

Princípios e Aplicações da Computação no Brasil 2

Ernane Rosa Martins
(Organizador)



Atena
Editora

Ano 2019

Ernane Rosa Martins

(Organizador)

**Princípios e Aplicações da Computação
no Brasil
2**

Atena Editora
2019

2019 by Atena Editora

Copyright © da Atena Editora

Editora Chefe: Profª Drª Antonella Carvalho de Oliveira

Diagramação e Edição de Arte: Geraldo Alves e Natália Sandrini

Revisão: Os autores

Conselho Editorial

- Prof. Dr. Alan Mario Zuffo – Universidade Federal de Mato Grosso do Sul
Prof. Dr. Álvaro Augusto de Borba Barreto – Universidade Federal de Pelotas
Prof. Dr. Antonio Carlos Frasson – Universidade Tecnológica Federal do Paraná
Prof. Dr. Antonio Isidro-Filho – Universidade de Brasília
Profª Drª Cristina Gaio – Universidade de Lisboa
Prof. Dr. Constantino Ribeiro de Oliveira Junior – Universidade Estadual de Ponta Grossa
Profª Drª Daiane Garabeli Trojan – Universidade Norte do Paraná
Prof. Dr. Darllan Collins da Cunha e Silva – Universidade Estadual Paulista
Profª Drª Deusilene Souza Vieira Dall’Acqua – Universidade Federal de Rondônia
Prof. Dr. Eloi Rufato Junior – Universidade Tecnológica Federal do Paraná
Prof. Dr. Fábio Steiner – Universidade Estadual de Mato Grosso do Sul
Prof. Dr. Gianfábio Pimentel Franco – Universidade Federal de Santa Maria
Prof. Dr. Gilmei Fleck – Universidade Estadual do Oeste do Paraná
Profª Drª Girlene Santos de Souza – Universidade Federal do Recôncavo da Bahia
Profª Drª Ivone Goulart Lopes – Istituto Internazionele delle Figlie de Maria Ausiliatrice
Profª Drª Juliane Sant’Ana Bento – Universidade Federal do Rio Grande do Sul
Prof. Dr. Julio Candido de Meirelles Junior – Universidade Federal Fluminense
Prof. Dr. Jorge González Aguilera – Universidade Federal de Mato Grosso do Sul
Profª Drª Lina Maria Gonçalves – Universidade Federal do Tocantins
Profª Drª Natiéli Piovesan – Instituto Federal do Rio Grande do Norte
Profª Drª Paola Andressa Scortegagna – Universidade Estadual de Ponta Grossa
Profª Drª Raissa Rachel Salustriano da Silva Matos – Universidade Federal do Maranhão
Prof. Dr. Ronilson Freitas de Souza – Universidade do Estado do Pará
Prof. Dr. Takeshy Tachizawa – Faculdade de Campo Limpo Paulista
Prof. Dr. Urandi João Rodrigues Junior – Universidade Federal do Oeste do Pará
Prof. Dr. Valdemar Antonio Paffaro Junior – Universidade Federal de Alfenas
Profª Drª Vanessa Bordin Viera – Universidade Federal de Campina Grande
Profª Drª Vanessa Lima Gonçalves – Universidade Estadual de Ponta Grossa
Prof. Dr. Willian Douglas Guilherme – Universidade Federal do Tocantins

Dados Internacionais de Catalogação na Publicação (CIP) (eDOC BRASIL, Belo Horizonte/MG)

P957 Princípios e aplicações da computação no brasil 2 [recurso eletrônico] / Organizador Ernane Rosa Martins. – Ponta Grossa (PR): Atena Editora, 2019. – (Princípios e aplicações da computação no brasil; v. 2)

Formato: PDF

Requisito de sistema: Adobe Acrobat Reader

Modo de acesso: World Wide Web

Inclui bibliografia

ISBN 978-85-7247-048-3

DOI 10.22533/at.ed.483191601

1. Computação. 2. Informática. 3. Programação de computador.
I. Martins, Ernane Rosa. II. Título. III. Série.

CDD 004

Elaborado por Maurício Amormino Júnior – CRB6/2422

O conteúdo dos artigos e seus dados em sua forma, correção e confiabilidade são de responsabilidade exclusiva dos autores.

2019

Permitido o download da obra e o compartilhamento desde que sejam atribuídos créditos aos autores, mas sem a possibilidade de alterá-la de nenhuma forma ou utilizá-la para fins comerciais.

www.atenaeditora.com.br

APRESENTAÇÃO

O volume 2 desta obra aborda mais 16 capítulos sobre o panorama atual da computação no Brasil. Tendo como alguns dos assuntos abordados nos capítulos: ensino de raciocínio lógico, desenvolvimento de sistema computacional, micromobilidade em redes sem fio, usabilidade e acessibilidade de sistemas, qualidade da informação, tecnologias de análise de aprendizagem, redes neurais artificiais, análise de vibração, algoritmos evolucionários, sistemas inteligentes e acessibilidade móvel.

Deste modo, esta obra reúne debates e análises acerca de questões relevantes, tais como: Como está o estado da arte da análise de aprendizagem preditiva, nova proposta de um framework para previsão de desempenhos em programação e quais os caminhos para avançar nessas pesquisas? É possível realizar uma modelagem computacional, analisando os parâmetros espaciais relevantes na tomada de decisão, utilizando técnicas de redes neurais artificiais? Quais são os principais desafios, no cenário nacional, a fim de estabelecer e manter um Sistema de Gestão de Segurança da Informação? Uma proposta de um agente testador que realiza busca local no espaço de estados de casos de teste orientado por utilidade e que utiliza os algoritmos evolucionários multiobjetivos, NSGAI, SPEA2, PAES e MOCeII pode identificar quais deles são mais eficientes na geração de casos de testes para agentes racionais? Como realizar uma pesquisa científica que identifique os requisitos desejáveis para desenvolver uma aplicação móvel touch screen, que vise auxiliar a alfabetização de deficientes visuais?

Nesse sentido, este material tem grande relevância por constituir-se numa coletânea de referência para pesquisas e estudos da computação, tendo como objetivo reunir trabalhos acadêmicos que permitam contribuir com análises e discussões sobre assuntos pertinentes à área. Os organizadores da Atena Editora, agradecem especialmente aos autores dos diversos capítulos apresentados, parabenizam a dedicação e esforço de cada um, os quais viabilizaram a construção dessa obra no viés da temática apresentada. Por fim, desejamos aos leitores que esta obra, seja de extrema importância para todos que vierem a utilizá-la.

Ernane Rosa Martins

SUMÁRIO

CAPÍTULO 1	1
ENSINO DE RACIOCÍNIO LÓGICO E COMPUTAÇÃO PARA CRIANÇAS: EXPERIÊNCIAS, DESAFIOS E POSSIBILIDADES (XXXVII CONGRESSO DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO 250 WEI - WORKSHOP SOBRE EDUCAÇÃO EM COMPUTAÇÃO)	
<i>Thâmillys Marques de Oliveira</i> <i>Willmara Marques Monteiro</i> <i>Fábio Cristiano Souza Oliveira</i> <i>Danielle Juliana Silva Martins</i> <i>Alessandra da Silva Luengo Latorre</i>	
DOI 10.22533/at.ed.4831916011	
CAPÍTULO 2	12
DESENVOLVIMENTO DE SISTEMA COMPUTACIONAL PARA AQUISIÇÃO E ANÁLISE DE DADOS AMBIENTAIS REMOTAMENTE.	
<i>Jucivaldo Araujo Ferreira Junior</i> <i>Rardiles Branches Ferreira</i> <i>Rodrigo Da Silva</i> <i>Julio Tota da Silva</i> <i>Samuel Alves de Souza</i>	
DOI 10.22533/at.ed.4831916012	
CAPÍTULO 3	19
CARACTERIZAÇÃO DA MICROMOBILIDADE EM REDES SEM FIO INFRAESTRUTURADAS PELA VARIAÇÃO DA RELAÇÃO SINAL-RUÍDO	
<i>Kerlla Souza Luz Prates</i> <i>Priscila América Solís Mendez</i> <i>Barreto Henrique Domingues Garcia</i> <i>Mylène Christine Queiroz de Farias</i>	
DOI 10.22533/at.ed.4831916013	
CAPÍTULO 4	30
AVALIAÇÃO DE USABILIDADE E ACESSIBILIDADE DO SISTEMA DE GERENCIAMENTO DE REFEITÓRIOS DO IFPI – CAMPUS FLORIANO	
<i>Samuel de Araújo Fonseca</i> <i>Antonio Rodrigues de Araújo Costa</i> <i>Neto Carlos Eduardo Moreira Borges</i> <i>Hugo Araújo Gonçalves</i> <i>Paulo Miranda e Silva Sousa</i> <i>Rennê Stephany Ferreira dos Santos</i>	
DOI 10.22533/at.ed.4831916014	
CAPÍTULO 5	39
AVALIAÇÃO DA APREENSIBILIDADE E DA QUALIDADE DA INFORMAÇÃO EM SAÚDE COM O SOFTWARE SPINEFIND	
<i>Carine Geltrudes Webber</i> <i>Asdrubal Falavigna</i> <i>Caio Rodrigues da Silva</i> <i>Marco Antonio Koff</i> <i>Natália Lisboa</i>	
DOI 10.22533/at.ed.4831916015	

CAPÍTULO 6 54

AS TECNOLOGIAS DE ANÁLISE DE APRENDIZAGEM E OS DESAFIOS DE PREVER DESEMPENHOS DE ESTUDANTES DE PROGRAMAÇÃO

Márcia Gonçalves de Oliveira

DOI 10.22533/at.ed.4831916016

CAPÍTULO 7 67

ANÁLISE E MODELAGEM DA RELAÇÃO INTERPESSOAL EM ESPORTES COLETIVOS UTILIZANDO REDES NEURAIS ARTIFICIAIS

Tadeu Nogueira Costa de Andrade

Marcos Rodrigo Trindade Pinheiro

Menuchi Paulo Eduardo Ambrósio

DOI 10.22533/at.ed.4831916017

CAPÍTULO 8 75

ANÁLISE DOS DESAFIOS PARA ESTABELECEER E MANTER SISTEMA DE GESTÃO DE SEGURANÇA DA INFORMAÇÃO NO CENÁRIO BRASILEIRO

Rodrigo Valle Fazenda

Leonardo Lemes Fagundes

DOI 10.22533/at.ed.4831916018

CAPÍTULO 9 87

ANÁLISE DE VIBRAÇÃO COM CONTROLE DE MEDIÇÃO UTILIZANDO O FILTROS ESTATÍSTICOS

Karla Melissa dos Santos Leandro

Iago Ferreira Lima

Werley Rafael da Silva

Marco Paulo Guimarães

Marcos Napoleão Rabelo

DOI 10.22533/at.ed.4831916019

CAPÍTULO 10 96

ANÁLISE DE REDE COLABORAÇÃO CIENTÍFICA COMO FERRAMENTA NA GESTÃO DE PROGRAMAS DE PÓS-GRADUAÇÃO

Aurelio R. Costa

Celia Ghedini Ralha

DOI 10.22533/at.ed.48319160110

CAPÍTULO 11 109

ALGORITMOS EVOLUCIONÁRIOS MULTI OBJETIVOS PARA A SELEÇÃO DE CASOS DE TESTE PARA SISTEMAS INTELIGENTES

Daniel Victor Saraiva

Francisca Raquel de Vasconcelos Silveira

DOI 10.22533/at.ed.48319160111

CAPÍTULO 12 124

ACESSIBILIDADE MÓVEL PARA ALFABETIZAÇÃO DE DEFICIENTES VISUAIS: PROPOSTA INICIAL DE UM PROTÓTIPO

Jenifer Melissa de Paula

José Valter Amaral de Freitas

Thatiane de Oliveira Rosa

DOI 10.22533/at.ed.48319160112

CAPÍTULO 13..... 129

AÇÃO PARA INCENTIVAR MENINAS DO ENSINO MÉDIO A CURSAR CARREIRAS TECNOLÓGICAS DA UNIVERSIDADE FEDERAL DE RIO GRANDE DO NORTE

Idalmis Milián Sardina
Cristiano Maciel
Midori Hijjoka Camelo
Hortensia Sardina Miranda

DOI 10.22533/at.ed.48319160113

CAPÍTULO 14..... 137

A TÉCNICA OC2-RD2 COMO UMA PRÁTICA METODOLÓGICA PARA O ENSINO DE PROGRAMAÇÃO DE COMPUTADORES

Karina Buttignon
Ítalo Santiago Vega
Jonhson de Tarso Silva
Adriano Carlos Moraes Rosa

DOI 10.22533/at.ed.48319160114

CAPÍTULO 15..... 149

A DECADE OF SOFTWARE ENGINEERING BEST PRACTICES ADOPTION IN SMALL COMPANIES:
A QUASI-SYSTEMATIC MAPPING

Alex Juvêncio Costa
Juliana De Albuquerque Gonçalves
Saraiva Yuska Paola Costa Aguiar

DOI 10.22533/at.ed.48319160115

CAPÍTULO 16..... 162

INVENTORYIOT I²OT: UMA PLATAFORMA DE GERENCIAMENTO AUTOMATIZADO DE INVENTÁRIO

Jauberth Weyll Abijaude
Péricles de Lima Sobreira
Aprígio Augusto Lopes Bezerra
Fabiola Greve

DOI 10.22533/at.ed.48319160116

SOBRE O ORGANIZADOR 177

ANÁLISE DE REDE COLABORAÇÃO CIENTÍFICA COMO FERRAMENTA NA GESTÃO DE PROGRAMAS DE PÓS-GRADUAÇÃO

Aurelio R. Costa

Departamento de Ciencia da Computação
Instituto de Ciencias Exatas
Universidade de Brasília
arcosta@gmail.com

Celia Ghedini Ralha

Departamento de Ciencia da Computação
Instituto de Ciencias Exatas
Universidade de Brasília
ghedini@unb.br

RESUMO: O desempenho de um programa de pós-graduação é aferido pela CAPES em parte pelo seu nível de publicações, para tanto, faz-se necessário que a gestão do programa seja feita de forma a maximizar a qualidade das publicações que são realizadas pelos pesquisadores associados. No contexto dos relacionamentos de coautoria em publicações, a análise de rede se mostra uma ferramenta adequada para avaliar as relações já formadas, bem como para estimular a formação de novas relações. Neste artigo é apresentada uma ferramenta de análise de rede aplicada à gestão do programa de mestrado acadêmico de uma Universidade brasileira. Os dados coletados do programa de pós-graduação foram modelados em um banco NoSQL orientado a grafo incluindo todas as publicações dos docentes vinculados. Ressaltamos a utilidade

da ferramenta na medida em que apresenta um módulo de recomendação de parcerias, o qual pode ser utilizado pelo gestor do programa bem como pelos demais pesquisadores. O desenvolvimento da pesquisa utilizou a metodologia *Design Science Research* para guiar tanto a construção da ferramenta quanto para elaborar a documentação associada. Como resultado, constatou-se que, mesmo utilizando apenas os dados do programa isolado das demais instituições, há um grande potencial para realização de novas parcerias de coautoria.

PALAVRAS-CHAVE: Sistemas de recomendação, Predição de enlace, NoSQL, Análise de rede sociais, Design Science Research.

ABSTRACT: The performance of a graduate program is assessed by CAPES partly by their level of publications, therefore, it is necessary that the program management is made to maximize the quality of publications that are carried out by research associates. In the context of co-authoring relationships in publications, network analysis shown an appropriate tool to evaluate the relationships already formed, and to stimulate the formation of new relationships. This paper presents a network analysis tool applied to the management of an academic master program of a Brazilian university. The

data collected from the graduate program were modeled in a NoSQL graph oriented database including all publications of the linked professors. We emphasize the usefulness of the tool in as much as it has a partner recommendation module, which can be used by the program manager and by other researchers. The development of the research used the *Design Science Research* methodology to guide both the construction of the tool and to develop associated documentation. As a result, it was found that even using only the data from the program isolated from other institutions, there is great potential for new co-authoring partnerships.

KEYWORDS: Recommending Systems, Link Prediction, NoSQL, Social Network Analysis, Design Science Research.

1 | INTRODUÇÃO

Um dos fundamentos da gestão de um programa de pós-graduação baseia-se na maximização dos indicadores de produção acadêmica, especialmente a publicações em periódicos bem qualificados com abrangência internacional. Os indicadores de produção científica e acadêmica são elementos fundamentais nas avaliações que são realizadas pela CAPES¹, fundação vinculada ao ministério da educação responsável por avaliar os programas de pós-graduação em todos os estados da Federação. Nesse sentido, se faz necessário que os gestores desses programas disponham de uma ferramenta que os permita visualizar a distribuição das publicações que foram realizadas pelos pesquisadores. Tal ferramenta auxiliaria, dessa forma, no direcionamento de ações objetivando a melhoria dos indicadores de produção científica e acadêmica.

Este trabalho objetiva apresentar uma ferramenta de auxílio à gestão dos indicadores de produção acadêmica de um programa de pós-graduação através da análise da rede de colaboração científica formada pelos pesquisadores vinculados ao programa, assim como detalhar a implementação da funcionalidade de recomendação de parcerias nas futuras publicações.

Como trabalhos correlatos destacamos o realizado em [1] o qual apresenta uma ferramenta para análise das informações de publicação dos pesquisadores vinculados ao Instituto Nacional de Ciência e Tecnologia para Controle do Câncer (INCTCC) que embora desenvolva um ambiente no qual seja possível a análise de produção científica, não implementa a funcionalidade de recomendação de parcerias. O trabalho realizado em [2] propõe uma modificação na plataforma de rede de colaboração utilizada pelo C.E.S.A.R. de forma a dota-la de um componente de recomendação de conteúdo/ pessoas e serviu como ponto de partida para o desenvolvimento do módulo de recomendação do presente trabalho. O trabalho realizado em [3] aplica a metodologia *Design Science Research*, também utilizada neste trabalho, na construção de um artefato para calcular o tamanho ideal de participantes de um bate-papo no contexto

1 <http://www.capes.org.br>

de educação a distância. Destacamos ainda o trabalho realizado por [4], o qual introduz várias métricas para a análise de redes de colaboração científica algumas delas utilizadas neste trabalho.

As demais seções desse trabalho estão assim distribuídas: Na Seção 2, são apresentados os fundamentos teóricos que deram base para a pesquisa; na Seção 3, a proposta de solução é descrita; na Seção 4 é realizada uma análise dos dados das publicações presentes nos currículos lattes² dos pesquisadores vinculados aos programas, limitando-se às publicações realizadas nos últimos dez anos; na Seção 5, a ferramenta de recomendação de parcerias em publicações é detalhada; e finalmente, na Seção 6, são apresentadas as conclusões e os trabalhos futuros.

2 | MÉTODOS E TECNOLOGIAS

Algumas áreas de pesquisa estão diretamente relacionadas a este trabalho. Nesta seção, será descrito brevemente cada uma destas áreas para que seja possível relacioná-las na proposta de solução apresentada na Seção 3.

2.1 Design Science Research

O paradigma *Design Science Research* é adotado nessa pesquisa por possibilitar a construção de um artefato que materialize as conjecturas teóricas e cuja avaliação permita refutar ou comprovar as teorias iniciais estabelecidas na pesquisa. De acordo com [5], *Design Science Research* constitui uma outra “lente”, um conjunto de técnicas com perspectivas analíticas e sintéticas, para realização de pesquisa em sistemas de informação.

Nesse contexto, *Design Science Research* envolve a geração de conhecimento novo através do desenvolvimento de artefatos inovadores e da análise do uso e/ou do desempenho de tais artefatos por meio de reflexão e abstração. Tais artefatos incluem, mas não se limitam a, interfaces homem/máquina, algoritmos, metodologias de projeto de sistemas e linguagens. Esta pesquisa tem o objetivo de construção de um artefato que modele as parcerias entre pesquisadores através de relações de coautoria.

Para a construção do grafo de relacionamento, construído inicialmente do *Design Science*, foram adotados dois tipos de nós, o nó pesquisador e o nó publicação. Também foi definido um tipo de aresta, o qual representa a relação de autoria de uma publicação. O ponto de partida para compreender a relação de parceria entre pesquisadores foi realizado através da coleta de dados dos currículos Lattes dos professores vinculados ao departamento responsável pelo programa de pós-graduação, os quais se relacionam através de autoria em artigos completos publicados em periódicos. As relações de colaboração foram extraídas através dos autores dos artigos que foram cadastrados na plataforma Lattes. Foi utilizado como chave de acesso no banco de

2 Plataforma Lattes do CNPQ. Disponível em: <http://lattes.cnpq.br>

dados os nomes dos autores dos artigos e os nomes utilizados para referência de cada autor na Seção Identificação do currículo Lattes. Assim, com os nomes dos autores dos artigos e os nomes utilizados para referência de cada autor foi possível a geração das arestas, representando as autorias, do grafo permitindo então a visualização das parcerias já realizadas.

2.2 Redes sociais e redes de colaboração científica

Segundo [4], uma rede é um conjunto de objetos conectados entre si. Pode-se representar uma rede por um grafo, onde os nós ou vértices estão relacionados/conectados por arestas. Uma rede social reflete uma estrutura social, a qual pode ser representada por indivíduos ou organizações e suas relações. Em geral, as relações representam um ou mais tipos de interdependência – como ideias ou religiões, ou relacionamentos mais específicos – como troca de conhecimento/informação ou amizade. Através dessa estrutura social é possível estudar a troca de informação entre indivíduos e organizações[4].

Redes de colaboração científica são redes sociais cujos relacionamentos representam algum tipo de interação científica. Segundo [6], uma rede social pode ser representada por um grafo $G=(V,E)$, onde V denota um conjunto de indivíduos em consideração e E um conjunto de relacionamentos existentes entre esses indivíduos. Uma rede social pode suportar um ou mais tipos de relacionamentos, como amizade, paternidade ou colaboração profissional. Intuitivamente, relacionamentos possuem diferentes intensidades que refletem a força das ligações sociais. A intensidade de um relacionamento é usualmente representada por uma função $w(e)$ com $e \in E$, que associa um peso para aresta na rede.

2.3 Bancos de dados NoSQL

Com o desenvolvimento de tecnologias de armazenamento de dados NoSQL, as quais buscam formas alternativas à abordagem relacional para problemas específicos, surgiu o interesse em não apenas modelar uma rede social como um grafo, mas também em armazená-la dessa forma[7].

Neste trabalho, foram estudadas diversas soluções de banco de dados NoSQL e optou-se por empregar o *Banco de dados em Grafo Neo4j*[8]. O Neo4j possui licença GPL, com bastante documentação disponível *on-line*, ao contrário de soluções como OrienteDB[9], AragoDB[10], VelocityGraph[10] ou mesmo bancos de dados relacionais como Oracle através da sua biblioteca Spatial[11] e PostgreSQL usando a PostGIS [12]. Tais soluções utilizam armazenamento baseado em documento ou *Resource Description Framework*, RDF[13] (metamodelo para armazenamento de informações implementado em recursos web pelo consórcio W3C). Além disso, os bancos de dados Oracle e PostgreSQL, usando o paradigma relacional para armazenamento dos dados em tabelas, apenas simulam o modelo em grafo por meio de *stored procedures*. Essa

abordagem aumenta a complexidade das consultas uma vez que é necessário realizar uma tradução da consulta no grafo para uma consulta relacional.

2.4 Sistemas de Recomendação

Segundo [14], o objetivo de um sistema de recomendação é gerar recomendações significativas a um conjunto de usuários para itens ou produtos que lhe possa interessar. Sugestões de livros na Amazon, amigos no Facebook ou filmes no Netflix são exemplos reais do uso de sistemas de recomendação na indústria. O projeto de um sistema de recomendação depende do domínio de características particulares dos dados disponíveis. Tais dados podem armazenar a relevância de um determinado item para um usuário, tal quando se classifica um item com estrelas em uma loja virtual ou quando se aponta o gostar de um comentário em uma rede social. Adicionalmente, o sistema pode ter acesso a atributos específicos de usuários ou itens como informações demográficas e descrição do item respectivamente. Sistemas de recomendação diferem quanto a forma que eles analisam os dados para desenvolver noção de afinidade entre usuários e itens que pode ser usada para identificar o casamento de pares usuário-item.

Existem duas formas mais difundidas de sistemas de recomendação, a filtragem colaborativa, a qual analisa apenas interações históricas e a filtragem baseada em conteúdo, que usa os dados disponíveis nos perfis de usuários e nos atributos dos itens. Há ainda técnicas híbridas que combinam características de ambas as abordagens [14].

3 | PROPOSTA DE SOLUÇÃO

A solução proposta iniciou-se com a modelagem das informações contidas no currículo dos pesquisadores referentes as publicações em periódicos, para tanto gerou-se o modelo apresentado na figura 1.

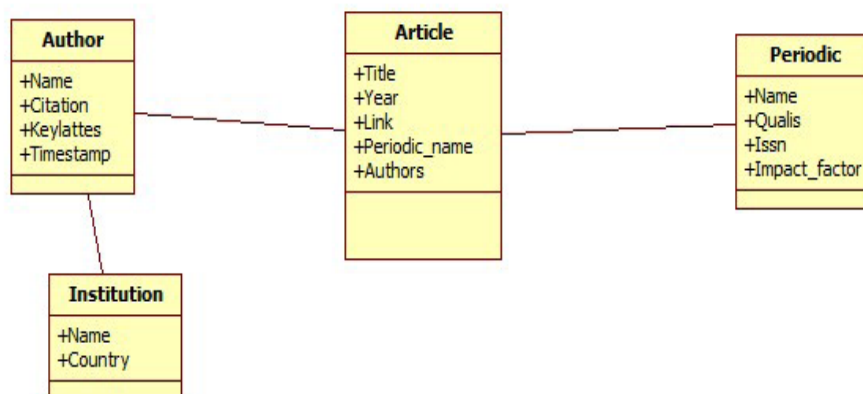


Figura 1: Diagrama de classes representando o modelo de dados

A implementação do artefato iniciou com a construção de um *script* para realizar uma busca dos dados do currículo *Lattes* de cada pesquisador vinculado ao programa de pós-graduação, salvando os dados da carga em arquivos de texto para possibilitar um pré-processamento antes da carga no banco Neo4j. Durante essa fase, algumas medidas corretivas precisaram ser adotadas, como o tratamento de homônimos, abreviações e inconsistências léxicas nos dados de origem. Com o resultado da busca de currículos, foram obtidos os dados de identificação dos pesquisadores, assim como, as publicações deixando essas entidades, autores e publicações, ainda sem definição de relacionamento.

Procedeu-se a carga do grafo com os dados de autores e de publicações obtidos na etapa anterior para então realizar o casamento entre as informações de identificação de cada autor como as informações de autoria de cada publicação. Na sequência, foi construído um artefato de visualização, o qual foi implementado utilizando o padrão de projeto MVC – *Model-View-Controller*[15]. Os elementos foram implementados da seguinte forma:

- *View* – desenvolvido na linguagem Javascript utilizando a biblioteca D3JS[16], adequada para visualização de dados;
- *Controller* – responsável por manipular os objetos do grafo para serem apresentados ao elemento de visualização, utilizando o *framework* web2py[17], escrito na linguagem Python;
- *Model* – implementado através do banco de dados em grafo Neo4j, o qual define a forma de persistência dos objetos de dados.

Uma das vantagens em utilizar um banco de dados orientado a grafo nesse contexto é a facilidade em realizar consultas no banco, como se estas estivessem sendo realizadas no próprio modelo conceitual. Os experimentos realizados utilizaram uma massa de teste com 401 vértices (entre autores e publicações) e 376 arestas, o que não apresentou problema de desempenho no grafo gerado.

A Figura 2 apresenta o processo da proposta, incluindo desde a coleta dos nomes dos professores até a visualização do grafo. Note que o primeiro elemento é o sub-processo de pré-processamento, o qual inclui três atividades:

- busca lattes – realiza a busca do currículo Lattes do autor;
- trata identificação de autoria- - extrai do currículo Lattes os nomes usados pelo autor em citações;
- captura artigos completos publicados em periódicos – extrai as publicações que serão usadas posteriormente para construção da rede de colaboração.

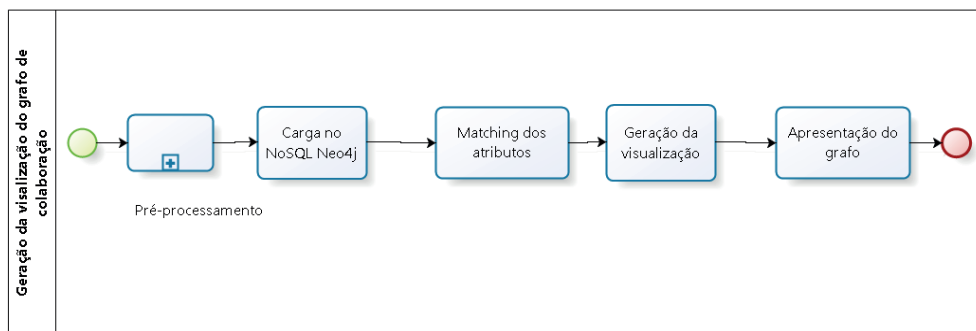


Figura 2: Processo de carga e visualização

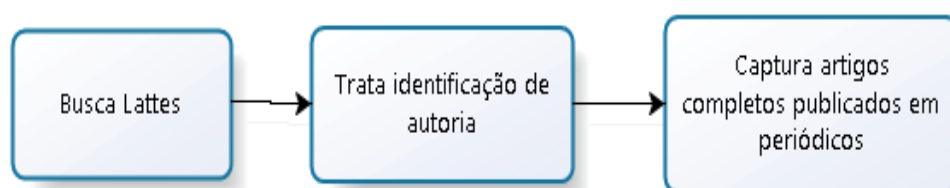


Figura 3: Sub processo de pré processamento

Após o sub-processo de pre-processamento, detalhado na Figura 3, a tarefa de carga no Neo4j é iniciada, quando será realizado o *matching* dos atributos de autoria para criação da rede de colaboração. Na sequencia, ocorrem a geração da visualização, quando são construídos os gráficos. Posteriormente é apresentado o grafo para o usuário.

Uma limitação identificada no decorrer do desenvolvimento do artefato foi a atualização dos dados de publicação nos currículos dos pesquisadores. Bem como, o campo de nomes usados em citações para que haja a correta análise das informações, uma vez que citações constituem a base para o relacionamento entre autores e publicações. A presença de homônimos na base dos currículos Lattes também foi uma dificuldade, a qual foi superada inicialmente através de intervenções manuais através da inclusão dos currículos que a ferramenta não conseguiu carregar. Restou claro que estas intervenções foram viáveis tendo em vista a pequena quantidade de currículos processados, 48 no total. Houve também dificuldade no processamento de caracteres latinos, como letras acentuadas, o que gera a necessidade de uma maior atenção na codificação de caracteres utilizados, para tanto foi necessário realizar uma normalização dos caracteres acentuados reduzindo-os aos caracteres presentes no conjunto de caracteres ASCII³.

Um outro aspecto que deve ser considerado é que embora já existam ferramentas para auxiliar a realização de análise de redes sociais, como o Pajek[18], o NVIVO[19] e o UCINET[20], optou-se pela não utilização dessas ferramentas uma vez que

3 Singla em inglês para Código Padrão Americano para o Intercâmbio de Informação

nenhuma delas possui a capacidade de consultar diretamente o banco de dados em grafo, no qual os dados são persistidos. O uso destas ferramentas traz a necessidade de desenvolver um artefato para organizar os dados de forma que as ferramentas pudessem tratar. Dessa forma, o esforço necessário foi direcionado para a exploração da capacidade de realizar a análise de um modelo de rede de colaboração já persistido em um grafo. Salieta-se que os construtos implementados representam o ponto de partida para o desenvolvimento da pesquisa cujo objetivo é a construção de um modelo de recomendação que utilizará os dados persistidos no banco de dados em grafo já implementado.

4 | ANÁLISE DOS DADOS DE PUBLICAÇÃO

Através dos dados obtidos, tendo por base as métricas definidas em [4], obtivemos o conjunto de valores que caracterizam a rede de colaboração científica e nos permite também realizar inferências para tornar possível a construção de um modelo para realização de recomendações de novas parcerias. Nos últimos dez anos as publicações seguiram o padrão demonstrado na Figura 4, que exhibe graficamente as informações da Tabela 1.

Na Tabela 1, é detalhado a evolução na quantidade de publicações como também a quantidade de coautorias internas, ou seja entre autores do mesmo departamento, e coautorias externas, na qual a relação de coautoria é realizada com um autor não vinculado ao departamento. É possível verificar na coluna *Fator de coautoria* a relação entre os relacionamentos de coautoria internos e externo, o qual foi calculado usando a Equação 1, para cada ano:

$$FC_i = \frac{\sum_{a \in A} (numAutInter_a - numAutExt_a)}{\sum numPublicacoes_i} \quad (1)$$

onde: *numAutInter* representa a quantidade de autorias internas, *numAutExt* a quantidade de autorias externas, *numPublicacoes* a quantidade total de publicações em um determinado ano e *A* é o conjunto de autores de uma determinada publicação. Note que quanto mais negativo for o valor de FC mais coautores externos ao departamento existem nas publicações, demonstrando maior interação com grupos de pesquisa externos.

Ano	Publicações	Coautorias internas	Coautorias externas	Fator de coautoria
2004	15	17	38	-1.4
2005	26	27	81	-2.0769
2006	19	22	45	-1.2105

2007	15	22	24	-0.1333
2008	26	29	62	-1.2692
2009	20	23	48	-1.25
2010	20	26	43	-0.85
2011	19	23	48	-1.3157
2012	21	27	51	-1.1428
2013	36	39	99	-1.6666
2014	25	29	56	-1.08

Tabela 1: Quantidade de publicações

A Figura 4 apresenta o gráfico correspondente aos dados da Tabela 1, sendo o tracejado o total de publicações realizadas, os triângulos as autorias entre os pesquisadores do departamento, os círculos as autorias externas e a linha contínua o fator de coautoria que para melhorar a visualização foi multiplicado por dez. Na mesma figura é possível identificar que no ano de 2013 houve um aumento das publicações com um expressivo incremento na participação de autores externo.

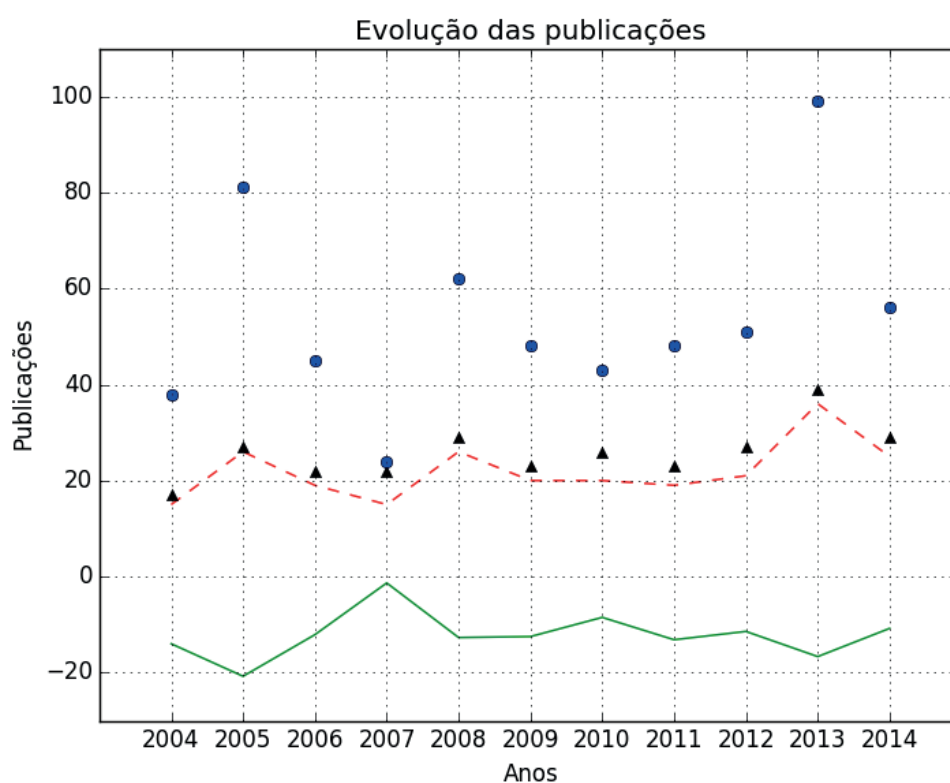


Figura 4: Publicações por ano

5 | RECOMENDAÇÃO DE PARCERIAS NA AUTORIA DE PUBLICAÇÕES

Um aspecto identificado na análise dos dados é que a formação de parceria na coautoria de publicações melhora os indicadores de produção. Nesse sentido, foi desenvolvido um sistema de recomendação para auxiliar a identificar possíveis parcerias na realização de pesquisas.

Conforme [4], recomendação de parcerias, ou predição de link no contexto de um grafo social, pode ser realizada de duas diferentes formas, porém complementares. A primeira examina um sociograma desbalanceado de estruturas sociais buscando transformá-lo em uma estrutura social balanceada. Como exemplo, se duas pessoas possuem muitos amigos em comum então essas duas pessoas devem se conhecer, em uma abordagem usando filtragem colaborativa. A segunda abordagem analisa o conteúdo da comunicação entre os indivíduos, buscando por indivíduos que compartilham o mesmo interesse. Como exemplo, se dois pesquisadores publicam sobre um mesmo assunto então eles deveriam se encontrar, nesse caso o método mais adequado para o sistema de recomendação é a filtragem baseada em conteúdo.

Na implementação realizada nesse trabalho, foi utilizada a metodologia de filtragem colaborativa baseada em vizinhança, nela um subgrupo de usuários são escolhidos baseado na similaridade desses usuários com o usuário ativo, e a combinação ponderada de suas relações usuário/item é usada para inferir predições para o usuário ativo. No contexto desse trabalho, os itens a serem considerados são os periódicos nos quais os pesquisadores realizam suas publicações e a ponderação dos pares pesquisador/periódico será baseada no decurso de tempo entre o ano da publicação e o ano atual em uma abordagem similar à utilizada por [4].

O Fator de influência para cada publicação é dado por:

$$\rho_i = \frac{1}{e^{BY-PY}}$$

onde: BY é o ano no qual a análise foi realizada, no caso 2014 e PY é o ano de publicação do trabalho no periódico

Dessa forma é possível calcular as recomendações conforme o somatório da Equação 2:

$$Rec = \sum_{i=1}^n \rho_i \quad (2)$$

Onde n representa o número de publicações do autor ativo.

Assim, escolhendo aleatoriamente um usuário como ativo, com um código de identificação 3251, foi gerado um ranking com os dez autores cujas probabilidades do usuário ativo formar parceria são as maiores.

#	Id	Rec
1	*3281	0.4193727265647525
2	*3270	0.36787944117144233
3	*3260	0.36787944117144233
4	3271	0.050122530995766454
5	*3295	0.049787068367863944
6	*3287	0.001292744523219513

7	3275	0.00033546262790251185
8	3284	0.00033546262790251185
9	3293	0.00033546262790251185

Tem tempo, o pesquisadores cujas identificações foram marcadas com asterisco já realizaram pelo menos uma publicação em conjunto com o usuário ativo, o que pode ser observado através da Figura 5 que apresenta a rede de colaboração já formada pelo usuário ativo.

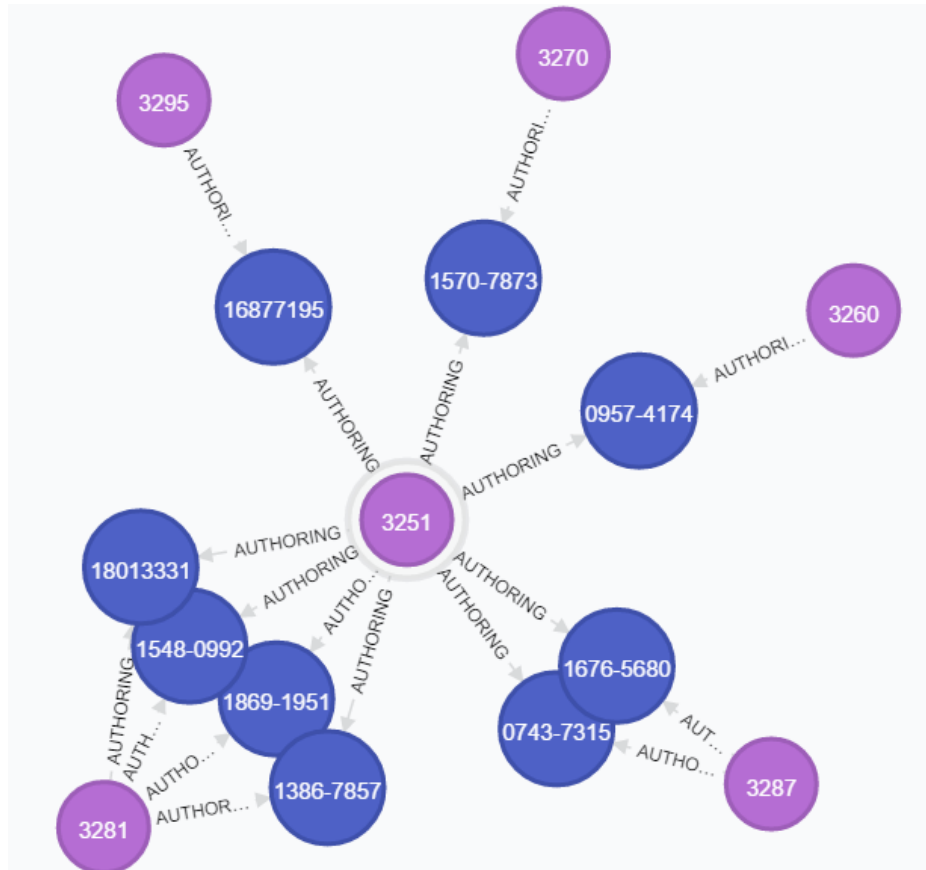


Figura 5: Rede de colaboração centrada um usuário aleatório

6 | CONCLUSÃO

A recomendação de parcerias como ferramenta de melhoria nos índices de publicações acadêmicas se mostra uma solução viável para ser utilizada em diversos cenários nos quais haja necessidade de melhoria dos indicadores de produção através do aumento das relações de coautoria entre os pesquisadores. Dessa forma, o artefato mostrado nesse trabalho pode assumir um papel importante no auxílio a gestão de um programa de pós-graduação, principalmente àqueles submetidos a avaliação da CAPES, por sugerir aos gestores e demais pesquisadores possibilidades de realização de parcerias baseadas nas publicações já realizadas pelos pesquisadores.

A ferramenta aqui apresentada é facilmente adaptada a outros programas de pós-graduação no Brasil uma vez que ela depende apenas dos dados presentes nos

currículos lattes dos pesquisadores vinculados aos programas.

Como trabalhos futuros pretende-se expandir a base de dados utilizada de forma a identificar melhor as parcerias externas e através dessa mudança viabilizar a experimentação de outras abordagens no modelo de recomendação.

REFERÊNCIAS

Hector N. Melo, Ruben A. Perorazio, and Jonice Oliveira. Ambiente analítico web para análise da colaboração científica no cenário médico. *Simpósio Brasileiro de Sistemas de Informação*, 2014.

Edeilson M. Silva, Ricardo A. Costa, Mario Godoy Neto, Robson Y. S. Oliveira, and Silvio R. L. Meira. Promovendo melhorias na comunicação e colaboração em uma plataforma de gestão de conhecimento através de recomendações. *Simpósio Brasileiro de Sistemas de Informação*, 2009.

Edmilson Barcelos Rocha, Mariano Pimentel, and Morganna Carmem Diniz. Desenvolvimento de um modelo da participação em batepapo seguindo a abordagem design science research. *Simpósio Brasileiro de Sistemas de Informação*, 2014.

Victor Ströele, Geraldo Zimbrão, and Jano M. Souza. Group and link analysis of multi-relational scientific social networks. *The Journal of Systems & Software*, 2013.

Vijay K. Vaishnavi and William Kuechler Jr. Design science research in information systems, 2004.

Vinícius P. Freire and Daniel R. Figueiredo. Ranking in collaboration networks using a group based metric. *Journal of Brazilian Computer Society*, 17(4):255–266, November 2011.

Ian Robinson, Jim Webber, and Emil Eifrem. *Graph Databases*. O'Reilly, 2013.

Jim Webber. A programmatic introduction to neo4j. In *Proceedings of the 3rd Annual Conference on Systems, Programming, and Applications: Software for Humanity*, SPLASH '12, pages 217–218, New York, NY, USA, 2012. ACM.

Claudio Tesoriero. *Getting Started with OrientDB*. Packt Publishing Ltd, 2013.

Matías Javier Antiñanco. *Bases de Datos NoSQL: escalabilidad y alta disponibilidad a través de patrones de diseño*. PhD thesis, Facultad de Informática, 2014.

Ravi Kothuri, Albert Godfrind, and Euro Beinat. *Pro oracle spatial for oracle database 11g*. Dreamtech Press, 2008.

Regina Obe and Leo Hsu. *PostGIS in action*. Manning Publications Co., 2011.

Dan Brickley and Ramanathan V Guha. Resource description framework (rdf) schema specification 1.0: W3c candidate recommendation 27 march 2000. 2000.

rem Melville and Vikas Sindhwani. Recommender systems. In *Encyclopedia of machine learning*, pages 829–838. Springer, 2010.

Trygve Reenskaug. Model-view-controller.

Mike Bostock. Data-driven documents (d3. js), a visualization framework for internet browsers running javascript, 2012.

Massimo DiPierro. *Web2Py Enterprise Web Framework, 2Nd Ed.* Wiley Publishing, 2nd edition, 2009.

V. Batagelj and A. Mrvar. Pajek-program for large network analysis. *Connections*, 21:47–57, 1998.

Graham R Gibbs. *Qualitative data analysis: Explorations with NVivo.* Open University, 2002.

Analytic Technologies. Ucinet 6 social network analysis software. 2013.

Agência Brasileira do ISBN
ISBN 978-85-7247-048-3



9 788572 470483