação na equação de cálculo da contribuição para levar em consideração essa variável.

R3: três colaboradores levantaram questões quanto à eficiência de atribuir os pontos de contribuição apenas com base na Distância de Levenshtein. Eles sugeriram que para alcançar maior precisão nos valores obtidos, seria necessário considerar outras metodologias, como filtragem de texto e lógica do código, para avaliar precisamente a contribuição entre diferentes níveis de conhecimento dos colaboradores. Apesar dessas sugestões de aprimoramento, eles concordaram que, para a avaliação apresentada, os resultados gerais se mostraram promissores para retratar as características reais e atuais da reputação individual no projeto.

5.6 Considerações Finais do Capítulo

A combinação dos gráficos das reputações dos colaboradores, juntamente com a pesquisa de confiabilidade e assertividade realizada com os atuais colaboradores, oferece uma visão completa e abrangente dos resultados da análise do ecossistema de trabalho. Essa abordagem integrada fornece uma base sólida para a tomada de decisões informadas e estratégicas em relação à gestão dos projetos e dos colaboradores envolvidos.

Os gráficos das reputações dos colaboradores proporcionam uma representação visual clara e intuitiva das contribuições individuais ao longo do tempo. Isso permite identificar padrões, tendências e discrepâncias na participação e no desempenho dos colaboradores. Com essa visualização, os gestores e as equipes podem avaliar facilmente o impacto de cada colaborador no sucesso do projeto e tomar medidas adequadas para otimizar a alocação de recursos e promover uma maior eficiência nas equipes de trabalho.

Além disso, a pesquisa de confiabilidade e assertividade realizada com os atuais colaboradores valida a precisão e a confiabilidade dos dados gerados pela análise do ecossistema. Essa pesquisa captura as percepções dos colaboradores em relação aos resultados obtidos, garantindo que os mesmos estejam alinhados com a realidade e que sejam

considerados uma fonte confiável de informações para embasar as decisões. Com bases nesse *feedback*, é possível ajustar e aprimorar continuamente o sistema de análise, a fim de torná-lo cada vez mais preciso e relevante para o contexto específico de trabalho.

6 Considerações Finais e Trabalhos Futuros

Este trabalho de conclusão de curso apresentou um processo para análise de um Ecosistema real, considerando uma empresa privada de desenvolvimento de software. Para propor esse processo, foram investigadas técnicas de análise de dados relacionadas a sistemas de reputação, que foram adotadas em ECOS de plataformas abertas. O processo tem como objetivo principal possibilitar a análise dos colaboradores do ECOS da empresa, destacando aqueles mais influentes, suas especialidades e seus impactos nos projetos.

Após a definição, implementação e implantação do processo na plataforma *SY-DLE ONE* da empresa foco deste estudo, foi realizada uma prova de conceito para avaliar a solução proposta. A solução foi desenvolvida e incorporada na plataforma de gestão de projetos da empresa, facilitando o seu uso por parte dos colaboradores. Os resultados obtidos apontam para a viabilidade da solução.

6.1 Contribuição

Este trabalho fornece uma nova perspectiva para o estudo de ECOS em ambientes empresariais para auxiliar a gestão de recursos humanos. As principais contribuições são:

- Fornecer evidências, baseadas na literatura técnica, sobre como os critérios de reputação exercem influência em um ecossistema. Ao examinar as interdependências e inter-relações entre os critérios de reputação, esta pesquisa aprimora o entendimento de como a reputação é formada e como ela impacta a dinâmica do ecossistema.
- A exploração da literatura fornece percepções sobre a interconexão dos critérios de reputação e oferece orientações valiosas para o desenvolvimento de metodologias robustas e abrangentes de mensuração da reputação. A pesquisa identifica abordagens e técnicas eficazes para quantificar o valor da reputação no contexto do ecossistema, propondo um conjunto de métricas capazes de se adaptar a características do ecossistema.

- Propõe um processo que fornece uma estrutura robusta e sistemática para avaliar e analisar os dados de ecossistemas centrados na plataforma *SYDLE ONE*. Sua adaptabilidade permite sua aplicação em outros ecossistemas, promovendo uma abordagem eficiente e escalável para a avaliação de dados em diferentes contextos.
- Apresentação de uma prova de conceito e a execução de um experimento utilizando o processo proposto, mediante uma amostra única de um projeto privado, que mostrou ser de difícil acesso em trabalhos relacionados. A análise realizada fornece informações essenciais para validar o processo e avaliar sua aplicabilidade no auxílio da gestão de recursos humanos dentro da empresa. Essa amostra permite uma visão inicial do potencial do processo em fornecer *insights* relevantes, através da identificação de colaboradores-chave e a avaliação da contribuição individual dos colaboradores. Essa validação inicial é fundamental para orientar melhorias futuras e garantir que o processo seja eficaz e adequado às demandas e objetivos da organização.
- A aplicação comercial do processo proposto nesse mercado em crescimento traz novas oportunidades de negócio e desenvolvimento de soluções inovadoras. Através dele, empresas podem adotar uma abordagem mais precisa e eficiente na gestão de recursos humanos, utilizando dados e análises para otimizar a alocação de talentos, identificar colaboradores-chave e promover a colaboração e inovação dentro do ecossistema. Essa análise promove a discussão sobre a comparação dos resultados obtidos em ecossistemas abertos e privados, abrindo novas possibilidades para aplicações comerciais.

6.2 Limitações

A principal limitação desta pesquisa está relacionada à fonte dos dados amostrais, devido à complexidade de adaptação interna da plataforma, tanto em termos de diferentes projetos, quanto da análise completa de um único projeto. Essas complexidades exigem esforços adicionais para coletar e analisar dados de forma abrangente e representativa.

No aspecto interno de um projeto, a limitação se deve à estrutura da arquitetura.

Como a plataforma apresenta uma característica de desenvolvimento *low code*, o sistema de versionamento é individual a cada classe, e não há um registro detalhado das alterações em cada versão. Portanto, foi necessário estabelecer um sistema de identificação e filtragem dos dados referentes às alterações entre as versões. Nesta pesquisa, apenas o critério das classes foi analisado, porém, a plataforma também inclui o desenvolvimento de processos internos e outros serviços, o que requer uma adaptação e um estudo mais aprofundado em colaboração com a organização responsável pelo desenvolvimento da plataforma, a fim de viabilizar um sistema semelhante de coleta de dados. Como resultado, também identificado nos *feedbacks* coletados dos colaboradores do projeto, não foi possível abranger toda a gama de reputação do projeto, o que diminui a precisão e limita resultados encontrados.

Quanto ao aspecto da análise de múltiplos projetos, a limitação significativa está relacionada à natureza individual de cada projeto. A forma como as classes ou objetos na plataforma estabelecem uma conexão abstrata com cada setor de atuação, requer uma compreensão aprofundada do contexto de cada projeto para uma avaliação precisa da contribuição dos colaboradores. Compreensão que só pode ser obtida com consulta e coleta de informações com os colaboradores ativos do projeto. Essa complexidade individualizada dos projetos dificulta a generalização e aplicação direta do processo proposto. O processo é flexível e adaptável, portanto, para superar essa limitação, é fundamental investir uma comunicação contínua com os colaboradores do projeto, para definir a melhor configuração de análise em cada projeto.

6.3 Trabalhos futuros

Embora este estudo tenha contribuído significativamente ao propor um processo para compreensão da reputação em um ECOS, existem várias áreas que poderiam ser exploradas em trabalhos futuros. Algumas sugestões para trabalhos futuros são:

- Aprimorar e comparar as métricas do sistema de reputação: É recomendado aprimorar o sistema de métricas utilizado no processo de avaliação da reputação, a fim de reconhecer o valor real de impacto em cada contribuição. O estudo atual propôs o uso da Distância de Levenshtein como método de cálculo da reputação

que consideram apenas as alterações, remoções e adições de caracteres nos *scripts* de desenvolvimento, mas não levam em conta aspectos como a lógica e complexidade do código. No entanto, é importante investigar e comparar outros métodos de cálculo para verificar sua eficácia e precisão na avaliação da reputação em um ECOS. Uma abordagem mais abrangente que considere esses fatores poderia fornecer uma avaliação mais precisa e completa da contribuição dos colaboradores.

- Investigar outras dimensões de avaliação da reputação: Além das métricas de código, é interessante explorar outras dimensões que podem influenciar a reputação dos colaboradores em um ECOS. Por exemplo, considerar a qualidade das interações sociais, a participação em discussões relevantes, a capacidade de resolver problemas e a habilidade de compartilhar conhecimento. Essas dimensões podem ser incorporadas ao sistema de avaliação da reputação, proporcionando uma visão mais abrangente do desempenho dos colaboradores.
- Relacionar ECOS públicos e privados: Ao realizar uma análise comparativa, é possível obter validações quanto à aplicabilidade de metodologias desenvolvidas em um ambiente e sua transferência para o outro. Essa abordagem permitirá identificar os princípios fundamentais que são comuns a ambos os contextos, bem como os desafios específicos enfrentados em cada um deles. Além disso, essa pesquisa pode contribuir para a troca de conhecimento entre os domínios público e privado, enriquecendo a compreensão dos ECOS como um todo.
- Abordagem completa da plataforma SYDLE ONE: Realizar as adaptações necessárias para a coleta de dados da parte da estrutura da plataforma que ainda não foi contemplada. Isso envolveria, por exemplo, a inclusão dos processos interno do projeto, permitindo uma análise mais completa da contribuição dos colaboradores. Com essa abordagem aprimorada, seria possível obter uma visão mais abrangente e precisa.

Esses trabalhos futuros podem fornecer avanços significativos no campo dos ECOS, contribuindo para a validação de metodologias, a compreensão das peculiaridades dos ambientes públicos e privados, bem como uma análise abrangente e precisa da reputação em um ECOS.

Referências

- ARAKAKI, M. G. *UM PROCESSO PARA O DESENVOLVIMENTO DE COMPONENTES DE COLABORAÇÃO PARA UM ECOSISTEMA DE E-SCIENCE*. Tese (Doutorado) — Federal University of Juiz de Fora, 2016. Available at <http://monografias.ice.ufjf.br/tcc-web/tcc?id=246>.
- AUDY, J. L. N. *Desenvolvimento distribuído de software*. [S.l.]: Elsevier, 2007.
- BADASHIAN, A. S.; ESTEKI, A.; GHOLIPOUR, A.; HINDLE, A.; STROULIA, E. Involvement, contribution and influence in github and stack overflow. In: *CASCON*. [S.l.: s.n.], 2014. v. 14, p. 19–33.
- BARBOSA, O.; SANTOS, R. P. dos; ALVES, C.; WERNER, C.; JANSEN, S. A systematic mapping study on software ecosystems from a three-dimensional perspective. *Software ecosystems*, Edward Elgar Publishing, 2013.
- BERK, I. V. D.; JANSEN, S.; LUINENBURG, L. Software ecosystems: a software ecosystem strategy assessment model. In: *Proceedings of the fourth european conference on software architecture: Companion volume*. [S.l.: s.n.], 2010. p. 127–134.
- BLINCOE, K.; HARRISON, F.; DAMIAN, D. Ecosystems in github and a method for ecosystem identification using reference coupling. In: IEEE. *2015 IEEE/ACM 12th Working Conference on Mining Software Repositories*. [S.l.], 2015. p. 202–211.
- BOEHM, B. A view of 20th and 21st century software engineering. In: *Proceedings of the 28th international conference on Software engineering*. [S.l.: s.n.], 2006. p. 12–29.
- BOSCH, J. From software product lines to software ecosystems. In: *SPLC*. [S.l.: s.n.], 2009. v. 9, p. 111–119.
- BRINKKEMPER, S.; SOEST, I. v.; JANSEN, S. Modeling of product software businesses: Investigation into industry product and channel typologies. In: *Information Systems Development*. [S.l.]: Springer, 2009. p. 307–325.
- CAMPBELL, P. R.; AHMED, F. A three-dimensional view of software ecosystems. In: *Proceedings of the Fourth European Conference on Software Architecture: Companion Volume*. [S.l.: s.n.], 2010. p. 81–84.
- FUKS, H.; RAPOSO, A. B.; GEROSA, M. A.; LUCENA, C. J. P. Do modelo de colaboração 3c à engenharia de groupware. *Simpósio Brasileiro de Sistemas Multimídia e Web-Webmidia*, p. 0–8, 2003.
- GUERCIO, H. *sSECO-Process: Avaliando a Dimensão Social em Ecossistemas de Software*. Tese (Doutorado) — Ph. D. Dissertation. Universidade Federal de Juiz de Fora, 2018.
- GUÉRCIO, H.; STRÖELE, V.; DAVID, J. M. N.; BRAGA, R.; CAMPOS, F. Topological analysis in scientific social networks to identify influential researchers. In: IEEE. *2017 IEEE 21st International Conference on Computer Supported Cooperative Work in Design (CSCWD)*. [S.l.], 2017. p. 287–292.

- GUERCIO, H.; STROELE, V.; DAVID, J. M. N.; BRAGA, R.; CAMPOS, F. Complex network analysis in a software ecosystem: Studying the eclipse community. In: IEEE. *2018 IEEE 22nd International Conference on Computer Supported Cooperative Work in Design ((CSCWD))*. [S.l.], 2018. p. 618–623.
- HAGEL, J.; BROWN, J. S.; DAVISON, L. Shaping strategy in a world of constant disruption. *Harvard Business Review*, v. 86, n. 10, p. 80–89, 2008.
- HENDRIKX, F.; BUBENDORFER, K.; CHARD, R. Reputation systems: A survey and taxonomy. *Journal of Parallel and Distributed Computing*, Elsevier, v. 75, p. 184–197, 2015.
- IANSITI, M.; LEVIEN, R. *The keystone advantage: what the new dynamics of business ecosystems mean for strategy, innovation, and sustainability*. [S.l.]: Harvard Business Press, 2004.
- JANSEN, S.; FINKELSTEIN, A.; BRINKKEMPER, S. A sense of community: A research agenda for software ecosystems. In: IEEE. *2009 31st International Conference on Software Engineering-Companion Volume*. [S.l.], 2009. p. 187–190.
- JØSANG, A.; ISMAIL, R.; BOYD, C. A survey of trust and reputation systems for online service provision. *Decision support systems*, Elsevier, v. 43, n. 2, p. 618–644, 2007.
- JØSANG, A.; QUATTROCIOCCI, W. Advanced features in bayesian reputation systems. In: SPRINGER. *Trust, Privacy and Security in Digital Business: 6th International Conference, TrustBus 2009, Linz, Austria, September 3-4, 2009. Proceedings 6*. [S.l.], 2009. p. 105–114.
- KIPPER, M. M. Proposta de metodologia de engenharia de domínio para o desenvolvimento de sistemas de automação industrial. 2010.
- LÉLIS, C. A. S. *Um modelo dinâmico de reputação para apoiar a manutenção colaborativa de software*. Tese (Doutorado) — Ph. D. Dissertation. Universidade Federal de Juiz de Fora, 2017.
- LIMA, T. M. P. Uma abordagem socio-técnica para apoiar modelagem e análise de ecossistemas de software. *Proj. Final. Curso Eng. Comput. e Informação. UFRJ*, p. 83, 2015.
- MANIKAS, K.; HANSEN, K. M. Software ecosystems—a systematic literature review. *Journal of Systems and Software*, Elsevier, v. 86, n. 5, p. 1294–1306, 2013.
- RAPPORT, D. J.; COSTANZA, R.; MCMICHAEL, A. J. Assessing ecosystem health. *Trends in ecology & evolution*, Elsevier, v. 13, n. 10, p. 397–402, 1998.
- SANTOS, R.; VIANA, D. Ecossistemas de software no desenvolvimento de plataformas para web, redes sociais e multimídia. *Sociedade Brasileira de Computação*, 2016.
- SANTOS, R. P. dos. *Managing and monitoring software ecosystem to support demand and solution analysis*. Tese (Doutorado) — Ph. D. Dissertation. Universidade Federal do Rio de Janeiro, 2016.
- SANTOS, R. P. dos; WERNER, C. M. L. A proposal for software ecosystems engineering. In: *IWSECO@ IC SOB*. [S.l.: s.n.], 2011. p. 40–51.

SILVA, T. L. *PNETSCAN: UMA ABORDAGEM DE AGRUPAMENTO EM PARALELO PARA RESOLUÇÃO DE PROBLEMAS BIG DATA*. Tese (Doutorado) — Federal University of Juiz de Fora, 2018. Available at <http://monografias.ice.ufjf.br/tcc-web/tcc?id=407>.

YUJIAN, L.; BO, L. A normalized levenshtein distance metric. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 29, n. 6, p. 1091–1095, 2007.