

**UNIVERSIDADE FEDERAL DE SANTA CATARINA  
DEPARTAMENTO DE INFORMÁTICA E ESTATÍSTICA**

Gabriela Bussolo Colonetti

**AGRUPANDO PESQUISADORES POR COAUTORIA DE  
PUBLICAÇÕES SEGUNDO CURRÍCULO LATTES**

Florianópolis

2016



Gabriela Bussolo Colonetti

**AGRUPANDO PESQUISADORES POR COAUTORIA DE  
PUBLICAÇÕES SEGUNDO CURRÍCULO LATTES**

Trabalho de conclusão de curso apresentado como parte dos requisitos para a obtenção do grau de “Bacharel em Sistemas de Informação”.

Florianópolis, 12 de dezembro 2016.

**Banca Examinadora:**

---

Msc. Rodrigo Gonçalves  
Orientador

---

Prof. Dra. Carina Dorneles  
Coorientador

---

Prof. Dr. Ronaldo dos Santos Mello



---

Prof. Msc. Angelo Frozza



## RESUMO

Técnicas de agrupamento de dados são responsáveis por criar grupos, ou *clusters*, de objetos que são semelhantes para que se possa extrair informações desses grupos com facilidade. Para comparar pesquisadores, diferentes abordagens são possíveis. Uma delas é com base em competências extraídas de suas produções via palavras-chave, títulos etc. Outra possibilidade é considerar a coautoria, ou seja, se um profissional X publicou junto com outro profissional Z e Y também publicou com Z, há uma possível similaridade de interesse entre X e Y, sendo interessante sugerir uma colaboração entre eles. Este trabalho aborda a clusterização de pesquisadores por coautoria sobre o Currículo Lattes, comparando e agrupando (*clusterizando*) profissionais e utiliza-se trabalhos existentes como apoio e baseline.

**Palavras-chave:** Conhecimento. Agrupamento de dados. Lattes. Clusterização. Coautoria.





## LISTA DE FIGURAS

Figura 1	Fragmento de um Currículo Lattes em formato XML...	18
Figura 2	(a) Esquema de um dado XML e (b) um exemplo de dado XML. ....	20
Figura 3	(a) Exemplo de grafo direcionado, (b) exemplo de grafo não direcionado. ....	20
Figura 4	Exemplo de criação do grafo utilizando JGraphT. ....	21
Figura 5	Visão geral do agrupador. ....	25
Figura 6	Utilização do xPath no projeto ....	26
Figura 7	Esboço do diagrama de classes do projeto. ....	27
Figura 8	Exemplo de resultado da criação de Grafo. ....	28
Figura 9	Desenho de grafo gerado após a criação do Grafo. ....	29
Figura 10	Exemplos de nomes retirados dos currículos. ....	32
Figura 11	Representação de títulos diferentes do mesmo artigo... ..	33
Figura 12	Representação dos autores do artigo. ....	34
Figura 13	Similaridade entre alguns autores INE/UFSC. ....	34
Figura 14	Cluster gerado pelo método Jaccard. ....	37
Figura 15	Cluster gerado pelo método Overlap. ....	38
Figura 16	Cluster gerado pelo método Cosine. ....	39
Figura 17	Publicações com coautorias do Dr. Carlos Westphall... ..	40
Figura 18	Publicações com coautorias da Dra. Carla Westphall... ..	40
Figura 19	Execução algoritmo SCD. ....	42



## LISTA DE ABREVIATURAS E SIGLAS

CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico.....	17
W3C	World Wide Web Consortium.....	18
xPath	XML Path Language.....	19
SCD	Similaridade por Coautoria Direta.....	29



## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	13
1.1 OBJETIVO GERAL .....	13
1.2 OBJETIVOS ESPECÍFICOS .....	13
1.3 METODOLOGIA .....	14
1.4 ORGANIZAÇÃO DO TEXTO .....	14
<b>2 FUNDAMENTAÇÃO TEÓRICA</b> .....	17
2.1 CURRÍCULO LATTES .....	17
2.2 XML .....	18
2.3 XPATH .....	19
2.4 GRAFOS .....	19
<b>2.4.1 JGraphT</b> .....	20
2.5 URSA .....	21
2.6 TÉCNICAS DE AGRUPAMENTO .....	21
2.7 <i>COLLABSEER</i> .....	22
<b>3 FERRAMENTA PARA AGRUPAMENTO POR CO- AUTORIA UTILIZANDO O LATTES</b> .....	25
3.1 VISÃO GERAL DO PROCESSO .....	25
3.2 ARQUITETURA .....	26
3.3 EXTRAÇÃO DOS DADOS .....	26
3.4 CRIAÇÃO DO GRAFO NÃO DIRECIONADO .....	28
3.5 CLUSTERIZAÇÃO .....	29
<b>3.5.1 Análise da Similaridade entre os Vértices</b> .....	30
<b>3.5.2 Criação dos Grupos</b> .....	30
<b>4 EXPERIMENTOS</b> .....	35
4.1 RESULTADOS <i>BEST-STAR</i> .....	35
<b>4.1.1 <i>Best-star</i> Utilizando Todos Autores</b> .....	40
4.2 RESULTADOS SCD .....	41
<b>5 CONCLUSÃO E TRABALHOS FUTUROS</b> .....	43
<b>APÊNDICE A – Resultados da implementação</b> .....	47
<b>ANEXO A – Artigo</b> .....	93
<b>REFERÊNCIAS</b> .....	105



# 1 INTRODUÇÃO

O crescimento da comunidade acadêmica, tanto em membros quanto em áreas de pesquisa e publicações, torna complexo analisar e identificar profissionais com interesses semelhantes para sugerir colaborações. Desta forma, automatizar processos que façam essa análise é vantajoso, como por exemplo, extrair e analisar informações dos currículos dos pesquisadores.

No Brasil, o CNPQ implantou a plataforma Lattes para criação e gestão de currículos virtuais com intuito de integrar currículos profissionais de instituições e grupos de pesquisa (CNPQ, 2015).

Este trabalho utiliza a coautoria para clusterização de pesquisadores. Verificando as publicações dos profissionais e analisando os coautores. Busca-se verificar se pelas coautorias pode-se construir *clusters* de profissionais afins em sua pesquisa. A capacidade de definir grupos de profissionais com características ou interesses acadêmicos em comum é uma forma de incentivar o compartilhamento de experiências e conhecimentos (LI et al., 2014).

Pesquisas colaborativas são cada vez mais populares e importantes no ciclo acadêmico, porém não existem plataformas para recomendação desses potenciais colaboradores (CHEN et al., 2011). Neste trabalho será explorado a possibilidade de descobrir colaboradores com base na estrutura da rede de coautores e interesses de pesquisa de um usuário.

## 1.1 OBJETIVO GERAL

O objetivo principal do trabalho é oferecer uma ferramenta que, dada uma série de currículos da plataforma Lattes, em formato XML, defina-se *clusters* de pesquisadores com interesse semelhantes, com base nas coautorias. O intuito desses *clusters* é identificar pesquisadores com expertise semelhante. São feitas análises sobre os conteúdos dos currículos e para a criação dos *clusters* são aplicados diferentes critérios com base em artigos relacionados.

## 1.2 OBJETIVOS ESPECÍFICOS

Os seguintes objetivos específicos foram estabelecidos:

- Analisar os currículos Lattes, verificando os campos necessários para extração e implementação do trabalho;
- Implementar algoritmos para realizar a clusterização de pesquisadores com base nas coautorias;

### 1.3 METODOLOGIA

Para a realização deste trabalho, foram estudados os documentos do currículo Lattes, em formato XML, analisando quais elementos são relevantes para o objetivo do trabalho. Os currículos analisados e utilizados durante todo o trabalho foram coletados da plataforma Lattes.

Inicialmente, apenas alguns currículos foram utilizados para os experimentos, a fim de ter um maior domínio e controle sobre o problema. À medida que os testes avançaram, uma quantidade maior de currículos foi utilizada.

Após o conhecimento sobre os currículos estar mais sólido e os elementos necessários para a realização do trabalho definidos, foi utilizada uma linguagem de consulta para documentos em XML, com o intuito de recuperar os dados que são utilizados para a criação dos grupos. Para a criação dos grupos, foi estabelecida uma estrutura de grafo como organização da cadeia de relação entre os autores.

Posteriormente, foram realizados estudos para buscar na literatura algoritmos de clusterização. Por fim, desenvolveu-se métodos baseados em Chen et al. (2011) e algoritmos próprios para geração dos grupos.

### 1.4 ORGANIZAÇÃO DO TEXTO

O trabalho está organizado em cinco capítulos. O primeiro apresenta os objetivos do projeto, citando os tópicos explorados e a metodologia utilizada.

O segundo capítulo descreve os conceitos necessários ao entendimento do projeto, como por exemplo, a plataforma Lattes e as técnicas e ferramentas utilizadas na implementação.

No capítulo três, detalha-se o desenvolvimento do projeto, incluindo as principais características da implementação. Apresenta-se os dados extraídos dos currículos Lattes, detalhes dos algoritmos utilizados para a criação dos grupos, além dos testes realizados para analisar



a eficácia da solução proposta.

O quarto capítulo trata dos resultados obtidos e no quinto encontram-se as conclusões e trabalhos futuros.



## 2 FUNDAMENTAÇÃO TEÓRICA

Nesse capítulo são descritas as técnicas e tecnologias utilizadas no projeto. É apresentada a estrutura do currículo Lattes, a linguagem de consulta para selecionar nodos em um documento XML (XPath) e uma explicação sobre a estrutura de dados em grafo.

Também é incluída uma descrição da biblioteca usada para a criação do grafo, o JGraphT, os métodos usados na implementação deste trabalho e o *framework* URSA utilizado para geração dos *clusters*.

### 2.1 CURRÍCULO LATTES

Neste trabalho, para realizar a criação dos grupos com base em coautorias, foram utilizados currículos Lattes dos profissionais e acadêmicos do Brasil, retirados da plataforma Lattes, gerida pelo CNPq. O Lattes é uma plataforma que contém currículos de pesquisadores de diversas instituições do Brasil.

Nos anos 80 surgiu a preocupação em definir um padrão para formulários de registro dos currículos dos pesquisadores brasileiros, o que além de facilitar a avaliação curricular dos pesquisadores, facilita à análise e seleção de especialistas. No final dos anos 90, o CNPq investiu no desenvolvimento de uma versão única de currículo que fosse capaz de integrar os pesquisadores. Assim, desde 1999 o Brasil possui um formulário único de currículo para o meio acadêmico (CNPQ, 2016).

Ao longo dos anos, a plataforma evoluiu e hoje possui diversas funcionalidades, como estatísticas e extração de dados. Extrair dados de um currículo é possível desde 2001 quando o CNPq e o instituto Stela<sup>1</sup>, organização privada sem fins econômicos dedicada à pesquisa, definiram os padrões da Plataforma Lattes, incluindo a funcionalidade de extração do currículo em formato XML. Por ser uma base de dados pública, qualquer pessoa pode extrair dados (STELA, 2016). Entretanto, a extração dos currículos é feita manualmente - não há nenhuma forma pública como, por exemplo, *web-services*, para a extração automatizada.

---

<sup>1</sup><http://www.stela.org.br/>

## 2.2 XML

XML é uma meta-linguagem, ou seja, ela oferece recursos para a definição de gramáticas que caracterizam linguagens para classes de documentos específicos, com conjunto de elementos, atributos e regras de composição bem determinados (W3C, 2008). Foi desenvolvida pela W3C, comunidade constituída por acadêmicos, pesquisadores e empresários, que tem como finalidade recomendar padrões e protocolos para a Web. De acordo com a W3C, entre os objetivos estabelecidos na especificação da linguagem XML estão as seguintes características: ser diretamente utilizável na Internet, ser legível por humanos, possibilitar um meio independente para publicação eletrônica e facilitar às pessoas o processamento de dados (ALMEIDA, 2002).

XML foi projetada para descrever a informação focando no que ela representa (W3C, 2008). A meta-linguagem é conhecida por ser uma representação textual de dados. O componente básico na XML é o elemento, ou tag, que é um texto intercalado pelos sinais < e > como por exemplo, <ARTIGO-PUBLICADO> para iniciar um elemento e </ARTIGO-PUBLICADO> para delimitar o fim dele. Dentro do elemento tem-se texto bruto, outro elemento ou uma mistura dos dois (ALMEIDA, 2002).

A descrição das informações é feita de maneira estruturada, similar a de uma árvore. Um documento XML possui um elemento raiz que contém todos os outros elementos, denominados filhos.

Figura 1 – Fragmento de um Currículo Lattes em formato XML.

```
<ARTIGOS-PUBLICADOS>
<ARTIGO-PUBLICADO SEQUENCIA-PRODUCAO="1" ORDEM-IMPORTANCIA="">
  <DADOS-BASICOS-DO-ARTIGO NATUREZA="COMPLETO"
  TITULO-DO-ARTIGO="Efficient scheduling of MPI applications on networks of workstations"
  ANO-DO-ARTIGO="1998" PAIS-DE-PUBLICACAO="" IDIOMA="Inglês" MEIO-DE-DIVULGACAO="MEIO_DIGITAL"
  NOME-PAGE-DO-TRABALHO="" FLAG-RELEVANCIA="SIM" DOI="10.1016/S0167-739X(97)00012-5"
  TITULO-DO-ARTIGO-INGLES="Efficient scheduling of MPI applications on networks of workstations"
  FLAG-DIVULGACAO-CIENTIFICA="NAO"/>
  <DETALHAMENTO-DO-ARTIGO TITULO-DO-PERIODICO-OU-REVISTA="Future Generation Computer Systems"
  ISSN="0167739X" VOLUME="13" FASCICULO="6" SERIE="" PAGINA-INICIAL="489" PAGINA-FINAL="499"
  LOCAL-DE-PUBLICACAO="North-Holland"/>
  <AUTORES NOME-COMPLETO-DO-AUTOR="Dantas, M" NOME-PARA-CITACAO="Dantas, M" ORDEM-DE-AUTORIA="1"
  NRO-ID-CNPQ="2900995280822495"/>
  <AUTORES NOME-COMPLETO-DO-AUTOR="Mario Antonio Ribeiro Dantas" NOME-PARA-CITACAO="DANTAS, M. A. R."
  ORDEM-DE-AUTORIA="2" NRO-ID-CNPQ=""/>
  <PALAVRAS-CHAVE PALAVRA-CHAVE-1="Message Passing" PALAVRA-CHAVE-2="Parallel Environment"
  PALAVRA-CHAVE-3="MPI" PALAVRA-CHAVE-4="Distributed Processing" PALAVRA-CHAVE-5="Load balancing"
  PALAVRA-CHAVE-6="Network of Workstations"/>
  <AREAS-DO-CONHECIMENTO>
  <AREA-DO-CONHECIMENTO-1 NOME-GRANDE-AREA-DO-CONHECIMENTO="CIENCIAS EXATAS E DA TERRA"
  NOME-DA-AREA-DO-CONHECIMENTO="" NOME-DA-SUB-AREA-DO-CONHECIMENTO="Sistemas de Computação"
  NOME-DA-ESPECIALIDADE="Arquitetura de Sistemas de Computação"/>
  </AREAS-DO-CONHECIMENTO>
</ARTIGO-PUBLICADO>
<ARTIGO-PUBLICADO SEQUENCIA-PRODUCAO="2" ORDEM-IMPORTANCIA="">
```

Fonte: Currículo Lattes (Cnpq).

Na Figura 1 temos um fragmento de documento XML - uma parte de um currículo Lattes. A tag ARTIGOS-PUBLICADOS contém um conjunto de tags ARTIGO-PUBLICADO onde estão inseridas algumas informações referentes às publicações dos autores que foram exploradas neste trabalho.

### 2.3 XPATH

A XML *Path Language* (XPath) , é uma linguagem de consulta que provê a projeção e seleção sobre nodos XML. A sintaxe adotada pelo XPath é bastante intuitiva, já que ela se baseia-se em caminhos. XPath pode retornar diversos tipos de dados, suas expressões podem retornar conjuntos de nodos ou atributos de tipos *string*, *boolean*, *integer*, dentre outros (VAZQUEZ, 2005).

Através da XPath é possível filtrar determinado elemento de acordo com a sua posição, realizar verificações na expressão, selecionar atributos, fazer comparações, selecionar mais de um caminho em uma mesma expressão, além da possibilidade de utilização de funções pré-estabelecidas, como *contains()*, *substring()*, *sum()*, dentre outras.

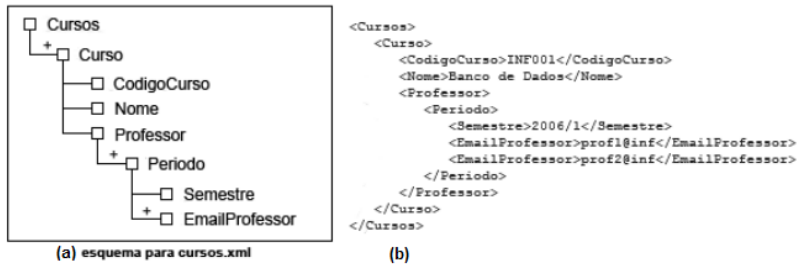
Há dois tipos de caminho para localização utilizando XPath: relativo e absoluto. Um caminho absoluto consiste em uma sequência de passos de uma ou mais localizações, iniciado e separado por barra simples, onde a busca inicia pelo elemento raiz passando por níveis da árvore em um dado XML. Já na localização relativa é possível especificar, em qualquer ponto do caminho da busca, elementos relativos ao elemento contexto.

Na Figura 2 temos um exemplo de estrutura XML. Para realizar a consulta do "Curso" utilizando o caminho absoluto usa-se a expressão `xPath"/Cursos/Curso"`. Já no relativo, a expressão pode ser escrita da seguinte maneira `"Cursos/Curso"` ou `"//Curso"` (SILVEIRA, 2006).

### 2.4 GRAFOS

Por definição, um grafo, ou dígrafo, é um conjunto de vértices e arestas, no qual os vértices são considerados objetos não vazios, que podem possuir atributos ou não, e as arestas são as relações entre esses objetos. As arestas de um grafo podem ter uma direção. Neste caso, um grafo é chamado de direcionado, caso contrário, o grafo é denominado não-direcionado. Para determinar a direção, representam-se as arestas

Figura 2 – (a) Esquema de um dado XML e (b) um exemplo de dado XML.

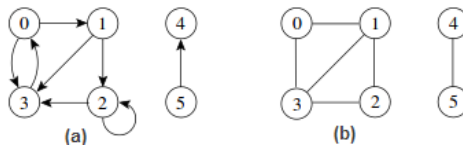


Fonte: Silveira (2006).

com setas na ponta destino. No caso de um grafo não-direcionado como na Figura 3b, as arestas  $(0, 1)$  e  $(1, 0)$  são idênticas, o que não acontece em um grafo direcionado (LINDEN, 2009).

Outra propriedade de grafos é a possibilidade de associar um valor para cada aresta, que pode ser considerado seu peso.

Figura 3 – (a) Exemplo de grafo direcionado, (b) exemplo de grafo não direcionado.



Fonte: Almeida (2004).

### 2.4.1 JGraphT

Para a implementação do trabalho, foi utilizado o JGraphT<sup>2</sup>, uma biblioteca Java gratuita que fornece objetos e algoritmos matemáticos da teoria dos grafos. É uma biblioteca extensa, porém de uso simples. Os vértices do grafo podem ser qualquer tipo de objeto (JGRAPHT, 2016).

A Figura 4 demonstra um uso do JGraphT no código deste pro-

<sup>2</sup><http://jgrapht.org/>

Figura 4 – Exemplo de criação do grafo utilizando JGraphT.

```
SimpleWeightedGraph<String, DefaultWeightedEdge> grafo =
new SimpleWeightedGraph<String, DefaultWeightedEdge>(DefaultWeightedEdge.class);
grafo.addVertex(p1);
grafo.addVertex(p2);
DefaultWeightedEdge e = new DefaultWeightedEdge();
grafo.setEdgeWeight(e, tupla.getPeso());
grafo.addEdge(p1, p2, e);
```

Fonte: Elaborada pela autora.

jeto. Para a criação dos *clusters* foram utilizados outros métodos implementados pela biblioteca, facilitando, por exemplo, extrair todas as arestas de um vértice ou retonar todos os vértices de um grafo.

## 2.5 URSA

O URSA<sup>3</sup> é um *framework* que implementa técnicas e algoritmos de agrupamento de dados. A ferramenta é uma extensão do trabalho desenvolvido por Ribacki (2013), de modo que o *framework* dê suporte a arquivos de áudio e vídeo e, também, implemente novas técnicas de agrupamento (GROSS, 2014).

O *framework* implementa diversos algoritmos de clusterização, como *Starts*, *Best-Star*, *Full-Stars*, *K-Médias*, *K-Medoids*, dentre outros.

## 2.6 TÉCNICAS DE AGRUPAMENTO

Técnicas ou algoritmos de agrupamento permitem criar grupos de dados para determinar padrões a partir dos dados (ABERNETHY, 2010). As técnicas de agrupamento implementadas pelo URSA podem ser separadas em três grandes grupos: agrupamento por partição total, agrupamento hierárquico e agrupamento baseado em densidade (GROSS, 2014), que são exploradas a seguir:

- Agrupamento por partição total: Técnica que gera grupos distintos sem relação entre eles - os objetos são agrupados de forma que todos os elementos de um mesmo aglomerado tenham um grau de similaridade mínima (*threshold*) entre eles (GROSS, 2014). A técnica de agrupamento por partição total foi adotada no desen-

---

<sup>3</sup><https://github.com/jlggross/URSA>

volvimento deste trabalho;

- **Agrupamento hierárquico:** Um algoritmo hierárquico que inicia com um agrupamento disjunto, ou seja, todos os elementos pertencem a um único grupo, colocando cada objeto em um grupo individual. Definidos os grupos iniciais, eles são analisados aos pares e, ao encontrar o maior grau de similaridade com outro grupo, um novo grupo é criado contendo os dois grupos. O processo é feito recursivamente até que se forme apenas um único grupo raiz (GROSS, 2014);
- **Agrupamento baseado em Densidade:** A ideia desse tipo de agrupamento é agrupar objetos vizinhos de um grupo de objetos em um *cluster* baseado em condições de densidade (GROSS, 2014). Esses algoritmos tipicamente consideram como aglomerados regiões densas de objetos no espaço de objetos.

## 2.7 COLLABSEER

*CollabSeer: A Search Engine for Collaboration Discovery* (CHEN et al., 2011) é um trabalho que consiste em uma plataforma de recomendação de colaboradores em pesquisas científicas, realizando clusterizações com base em coautoria, no qual também é possível encontrar um profissional de uma determinada expertise através de buscas por palavras-chaves referentes a área de pesquisa desejada.

Neste trabalho, a clusterização é feita com base no *dataset* do CiteSeerX, repositório para trabalhos científicos e acadêmicos com foco em ciência da computação e da informação (CITSEERX, 2015). O artigo realizou cálculos de similaridade entre os autores com base em coautorias utilizando três frações de similaridade: *Jaccard*, *Cosine* e *Overlap*, elas foram implementadas neste trabalho e são detalhadas à seguir:

1. **Coefficiente de *Jaccard*:** A equação define que a relação entre dois autores, ou seja, dois vértices de um grafo, é medida de acordo com os conjuntos vizinhos, onde a quantidade da intersecção dos vizinhos entre dois vértices é dividida pela quantidade da união dos vizinhos entre esses dois autores. A Equação 2.1 ilustra a fórmula da equação, em que  $m_i$  e  $m_j$  são os conjuntos de vizinhos dos vértices  $v_i$  e  $v_j$ ,  $\tau$  representa a quantidade,  $\cap$  intersecção e  $\cup$  a união.



$$S_{Jaccard}(v_i, v_j) = \frac{\tau(m_i \cap m_j)}{\tau(m_i \cup m_j)} \quad (2.1)$$

2. Coeficiente de *Cosine*: O cálculo do coeficiente de *Cosine* é similar ao de *Jaccard*, com a diferença no denominador, conforme mostra a Equação 2.2, a quantidade ( $\tau$ ) da intersecção ( $\cap$ ) de vizinhos  $m_i$  e  $m_j$  dos vértices  $v_i$  e  $v_j$  é dividida pela multiplicação da quantidade de vizinhos que cada vértice possui e, então, é aplicado a raiz quadrada sobre o valor.

$$S_{Cosine}(v_i, v_j) = \frac{\tau(m_i \cap m_j)}{\sqrt{\tau(m_i)\tau(m_j)}} \quad (2.2)$$

3. Coeficiente de *Overlap*: O coeficiente utiliza também os vizinhos para calcular a similaridade entre dois vértices, porém a quantidade ( $\tau$ ) da intersecção de vizinhos entre dois vértices é dividido pelo valor mínimo de vizinhos de um dos dois vértices analisados, ou seja, se o vértice  $v_i$  possuir 10 vizinhos e o vértice  $v_j$  possuir 5 vizinhos, o denominador da equação é o valor 5. A Equação 3 representa a equação.

$$S_{Overlap}(v_i, v_j) = \frac{\tau(m_i \cap m_j)}{\min(\tau(m_i)\tau(m_j))} \quad (2.3)$$



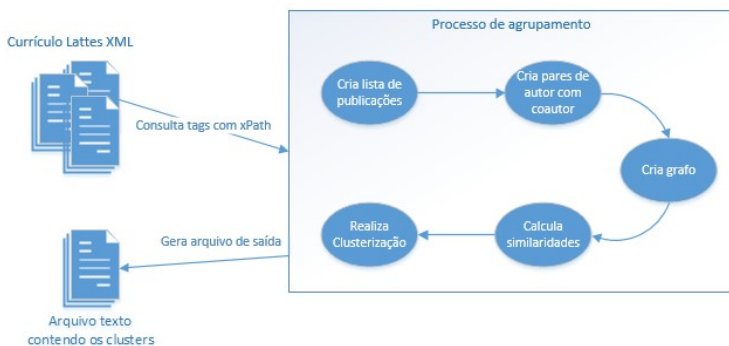
### 3 FERRAMENTA PARA AGRUPAMENTO POR COAUTORIA UTILIZANDO O LATTES

Este capítulo apresenta uma visão geral da implementação e funcionamento da ferramenta de clusterização de autores por coautoria desenvolvida neste projeto.

#### 3.1 VISÃO GERAL DO PROCESSO

A Figura 5 representa de forma resumida o funcionamento do agrupador. Há a entrada de uma série de currículos Lattes em formato XML, passando por um processo de agrupamento no qual são identificadas coautorias, até a criação de um grafo não direcionado, onde as arestas possuem pesos da quantidade de publicações distintas que os autores realizaram juntos. A saída é um arquivo com todos os *clusters* criados no processo. Para a clusterização, inicialmente, foi desenvolvido um algoritmo e, posteriormente, foram aplicados algoritmos baseados em trabalhos existentes.

Figura 5 – Visão geral do agrupador.



Fonte: Elaborada pela autora.

Foi utilizada a XPath para extrair os dados que foram considerados necessários para análise dos pesquisadores e seus artigos, assim como suas respectivas áreas de conhecimento. A extração de todos os artigos publicados de cada pesquisador foi feita utilizando os caminhos absolutos do XML. Foram realizados também consultas de atributos e consultas com mais de um caminho na mesma expressão, conforme

modelo na Figura 6.

Figura 6 – Utilização do xPath no projeto

```
XPathExpression exprAchaArea = xpath.compile("//*[starts-with(name(),
'AREA-DO-CONHECIMENTO')]");
XPathExpression exprAchaArtigos = xpath.compile("/CURRICULO-VITAE
/PRODUCAO-BIBLIOGRAFICA/ARTIGOS-PUBLICADOS/
ARTIGO-PUBLICADO|CURRICULO-VITAE/PRODUCAO-BIBLIOGRAFICA/TRABALHOS-EM-EVENTOS
/TRABALHO-EM-EVENTOS");
NodeList nlArtigos = (NodeList) exprAchaArtigos.evaluate(document, XPathConstants.NODESET);
XPathExpression exprAchaTitulo = xpath.compile("DADOS-BASICOS-DO-ARTIGO/@TITULO-DO-ARTIGO |
DADOS-BASICOS-DO-TRABALHO/@TITULO-DO-TRABALHO");
XPathExpression exprAchaAno = xpath.compile("DADOS-BASICOS-DO-ARTIGO/@ANO-DO-ARTIGO |
DADOS-BASICOS-DO-TRABALHO/@ANO-DO-TRABALHO");
XPathExpression exprAchaAutores = xpath.compile("AUTORES");
```

Fonte: Elaborada pela autora.

### 3.2 ARQUITETURA

A Figura 7 representa um esboço do diagrama de classes para organização do código, no qual foram criados três pacotes: *main*, *model* e *source*. No *model* estão concentradas as classes de modelo do projeto, como *Pessoa*, *Grafo*, *Publicacao*, *Tupla* e *Grupo*. No *source* tem-se o *ManipuladorXML* no qual é realizada a manipulação dos dados. Seus métodos são chamados pela classe *Main* localizada no pacote *main*. No *ManipuladorXML* também se concentra a lógica do xPath, utilizada para a busca dos dados necessários. Após a extração, foram criados os objetos para posteriormente o *Grafo* ser submetido ao algoritmo de criação de grupos.

A classe *Pessoa* é relacionada com a classe *Publicacao* através dos atributos *autor principal* e *lista de coautores*. Antes da criação do grafo uma lista de objetos *Tupla* é criada, no qual cada *Tupla* possui dois objetos *Pessoa*, referentes aos autores, a lista de publicações em comum e um valor inteiro que representa a quantidade de vezes que esses dois autores publicaram artigos distintos.

### 3.3 EXTRAÇÃO DOS DADOS

O objetivo desta primeira etapa é a extração de nodos relacionados aos artigos publicados dos autores, além de informações referentes às áreas de atuação de cada currículo processado. Foram considerados



cada autor e coautor, porém, após um estudo mais aprofundado, foi verificado a falta deste número em alguns profissionais. Por este motivo, o nome do pesquisador também foi extraído do XML e utilizado para sua identificação. No início do desenvolvimento não foi feito nenhum tratamento em relação aos nomes divergentes do mesmo profissional. No entanto, posteriormente foi verificado que as diferenças entre nomes referenciando a mesma pessoa poderia ocasionar alteração no resultado. Por isso, medidas foram implementadas e são explicadas na seção 3.5.

### 3.4 CRIAÇÃO DO GRAFO NÃO DIRECIONADO

Após a criação da lista de objetos `Tupla` com os atributos necessários, foi realizada a criação do grafo não direcionado, utilizando métodos da biblioteca `JGraphT`. O grafo criado dispõe dos nomes dos autores como vértices e as arestas são valoradas pelo número de coautorias. O cálculo feito na classe `Tupla` é realizado da seguinte forma: cada vez que o programa encontra uma publicação de ano e título diferente de uma publicação já inserida na lista de publicações da tupla analisada, esta publicação é adicionada na lista e o peso é incrementado.

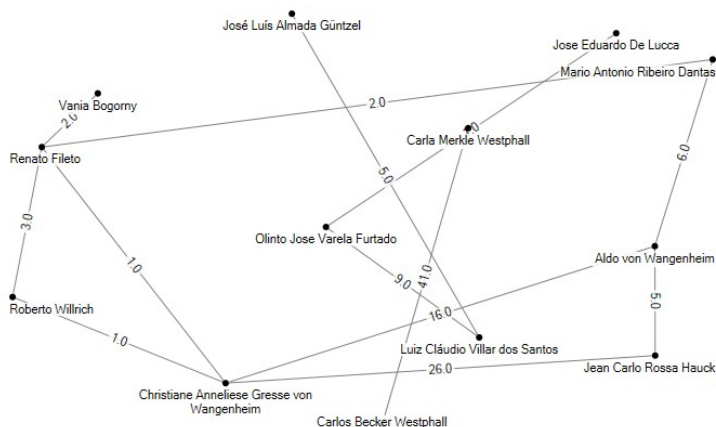
Figura 8 – Exemplo de resultado da criação de Grafo.

```
(José Luis Almada Güntzel : Luiz Cláudio Villar dos Santos) - 5.0
(Carla Merkle Westphall : Carlos Becker Westphall) - 41.0
(Roberto Willrich : Renato Fileto) - 3.0
(Roberto Willrich : Christiane Anneliese Gresse von Wangenheim) - 1.0
(Renato Fileto : Christiane Anneliese Gresse von Wangenheim) - 1.0
(Aldo von Wangenheim : Christiane Anneliese Gresse von Wangenheim) - 16.0
(Luiz Cláudio Villar dos Santos : Olinto Jose Varela Furtado) - 9.0
(Aldo von Wangenheim : Jean Carlo Rossa Hauck) - 5.0
(Christiane Anneliese Gresse von Wangenheim : Jean Carlo Rossa Hauck) - 26.0
(Aldo von Wangenheim : Mario Antonio Ribeiro Dantas) - 6.0
(Mario Antonio Ribeiro Dantas : Renato Fileto) - 2.0
(Renato Fileto : Vania Bogorny) - 2.0
(Jose Eduardo De Lucca : Olinto Jose Varela Furtado) - 1.0
```

Fonte: Elaborada pela autora.

Após os testes terem sido realizados, a saída dos dados quando organizados no grafo pode ser analisada na Figura 8, a Figura 9 representa um desenho do grafo gerado.

Figura 9 – Desenho de grafo gerado após a criação do Grafo.



Fonte: Elaborada pela autora.

### 3.5 CLUSTERIZAÇÃO

Para a clusterização dos pesquisadores com base em suas coautorias em publicações, foi implementado, em um primeiro momento, um algoritmo simples denominado "Similaridade por Coautoria Direta - SCD", da seguinte forma:

1. Foi estabelecido um *threshold*, ou seja, um valor mínimo para o relacionamento entre os autores, de valor três;
2. O algoritmo faz uma busca em todo o grafo para encontrar o maior peso entre dois pesquisadores, ou seja, o maior peso de uma aresta do grafo. Após descobrir quais são esses vértices (pesquisadores), a aresta é adicionada em um objeto *Grupo*, classe que possui uma lista de atributos do tipo *DefaultWeightedEdge*, que é uma classe referente ao vértice na biblioteca *JGraphT*;
3. Verifica-se então as arestas dos dois vértices (da aresta maior), as arestas que dispõem de um valor maior que o *threshold* será adicionado também ao *Grupo*. O processo é repetido até não existir mais ligações entre membros desse grupo inicial com valores superiores ao *threshold*;
4. Feito isso, o algoritmo repete o processo, criando um novo grupo,

até não existir mais nenhuma aresta com peso superior ao *threshold*;

5. Os pesquisadores que não estiveram em nenhum *cluster*, são inseridos um por um em grupos individuais;
6. Durante o processo, cada vez que alguma aresta é adicionada em um grupo, seu peso é alterado para zero. Assim em uma nova busca pelo grafo, aquela aresta não será incluída em nenhum outro grupo.

O pseudo-algoritmo da implementação pode ser visto no Algoritmo 1. O resultado da execução encontra-se no Apêndice A.1, onde é apresentado o arquivo de saída gerado pelo programa, com todos os grupos criados no processo. A saída é referente a um teste com doze currículos Lattes de profissionais.

### 3.5.1 Análise da Similaridade entre os Vértices

Os testes feitos com o algoritmo SCD não apresentaram resultados satisfatórios. Assim, para obter um melhor resultado de clusterização foi implementado, em um segundo momento, o algoritmo baseado no artigo *CollabSeer: A Search Engine for Collaboration Discovery* (CHEN et al., 2011).

Com base neste trabalho, foram implementados três métodos de similaridade (*Cosine*, *Jaccard* e *Overlap*) para, então, realizar o agrupamento dos autores com base em suas similaridades.

### 3.5.2 Criação dos Grupos

Após calculados todos os coeficientes de similaridade utilizando as três equações, foi aplicado o *framework* URSA. O algoritmo utilizado é implementado pela biblioteca, o *Best-Star*, tem como abordagem atribuir um objeto ao grupo de objetos que possui maior similaridade e realocar objetos não centróides, ou seja, que não são pontos iniciais da criação do grupo, para grupos com similaridade superior ao que está inserido.

O algoritmo obteve um resultado não muito satisfatório, em virtude de muitos autores colaboradores em artigos aparecerem com nomes divergentes em currículos distintos e dos autores sem currículo sendo



---

**Algoritmo 1:** Pseudo-algoritmo SCD.
 

---

```

início
  listaAresta = recuperaTodasArestasDoGrafo();
  threshold = 3;
  repita
    arestaMaior = null;
    grupo = new Grupo();
    para (cada listaAresta) faça
      peso = retornaPeso(aresta);
      se (peso > maior ∧ peso  $\geq$  threshold)
        então
          maior = peso;
          arestaMaior = aresta;
        fim
      grupo.add(arestaMaior);
      listaAresta.setPeso(arestaMaior,0);
      /* recupero arestas dos dois vértices da
         aresta maior. */
      arestasVertice = retornaArestas(arestaMaior);
      para (cada arestasVertice) faça
        peso = retornaPeso(arestasVertice);
        se (peso  $\geq$  threshold) então
          grupo.add(arestasVertice);
          listaAresta.setPeso(arestasVertice,0);
        fim
      fim
    fim
  até (!retornaPeso(listaAresta) > threshold);
  incluirGruposRestantes(listaAresta);
fim

```

---

processados gerarem muitos ruídos nos dados. Com isso, o algoritmo gerou uma série de grupos individuais.

Com o intuito de buscar melhores resultados para clusterização, foi realizada uma nova análise nos currículos e observado que os nomes que haviam muitas divergências em função de abreviações e inconsistências eram normalmente dos pesquisadores que não estavam no *dataset* de currículos Lattes utilizados para o trabalho, ou seja, quando o pesquisador insere uma nova pesquisa em seu currículo, seu nome é inserido de forma completa e correta, porém os de seus parceiros pesquisadores eram abreviados ou inseridos em formato de citação.

Desta forma, foi realizado um tratamento nos nomes que estavam em formato de citação, isto é, os nomes que estavam no formato "Sobrenome, Nome". No entanto, ainda assim, a clusterização não teve uma melhora considerável, em razão de que muitas vezes o autor possuir dois sobrenomes e seu nome era inserido no currículo de diversas formas. Na Figura 10 é possível observar alguns exemplos de nomes recuperados dos currículos de forma incorreta.

Figura 10 – Exemplos de nomes retirados dos currículos.

```
Dantas, M A R
Dantas, M
Mario Antonio Ribeiro Dantas
Carina Friedrich Dorneles
Friedrich Dorneles, Carina
JEAN CARLO R. HAUCK
JEAN CARLO HAUCK
C. Liu
Chao Liu
COMUNELLO, E
EROS COMUNELLO
ALDO VON WANGENHEIM
A VON WANGENHEIM
```

Fonte: Elaborada pela autora.

Muitos autores/coautores que não possuem currículos sendo processados geraram ruídos nos dados através de ligações unitárias, foi possível observar que os pesquisadores diversas vezes publicavam com um autor apenas uma vez o fato ocorre em decorrência de orientações de graduação ou mestrado, por exemplo, no qual a publicação realiza-se em virtude de um desenvolvimento de trabalho de conclusão de curso. Para minimizar o problema dos ruídos os autores/coautores que não possuíam currículos sendo processados geravam, foram alteradas as consultas do XPath na classe ManipuladorXML para, assim, de

cada elemento ARTIGO-PUBLICADO e TRABALHO-EM-EVENTOS ser retirado os títulos das respectivas publicações e apenas o nome do autor/coautor que corresponde ao pesquisador que o currículo pertence. As informações foram inseridas em um *HashMap* no qual o título da publicação é a chave do mapa e os autores são inseridos em uma lista nos valores. Dessa forma, toda vez que for encontrado em um currículo um título de trabalho ou publicação já disponível no mapa, o autor é inserido na lista de valores do *HashMap*.

Na análise dos currículos também foi observado que alguns títulos de mesmo artigo eram escritos com palavras diferentes. Testes foram feitos para calcular a similaridade entre títulos, porém o custo para processamento era alto e o resultado não trouxe melhoras para o problema, já que alguns títulos eram muito abreviados e a similaridade calculada era muito próxima a zero. Um exemplo de divergência de título de um mesmo artigo pode ser visto na Figura 11. Conforme a Figura 12 podemos observar que os autores do artigo são os mesmos.

Figura 11 – Representação de títulos diferentes do mesmo artigo.

<pre>&lt;DADOS-BASICOS-DO-ARTIGO NATUREZA="COMPLETO" TITULO-DO-ARTIGO="XML" ANO-DO-ARTIGO="2009" PAIS-DE-PUBLICACAO="" IDIOMA="Inglês" MEIO-DE-DIVULGACAO="MEIO_DIGITAL" HOME-PAGE-DO-TRABALHO="" FLAG-RELEVANCIA="SIM" DOI="10.1145/1815918.1815924" TITULO-DO-ARTIGO-INGLES="XML" FLAG-DIVULGACAO-CIENTIFICA="NAO"/&gt;</pre>	<pre>&lt;DADOS-BASICOS-DO-ARTIGO NATUREZA="COMPLETO" TITULO-DO-ARTIGO="XML: Some Papers in a Haystack" ANO-DO-ARTIGO="2009" PAIS-DE-PUBLICACAO="" IDIOMA="Inglês" MEIO-DE-DIVULGACAO="MEIO_DIGITAL" HOME-PAGE-DO-TRABALHO="" FLAG-RELEVANCIA="SIM" DOI="" TITULO-DO-ARTIGO-INGLES="" FLAG-DIVULGACAO-CIENTIFICA="NAO"/&gt;</pre>
---	--

Fonte: Currículo Lattes (Cnpq).

Após a retirada dos ruídos citados, os valores de similaridade entre os autores foram novamente calculados, sendo possível ver uma evolução nos resultados, conforme mostra a Figura 13. Esta abordagem é a forma final proposta e foi utilizada para os testes apresentados no próximo capítulo.

Figura 12 – Representação dos autores do artigo.

```

<AUTORES
NOME-COMPLETO-DO-AUTOR="Mirella Moro"
NOME-PARA-CITACAO="Moro, Mirella"
ORDEM-DE-AUTORIA="1"
NRO-ID-CNPQ="6408321790990372"/>
<AUTORES NOME-COMPLETO-DO-AUTOR="Vanessa de Paula Braganholo"
NOME-PARA-CITACAO="BRAGANHOLO, Vanessa de Paula"
ORDEM-DE-AUTORIA="2"
NRO-ID-CNPQ=""/>
<AUTORES NOME-COMPLETO-DO-AUTOR="Dênio Duarte"
NOME-PARA-CITACAO="DUARTE, D."
ORDEM-DE-AUTORIA="4"
NRO-ID-CNPQ="544229000261731"/>
<AUTORES NOME-COMPLETO-DO-AUTOR="Renata de Matos Galante"
NOME-PARA-CITACAO="GALANTE, Renata de Matos"
ORDEM-DE-AUTORIA="5"
NRO-ID-CNPQ=""/>
<AUTORES NOME-COMPLETO-DO-AUTOR="Ronaldo dos Santos Mello"
NOME-PARA-CITACAO="MELLO, R. S."
ORDEM-DE-AUTORIA="6"
NRO-ID-CNPQ="5011370918857999"/>
<AUTORES NOME-COMPLETO-DO-AUTOR="Carina Friedrich Dorneles"
NOME-PARA-CITACAO="DORNELES, C. F.;Dorneles, Carina F.
:Dorneles, Carina Friedrich"
ORDEM-DE-AUTORIA="1"/>
<AUTORES NOME-COMPLETO-DO-AUTOR="MORO, Mirella M."
NOME-PARA-CITACAO="Moro, Mirella M."
ORDEM-DE-AUTORIA="1"
NRO-ID-CNPQ="6408321790990372"/>
<AUTORES NOME-COMPLETO-DO-AUTOR="Vanessa de Paula Braganholo"
NOME-PARA-CITACAO="BRAGANHOLO, V. P."
ORDEM-DE-AUTORIA="2"
NRO-ID-CNPQ=""/>
<AUTORES NOME-COMPLETO-DO-AUTOR="Carina Friedrich Dorneles"
NOME-PARA-CITACAO="DORNELES, C. F."
ORDEM-DE-AUTORIA="3"
NRO-ID-CNPQ="0378897709136226"/>
<AUTORES NOME-COMPLETO-DO-AUTOR="Denio Duarte"
NOME-PARA-CITACAO="DUARTE, D."
ORDEM-DE-AUTORIA="4"
NRO-ID-CNPQ="544229000261731"/>
<AUTORES NOME-COMPLETO-DO-AUTOR="Renata de Matos Galante"
NOME-PARA-CITACAO="GALANTE, R. M."
ORDEM-DE-AUTORIA="5"
NRO-ID-CNPQ=""/>
<AUTORES NOME-COMPLETO-DO-AUTOR="Ronaldo dos Santos Mello"
NOME-PARA-CITACAO="MELLO, R. S."
ORDEM-DE-AUTORIA="6"/>

```

Fonte: Currículo Lattes (Cnpq).

Figura 13 – Similaridade entre alguns autores INE/UFSC.

```

Carina Friedrich Dorneles Vania Bogorny 0.5
Carina Friedrich Dorneles Carlos Becker Westphall 0.0
Carina Friedrich Dorneles Ronaldo dos Santos Mello 0.5
Ronaldo dos Santos Mello Vania Bogorny 0.5

```

Fonte: Currículo Lattes (Cnpq).

## 4 EXPERIMENTOS

Neste capítulo são descritos os experimentos realizados. Eles foram feitos com um conjunto de cinquenta currículos Lattes de professores do INE/UFSC, cinco testes foram feitos:

1. Aplicando o algoritmo *Best-Star* apenas nos pesquisadores que possuem currículos no *dataset* de currículos e calculando a similaridade através do método de *Jaccard*;
2. Aplicando o algoritmo *Best-Star* apenas nos pesquisadores que possuem currículos no *dataset* de currículos e calculando a similaridade através do método de *Cosine*;
3. Aplicando o algoritmo *Best-Star* apenas nos pesquisadores que possuem currículos no *dataset* de currículos e calculando a similaridade através do método de *Overlap*;
4. Aplicando o algoritmo *Best-Star* sobre os currículos dos autores de três grupos gerados por método (*Cosine*, *Overlap* e *Jaccard*), considerando todos os autores presentes no currículo;
5. Aplicando o algoritmo SCD apenas sobre os autores que possuem currículos no *dataset* de currículos.

### 4.1 RESULTADOS *BEST-STAR*

Com o objetivo de verificar o melhor método para cálculo do coeficiente de similaridade na geração dos *clusters*, foi executado o algoritmo *Best-Star* utilizando as três equações referentes ao artigo de Chen et al. (2011) com grau de similaridade de 0.1 até 0.9, para, assim, examinar qual o *threshold* de similaridade entre os autores seria mais adequado na geração de um bom resultado, ou seja, resultado no qual as relações entre os autores fossem coerentes e não houvesse uma grande geração de grupos unitários.

Para apurar a qualidade do resultado, os pesquisadores foram analisados antes da geração dos grupos com a finalidade de associar alguns autores do *dataset* utilizando como critério sua área de conhecimento e especialidade. As seguintes associações foram feitas:

1. Carla Merkle Westphall e Carlos Becker Westphall;

2. Ronaldo dos Santos Mello e Carina Friedrich Dorneles;
3. Christiane Anneliese Gresse von Wangenheim e Jean Carlo Rossa Hauck;
4. Jean Everson Martina e Ricardo Felipe Custodio;
5. Silvia Modesto Nassar e Renato Cislighi.

Dentre os resultados das três equações, a clusterização pelo método *Overlap* com o *threshold* de similaridade 0.5 gerou o melhor resultado. Ele gerou doze grupos nos quais apenas dois autores ficaram isolados. Utilizando este mesmo valor de similaridade, o coeficiente de *Jaccard* e *Cosine* geraram vinte e três e sete *clusters* com apenas um membro, respectivamente.

Utilizando o coeficiente de *Cosine*, o *threshold* de melhor resultado foi de valor 0.3 e no método de *Jaccard* o valor 0.2 apresentou o melhor ganho. O resultado do algoritmo por método pode ser analisado a partir do Apêndice A.2.

Um resumo da execução com as três equações pode ser visto nas Tabelas 1, 2, 3 e alguns *clusters* gerados podem ser visualizados na Figura 14, Figura 15 e Figura 16.

Tabela 1 – Resultado Coeficiente *Overlap*

Threshold	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Número de Clusters	12	12	12	12	12	19	22	22	22
Clusters com apenas um membro	2	2	2	2	2	9	14	14	14

Tabela 2 – Resultado Coeficiente *Jaccard*

Threshold	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Número de Clusters	12	13	20	26	29	32	33	33	33
Clusters com apenas um membro	2	2	9	18	23	28	30	30	30

Tabela 3 – Resultado Coeficiente *Cosine*

Threshold	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Número de Clusters	12	12	12	16	18	29	30	32	33
Clusters com apenas um membro	2	2	2	4	7	23	25	28	30

Pelo método *Overlap*, dos autores que foram associados anteriormente, apenas os pesquisadores Ronaldo dos Santos Mello com Carina Dorneles e Christiane Anneliese Gresse von Wangenheim com Jean Carlo Rossa Hauck ficaram no mesmo *cluster*. Isso ocorreu em virtude de que alguns autores que eram da mesma especialidade e publicaram juntos não possuíam membros em comum no *dataset*. No método de *Jaccard* e *Cosine* apenas os autores Ronaldo dos Santos Mello com Carina Dorneles ficaram no mesmo grupo.

Figura 14 – Cluster gerado pelo método Jaccard.

**Cluster 10**

&lt;Renato Cislighi&gt;

**Áreas:** [Administração, Ciência da Computação, Educação]**Especializações:** [Acompanhamento e avaliação de disciplinas e cursos, Administração de Unidades Educativas, Avaliação de Sistemas, Instituições, Planos e Programas Educacionais, Evasão discente, Organizações Públicas, Teleinformática]

&lt;Ricardo Azambuja Silveira&gt;

**Áreas:** [Administração, Ciência da Computação, Educação]**Especializações:** [Administração de Recursos Humanos, Arquitetura de Sistemas de Computação, Lógicas e Semântica de Programas, Tecnologia Educacional]

Fonte: Elaborada pela autora.

Figura 15 – Cluster gerado pelo método Overlap.

**Cluster 8**

<Carina Friedrich Dorneles>

**Áreas:** [Ciência da Computação]

**Especializações:** [Banco de Dados, Consulta Por Similaridade, Dados Semi Estruturados e Web, Documentos, Engenharia de Software, Extração de Informações, Interfaces, Sistemas de Informação, Xml]

<Ronaldo dos Santos Mello>

**Áreas:** [Ciência da Computação, Economia]

**Especializações:** [Arquitetura de Sistemas de Computação, Banco de Dados, Engenharia de Software, Information Retrieval, Linguagens de Programação, Métodos e Modelos Matemáticos, Econométricos e Estatísticos, Sistemas de Informação]

<Vania Bogorny>

**Áreas:** [Ciência da Computação]

**Especializações:** [Aprendizado de Máquina, Banco de Dados, Bancos de Dados Geográficos, Data Mining, Engenharia de Software, Integração Semântica, Mineração de Dados, Ontologias, Regras de Associação, Sistemas de Informação, Spatial Data Mining]

Fonte: Elaborada pela autora.



Figura 16 – Cluster gerado pelo método Cosine.

### Cluster 3

<Aldo von Wangenheim>

**Áreas:** [Ciência da Computação, Ciência da Informação, Comunicação, Educação, Engenharia Biomédica, Engenharia Elétrica, Engenharia Mecânica, Engenharia de Produção, Física, Geociências, Matemática, Medicina]

**Especializações:** [Arquitetura de Sistemas de Computação, Banco de Dados, Cancerologia, Cardiologia, Cirurgia Cardiovascular, Cirurgia Otorrinolaringológica, Climatologia, Educação Permanente, Educação de Adultos, Engenharia de Software, Garantia de Controle de Qualidade, Gerência do Projeto e do Produto, Instrumentação Meteorológica, Linguagem Formais e Autômatos, Linguagens de Programação, Matemática Discreta e Combinatória, Meteorologia Sinótica, Modelos Analíticos e de Simulação, Métodos e Técnicas de Ensino, Neurologia, Neuropatologia, Oftalmologia, Ortopedia, Otorrinolaringologia, Pediatria, Petrologia, Processamento Gráfico (Graphics), Processamento de Imagens, Processamento de Sinais Biológicos, Representação da Informação, Sensoriamento Remoto da Atmosfera, Sistemas de Informação, Sistemas de Telecomunicações, Tecnologia Educacional, Teleinformática, Telerradiologia, Videodifusão, prototipagem rápida, Ótica]

<Christiane Anneliese Gresse von Wangenheim>

**Áreas:** [Administração, Ciência da Computação, Educação, Engenharia de Produção, Psicologia]

**Especializações:** [Administração da Produção, Análise de Custos, Educação de Adultos, Engenharia de Software, Ensino e Aprendizagem na Sala de Aula, GERÊNCIA DE PROJETOS, Gestão de conhecimento, Inteligência Artificial, Interação Humano Computador, MELHORIA DE PROCESSO DE SOFTWARE, Métodos e Técnicas de Ensino, Sistemas de Informação, Software Básico]

<Vitório Bruno Mazzola>

**Áreas:** [Ciência da Computação, Engenharia Elétrica]

**Especializações:** [Controle de Processos Eletrônicos, Retroalimentação, Engenharia de Software, Linguagem Formais e Autômatos, Lógicas e Semântica de Programas, Redes de Computadores, Software Básico, Transmissão da Energia Elétrica, Distribuição da Energia Elétrica]

Fonte: Elaborada pela autora.

Outro exemplo a ser citado são os pesquisadores Carla Merkle West-phall e Carlos Becker Westphall. Ambos possuem juntos quarenta e uma publicações em comum, porém, ao realizar o cálculo de similaridade entre os dois, o valor resultou em zero devido ao fato dos autores não possuírem outros pesquisadores em comum no *dataset* além deles mesmos. Com isso, ao aplicar os métodos de similaridade, o numerador, ou seja, a intersecção entre eles, era zero e o cálculo resultava em zero. As publicações do Carlos com outros autores INE/UFSC presentes em seu currículo Lattes não incluía Carla Merkle Westphall como

coautora e vice-versa. Sendo assim, não existe intersecção de coautores entre eles. A Figura 17 mostra as publicações e os pesquisadores que o autor Carlos Becker Westphall realizou parceria e a Figura 18 mostra as publicações e os pesquisadores que Carla Merkle Westphall realizou em conjunto.

Figura 17 – Publicações com coautorias do Dr. Carlos Westphall.

vertice: Carlos Becker Westphall  
vizinhos: Carla Merkle Westphall, João Bosco Mangureira Sobral,  
Ricardo Felipe Custodio e Silvia Modesto Nassar

Carlos Becker Westphall e Silvia Modesto Nassar - 1  
1) GERÊNCIA DE TRÁFEGO DE REDES UTILIZANDO BASELINE BAYESIANA.

João Bosco Mangureira Sobral e Carlos Becker Westphall - 3  
1) SEGURANÇA E FLEXIBILIDADE DE APLICAÇÕES DE BANCO DE DADOS NA WEB  
2) IMPLEMENTAÇÃO DE GERENCIAMENTO DE SERVIÇOS DE TELECOMUNICAÇÕES UTILIZANDO CORBA, JAVA E REFLEXÃO COMPUTACIONAL  
3) GERENCIAMENTO DE SERVIÇOS DE TELECOMUNICAÇÕES UTILIZANDO CORBA E JAVA

Carlos Becker Westphall e Ricardo Felipe Custodio - 1  
1) ANALYTICAL MODEL TO EVALUATE THE PERFORMANCE OF MOBILE AGENTS  
IN A GENERIC NETWORK TOPOLOGY

Fonte: Elaborada pela autora.

Figura 18 – Publicações com coautorias da Dra. Carla Westphall.

vertice: Carla Merkle Westphall  
vizinhos: Carlos Becker Westphall, Lau Cheuk Lung e Paulo Jose de Freitas Filho.

Carla Merkle Westphall e Lau Cheuk Lung - 3  
1) COMUNICAÇÃO DE GRUPO SEGURA NO CORBA: INTEGRIDADE E CONFIDENCIALIDADE NO MIOP  
2) PROJETO, IMPLEMENTAÇÃO E AVALIAÇÃO DO JACOWEB SECURITY  
3) INTEGRATING SSL TO THE JACOWEB SECURITY FRAMEWORK: PROJECT AND IMPLEMENTATION

Paulo Jose de Freitas Filho e Carla Merkle Westphall - 1  
1) A SIMULATION MODEL FOR THE COMPARISON OF TWO MULTICOMPUTER ARCHITECTURES

Fonte: Elaborada pela autora.

#### 4.1.1 *Best-star* Utilizando Todos Autores

Para este teste, foram utilizados vinte e sete currículos de professores do INE/UFSC. A escolha dos pesquisadores foi feita através dos resultados dos testes com o algoritmo *Best-Star* da Seção 4.1, no qual foram extraídos os autores de três grupos não unitários de cada

método.

O resultado de todos os métodos, no entanto, não foi superior aos resultados da Seção 4.1, já que centenas de grupos unitários foram criados utilizando *threshold* de 0.1 à 0.9. Portanto, é vantajoso considerar apenas os autores dos currículos Lattes do *dataset*.

## 4.2 RESULTADOS SCD

Nesta seção são expostos os resultados referentes à execução utilizando o algoritmo SCD implementado na primeira parte deste trabalho. Os testes foram executados considerando apenas os autores dos currículos Lattes e com pesos de três à dez. Isso significa que o algoritmo agrupou apenas arestas (pares de pesquisadores) que possuem no mínimo três publicações em conjunto.

Os resultados dos testes também não foram superiores ao algoritmo *Best-Star*. Apesar de alguns grupos gerados nos dois métodos serem similares, muitos grupos unitários foram gerados no teste, o que deixa o resultado da Seção 4.1 ainda como o melhor resultado encontrado no trabalho. Na Figura 19 é possível visualizar algum dos *clusters* gerados com peso três no algoritmo SCD.

Os resultados obtidos neste método podem ser considerados mais adequado para achar autores que publicam juntos, no entanto, não para agrupar autores com "coautorias semelhantes".

## Grupo 0

(Carla Merkle Westphall : Carlos Becker Westphall) - 41.0  
(Carla Merkle Westphall : Lau Cheuk Lung) - 3.0  
(João Bosco Mangueira Sobral : Carlos Becker Westphall) - 3.0

## Grupo 1

(Christiane Anneliese Gresse von Wangenheim : Jean Carlo Rossa Hauck) - 26.0  
(Aldo von Wangenheim : Christiane Anneliese Gresse von Wangenheim) - 16.0  
(Aldo von Wangenheim : Jean Carlo Rossa Hauck) - 5.0  
(Aldo von Wangenheim : Rogerio de Almeida Richa) - 4.0  
(Aldo von Wangenheim : Mario Antonio Ribeiro Dantas) - 6.0

## Grupo 2

(Luciana de Oliveira Rech : Lau Cheuk Lung) - 11.0  
(Lau Cheuk Lung : Frank Augusto Siqueira) - 6.0

## Grupo 3

(Luiz Cláudio Villar dos Santos : Olinto Jose Varela Furtado) - 9.0  
(José Luís Almada Güntzel : Luiz Cláudio Villar dos Santos) - 5.0

## Grupo 4

(Jean Everson Martina : Ricardo Felipe Custodio) - 9.0

## Grupo 5

(Ronaldo dos Santos Mello : Carina Friedrich Dorneles) - 4.0

Figura 19 – Execução algoritmo SCD.

## 5 CONCLUSÃO E TRABALHOS FUTUROS

Este trabalho teve como objetivo implementar uma ferramenta que, ao receber uma série de currículos Lattes, em formato XML, gere grupos de pesquisadores através de suas coautorias. No primeiro momento do trabalho, os *clusters* foram gerados de acordo com o algoritmo SCD. Os resultados não foram satisfatórios devido ao fato do método ser mais adequado para localizar coautores que publicam juntos. Com isso, outros algoritmos de agrupamento foram executados visando uma aprimoração dos resultados.

Conforme pode-se concluir, foi necessário uma análise mais detalhada nos currículos, já que o conteúdo do documento diversas vezes não condiz com o esperado, uma vez que alguns pesquisadores não preenchem todos os campos de forma correta e completa em seu currículo Lattes. Com a utilização do XPath, foi possível retirar os dados necessários dos documentos em XML e aplicar métodos de similaridade com base em trabalhos relacionados, bem como, executar algoritmos de *clustering* implementados por bibliotecas já existentes.

Com isso, os grupos gerados continham os nomes dos pesquisadores, agrupados com a união da área de conhecimento e especialização de cada membro do grupo, o que permite concluir que os resultados da análise realizada na seção 4.1 é convincente, pois os grupos gerados através de cada método de similaridade resulta em *clusters* coerentes. Os testes das subseção e seção 4.1.1 e 4.2 não demonstraram um melhor ganho para o trabalho, em virtude do excesso de grupos unitários não compensa considerar todos os autores presentes nos currículos para clusterização com base em coautorias, assim como o algoritmo SCD implementado na primeira parte do trabalho.

Contudo, a capacidade de definir grupos de pesquisadores com base em suas coautorias em comum é uma forma de estimular o compartilhamento de interesses e conhecimento de profissionais. Como trabalho futuro, os experimentos podem ser mais amplos e considerando as expertises dos autores, visando verificar se comumente relacionam-se. Para isso o problema de escalabilidade precisa ser aprimorado, já que com muitos currículos o processo de clusterização tende a ser lento na implementação atual.



## **APÊNDICE A - Resultados da implementação**





## A.1 *CLUSTERS* GERADOS NA EXECUÇÃO COM ALGORITMO SCD

### Grupo 0

(Aline Pereira Soares : Marina Keiko Nakayama) - 5.0

(Aline Pereira Soares : Ricardo Azambuja Silveira) - 5.0

(Aline Pereira Soares : Luis Roque Klering)

- 3.0

### Grupo 1

(Lucimar Maria Fossatti de Carvalho : Silvia Modesto Nassar)

- 5.0

(Lucimar Maria Fossatti de Carvalho : Fernando Mendes Azevedo) - 3.0

(Lucimar Maria Fossatti de Carvalho : Hugo José Teixeira de Carvalho) - 3.0

(Cristiane Koehler : Silvia Modesto Nassar)

- 3.0

(Cristiane Koehler : Maria Marlene de Souza Pires) - 3.0

(Beatriz Wilges : Silvia Modesto Nassar) - 4.0

(Beatriz Wilges : Ricardo Azambuja Silveira) - 3.0

(Gustavo Pereira Mateus : Beatriz Wilges) - 4.0

(Lucia Hisako Takase Gonçalves :

Silvia Modesto Nassar) - 5.0

### Grupo 2

(Patrícia Cristiane de Souza : Raul Sidnei Wazlawick) - 3.0

### Grupo 3

(Júlia Marques Carvalho da Silva : Ricardo

Azambuja Silveira) - 3.0

### Grupo 4

(Vania Bogorny : Luis Otavio Alvares) - 3.0

### Grupo 5

(Abderrahim Sekkaki : Carlos Becker Westphall) - 3.0

(Paulo Fernando Silva : Carlos Becker Westphall) - 3.0

(Shirlei Aparecida de Chaves : Carlos Becker Westphall) - 3.0

(Carlos Becker Westphall : Jorge Werner) - 3.0

(Carlos Becker Westphall : Guilherme Arthur Geronimo) - 3.0

(Carlos Becker Westphall : Sergio Roberto Villarreal) - 3.0

### Grupo 6

- (Alexandre Lazaretti Zanatta : Patrícia Vilain) - 1.0  
 Grupo 7  
 (SCHEIDT, RAFAEL DE FARIA : Patrícia Vilain) - 1.0  
 Grupo 8  
 (SCHEIDT, RAFAEL DE FARIA : DANTAS, M A R) - 1.0  
 Grupo 9  
 (FERNANDO MARTINELI LOUREIRO : Paulo Jose de Freitas Filho) - 1.0  
 Grupo 10  
 (Nelkis De La Orden Medina : Paulo Jose de Freitas Filho) - 2.0  
 Grupo 11  
 (RIVALINO MATIAS JUNIOR : Paulo Jose de Freitas Filho) - 1.0  
 Grupo 12  
 (RIVALINO MATIAS JUNIOR : Pedro Alberto BARBETTA) - 1.0  
 Grupo 13  
 (RIVALINO MATIAS JUNIOR : Kishor S. Trivedi) - 1.0  
 Grupo 14  
 (Mariana Dehon Costa Lima : Silvia Nassar) - 1.0  
 Grupo 15  
 (Mariana Dehon Costa Lima : Paulo Jose de Freitas Filho) - 1.0  
 Grupo 16  
 (Mariana Dehon Costa Lima : Carlos Magno C Jacinto) - 2.0  
 Grupo 17  
 (Carlos Magno C Jacinto : Paulo Jose de Freitas Filho) - 1.0  
 Grupo 18  
 (Carlos Magno C Jacinto : Silvia Nassar) - 1.0  
 Grupo 19  
 (Carlos Magno C Jacinto : Mauro Roisenberg) - 1.0  
 Grupo 20  
 (Carlos Magno C Jacinto : Diego Garcia Rodrigues) - 1.0  
 Grupo 21  
 (Vânia Ribas Ulbricht : Neri dos Santos) - 1.0  
 Grupo 22  
 (Vânia Ribas Ulbricht : Raul Sidnei Wazlawick) - 1.0  
 Grupo 23  
 (Patrícia Cristiane de Souza : Augusto Bohner Hoffmann) - 1.0  
 Grupo 24  
 (Antonio Carlos da Rocha Costa : Raul Sidnei Wazlawick) - 1.0  
 Grupo 25

- (Sérgio Luiz de Medeiros Rivero : Bernd Heinrich Storb) - 1.0  
Grupo 26
- (Sérgio Luiz de Medeiros Rivero : Raul Sidnei Wazlawick) - 1.0  
Grupo 27
- (Adja Ferreira de Andrade : Antonio Carlos Mariani) - 1.0  
Grupo 28
- (Adja Ferreira de Andrade : Raul Sidnei Wazlawick) - 1.0  
Grupo 29
- (Ahmed Ali Abdala Esmin : Raul Sidnei Wazlawick) - 1.0  
Grupo 30
- (Marta Costa Rosatelli : Edla Maria Faust Ramos) - 1.0  
Grupo 31
- (Marta Costa Rosatelli : Walter de Abreu Cybis) - 1.0  
Grupo 32
- (Marta Costa Rosatelli : Bernd Heinrich Storb) - 1.0  
Grupo 33
- (Marta Costa Rosatelli : Vera Rejane Niederberg Schuhmacher)  
- 1.0  
Grupo 34
- (Marta Costa Rosatelli : Antonio Carlos Mariani) - 1.0  
Grupo 35
- (Marta Costa Rosatelli : Tereza Gonçalves Kirner) - 1.0  
Grupo 36
- (Marta Costa Rosatelli : Claudio Kirner) - 1.0  
Grupo 37
- (Marta Costa Rosatelli : Lea da Cruz Fagundes) - 1.0  
Grupo 38
- (Marta Costa Rosatelli : Raul Sidnei Wazlawick) - 1.0  
Grupo 39
- (Ned Kock : Robert Davison) - 1.0  
Grupo 40
- (Ned Kock : Rosalie Ocker) - 1.0  
Grupo 41
- (Ned Kock : Raul Sidnei Wazlawick) - 1.0  
Grupo 42
- (Tereza Gonçalves Kirner : Lea da Cruz Fagundes) - 1.0  
Grupo 43
- (Tereza Gonçalves Kirner : Bernd Heinrich Storb) - 1.0  
Grupo 44
- (Tereza Gonçalves Kirner : Raul Sidnei Wazlawick) - 1.0  
Grupo 45

- (Patrícia Cristiane de Souza : Marta Costa Rosatelli) - 1.0  
 Grupo 46  
 (Cláudio Magalhães de Oliveira : Raul Sidnei Wazlawick) - 1.0  
 Grupo 47  
 (Cristiano Tolfo : Raul Sidnei Wazlawick) - 2.0  
 Grupo 48  
 (Humberto Fioravante Ferro : Rogério Cid Bastos) - 1.0  
 Grupo 49  
 (Humberto Fioravante Ferro : Cláudio Magalhães de Oliveira) -  
 1.0  
 Grupo 50  
 (Humberto Fioravante Ferro : Raul Sidnei Wazlawick) - 1.0  
 Grupo 51  
 (Cristiano Tolfo : Fernando Antonio Forcelini) - 1.0  
 Grupo 52  
 (Cristiano Tolfo : Marcelo Gitirana Gomes Ferreira) - 1.0  
 Grupo 53  
 (Tolfo, Cristiano : Marcelo Gitirana Gomes Ferreira) - 1.0  
 Grupo 54  
 (Tolfo, Cristiano : Forcellini, Fernando Antonio) - 1.0  
 Grupo 55  
 (Tolfo, Cristiano : Raul Sidnei Wazlawick) - 1.0  
 Grupo 56  
 (Analúcia S. M. De Franceschi : K. S. Borges) - 1.0  
 Grupo 57  
 (Analúcia S. M. De Franceschi : Francisco Vasques) - 1.0  
 Grupo 58  
 (Analúcia S. M. De Franceschi : Ricardo Alexandre Reinaldo de  
 Moraes) - 1.0  
 Grupo 59  
 (Paulo Portugal : Francisco Vasques) - 1.0  
 Grupo 60  
 (Paulo Portugal : José Alberto Fonseca) - 1.0  
 Grupo 61  
 (Paulo Portugal : Ricardo Alexandre Reinaldo de Moraes) - 2.0  
 Grupo 62  
 (Ricardo Alexandre Reinaldo de Moraes : Vasques, Francisco) -  
 2.0  
 Grupo 63  
 (Ricardo Alexandre Reinaldo de Moraes : Portugal, P.) - 1.0  
 Grupo 64

- (Ricardo Alexandre Reinaldo de Moraes : Souto, Pedro F.) - 1.0  
 Grupo 65  
 (Ricardo Alexandre Reinaldo de Moraes : Francisco Vasques) -  
 2.0  
 Grupo 66  
 (Ricardo Alexandre Reinaldo de Moraes : Carreiro, Francisco  
 Borges) - 1.0  
 Grupo 67  
 (Ricardo Alexandre Reinaldo de Moraes : Bartolomeu, Paulo) -  
 1.0  
 Grupo 68  
 (Ricardo Alexandre Reinaldo de Moraes : Silva, Valter) - 1.0  
 Grupo 69  
 (Ricardo Alexandre Reinaldo de Moraes : Fonseca, José Alberto)  
 - 1.0  
 Grupo 70  
 (Ricardo Alexandre Reinaldo de Moraes : Portugal, Paulo) - 1.0  
 Grupo 71  
 (Ricardo Alexandre Reinaldo de Moraes : Ricardo Felipe Custó-  
 dio) - 1.0  
 Grupo 72  
 (VIEGAS, RAIMUNDO : AFFONSO, LUIZ) - 1.0  
 Grupo 73  
 (VIEGAS, RAIMUNDO : Vasques, Francisco) - 2.0  
 Grupo 74  
 (VIEGAS, RAIMUNDO : Portugal, Paulo) - 2.0  
 Grupo 75  
 (VIEGAS, RAIMUNDO : Ricardo Alexandre Reinaldo de Mo-  
 raes) - 2.0  
 Grupo 76  
 (VIEGAS, RAIMUNDO : GUEDES, LUIZ AFFONSO) - 1.0  
 Grupo 77  
 (Carlos M. D. Veigas : Francisco Vasques) - 1.0  
 Grupo 78  
 (Carlos M. D. Veigas : Paulo Portugal) - 1.0  
 Grupo 79  
 (Carlos M. D. Veigas : Ricardo Alexandre Reinaldo de Moraes)  
 - 1.0  
 Grupo 80  
 (Evandro César Freiburger : Ricardo Pereira e Silva) - 1.0  
 Grupo 81

- (Ricardo Azambuja Silveira : Rosa Vicari) - 1.0  
 Grupo 82  
 (André Bastos : Rita Cristina Galarraga Bernardi) - 1.0  
 Grupo 83  
 (André Bastos : Ricardo Azambuja Silveira) - 1.0  
 Grupo 84  
 (Mário Lúcio Mesquita Machado : Diego Gomes de Souza) - 1.0  
 Grupo 85  
 (Mário Lúcio Mesquita Machado : João Artur de Souza) - 1.0  
 Grupo 86  
 (Mário Lúcio Mesquita Machado : Gertrudes A Dandolini) - 1.0  
 Grupo 87  
 (Mário Lúcio Mesquita Machado : Ricardo Azambuja Silveira) -  
 1.0  
 Grupo 88  
 (Beatriz Wilges : Joel Pinho Lucas) - 2.0  
 Grupo 89  
 (Joel Pinho Lucas : Beatriz Widges) - 2.0  
 Grupo 90  
 (Joel Pinho Lucas : Samuel Elias da Silva Salomão) - 1.0  
 Grupo 91  
 (Joel Pinho Lucas : Ricardo Azambuja Silveira) - 1.0  
 Grupo 92  
 (Raymundo Carlos Machado Ferreira Filho : Nilo César Consoli)  
 - 1.0  
 Grupo 93  
 (Raymundo Carlos Machado Ferreira Filho : Rosa Maria Vicari)  
 - 1.0  
 Grupo 94  
 (Raymundo Carlos Machado Ferreira Filho : Fernando Schnaid)  
 - 1.0  
 Grupo 95  
 (Raymundo Carlos Machado Ferreira Filho : Ricardo Azambuja  
 Silveira) - 1.0  
 Grupo 96  
 (Beatriz Widges : Ricardo Azambuja Silveira) - 2.0  
 Grupo 97  
 (Kelly Hannel : Verônica Burmann da Silva) - 1.0  
 Grupo 98  
 (Kelly Hannel : Raymundo Carlos Machado Ferreira Filho) - 1.0  
 Grupo 99

- (Kelly Hannel : Ricardo Azambuja Silveira) - 1.0  
 Grupo 100  
 (Beatriz Widges : Silvia Modesto Nassar) - 1.0  
 Grupo 101  
 (Júlia Marques Carvalho da Silva : Natanael Bavaresco) - 1.0  
 Grupo 102  
 (Eduardo Rodrigues Gomes : Ricardo Azambuja Silveira) - 1.0  
 Grupo 103  
 (Néstor Darío Duque Méndez : Demetrio Arturo Ovalle Car-  
 ranza) - 1.0  
 Grupo 104  
 (Néstor Darío Duque Méndez : Rosa Vicari) - 1.0  
 Grupo 105  
 (Néstor Darío Duque Méndez : Ricardo Azambuja Silveira) - 1.0  
 Grupo 106  
 (Júlia Marques Carvalho da Silva : Rosa Maria Vicari) - 1.0  
 Grupo 107  
 (Aline Pereira Soares : Patrícia de Sá Freire) - 2.0  
 Grupo 108  
 (Aline Pereira Soares : Andressa Sasaki Vasques Pacheco) - 2.0  
 Grupo 109  
 (Aline Pereira Soares : Fernando Spanhol) - 1.0  
 Grupo 110  
 (Arnoldo Uber Júnior : Ricardo Azambuja Silveira) - 1.0  
 Grupo 111  
 (Andressa Sasaki Vasques Pacheco : Mauricio Rissi) - 1.0  
 Grupo 112  
 (Andressa Sasaki Vasques Pacheco : Marina Keiko Nakayama) -  
 1.0  
 Grupo 113  
 (Andressa Sasaki Vasques Pacheco : Ricardo Azambuja Silveira)  
 - 1.0  
 Grupo 114  
 (Andressa Sasaki Vasques Pacheco : Fernando Spanhol) - 1.0  
 Grupo 115  
 (Gustavo Pereira Mateus : Silvia Modesto Nassar) - 1.0  
 Grupo 116  
 (Gustavo Pereira Mateus : Ricardo Azambuja Silveira) - 1.0  
 Grupo 117  
 (Gustavo Pereira Mateus : Rogério Cid Bastos) - 1.0  
 Grupo 118

- (Roberto Rabello : Liliane Passerino) - 1.0  
 Grupo 119  
 (Roberto Rabello : Rosa Vicari) - 1.0  
 Grupo 120  
 (Roberto Rabello : Ricardo Azambuja Silveira) - 1.0  
 Grupo 121  
 (Rafaela Comarella : Ronaldo Campos) - 2.0  
 Grupo 122  
 (Rafaela Comarella : Ricardo Azambuja Silveira) - 1.0  
 Grupo 123  
 (Rafaela Comarella : Araci Hack Catapan) - 1.0  
 Grupo 124  
 (Aline Pereira Soares : Gustavo Lucas) - 1.0  
 Grupo 125  
 (Aline Pereira Soares : Fernando José Spanhol) - 1.0  
 Grupo 126  
 (Marina Keiko Nakayama : Bianca Smith Pilla) - 1.0  
 Grupo 127  
 (Marina Keiko Nakayama : Erlaine Binotto) - 1.0  
 Grupo 128  
 (Marina Keiko Nakayama : Ricardo Azambuja Silveira) - 1.0  
 Grupo 129  
 (Ronaldo Campos : Ricardo Azambuja Silveira) - 1.0  
 Grupo 130  
 (Paula Andrea Rodríguez Marín : ulián Moreno Cadavid) - 1.0  
 Grupo 131  
 (Paula Andrea Rodríguez Marín : Néstor Darío Duque Méndez)  
 - 1.0  
 Grupo 132  
 (Paula Andrea Rodríguez Marín : Demetrio Arturo Ovalle Car-  
 ranza) - 1.0  
 Grupo 133  
 (Paula Andrea Rodríguez Marín : Ricardo Azambuja Silveira) -  
 1.0  
 Grupo 134  
 (Konietschke Rainer : Zerbato D.) - 1.0  
 Grupo 135  
 (Konietschke Rainer : Tobergte A.) - 1.0  
 Grupo 136  
 (Konietschke Rainer : Poignet, P.) - 1.0  
 Grupo 137



- (Konietschke Rainer : Fröhlich, F.) - 1.0  
Grupo 138
- (Konietschke Rainer : Botturi, D.) - 1.0  
Grupo 139
- (Konietschke Rainer : Fiorini, P.) - 1.0  
Grupo 140
- (Konietschke Rainer : Hirzinger, G.) - 1.0  
Grupo 141
- (Konietschke Rainer : Rogerio de Almeida Richa) - 1.0  
Grupo 142
- (Bó, Antônio P.L. : Poignet, Philippe) - 1.0  
Grupo 143
- (Bó, Antônio P.L. : Rogerio de Almeida Richa) - 1.0  
Grupo 144
- (Poignet, P. : Chao Liu,) - 1.0  
Grupo 145
- (Poignet, P. : Rogerio de Almeida Richa) - 1.0  
Grupo 146
- (Balicki, M. : Sznitman, R.) - 1.0  
Grupo 147
- (Balicki, M. : Meisner, E.) - 1.0  
Grupo 148
- (Balicki, M. : Russell Taylor) - 1.0  
Grupo 149
- (Balicki, M. : Hager, G.) - 1.0  
Grupo 150
- (Balicki, M. : Rogerio de Almeida Richa) - 1.0  
Grupo 151
- (Raphael Sznitman : Bruno Jedynek) - 1.0  
Grupo 152
- (Raphael Sznitman : Russell Taylor) - 1.0  
Grupo 153
- (Raphael Sznitman : Gregory Hager) - 1.0  
Grupo 154
- (Raphael Sznitman : Rogerio de Almeida Richa) - 1.0  
Grupo 155
- (Rogerio de Almeida Richa : LINHARES, RODRIGO) - 1.0  
Grupo 156
- (Rogerio de Almeida Richa : COMUNELLO, EROS) - 1.0  
Grupo 157
- (Rogerio de Almeida Richa : VON WANGENHEIM, ALDO) -

## 1.0

- Grupo 158  
(Rogerio de Almeida Richa : SCHNITZLER, JEAN-YVES) - 1.0
- Grupo 159  
(Rogerio de Almeida Richa : WASSMER, BENJAMIN) - 1.0
- Grupo 160  
(Rogerio de Almeida Richa : GUILLEMOT, CLAIRE) - 1.0
- Grupo 161  
(Rogerio de Almeida Richa : THURET, GILLES) - 1.0
- Grupo 162  
(Rogerio de Almeida Richa : GAIN, PHILIPPE) - 1.0
- Grupo 163  
(Rogerio de Almeida Richa : HAGER, GREGORY) - 1.0
- Grupo 164  
(Rogerio de Almeida Richa : TAYLOR, RUSSELL) - 1.0
- Grupo 165  
(GUILLEMOT, C : Rogerio de Almeida Richa) - 1.0
- Grupo 166  
(GUILLEMOT, C : COMUNELLO, E) - 1.0
- Grupo 167  
(GUILLEMOT, C : VON WANGENHEIM, A) - 1.0
- Grupo 168  
(GUILLEMOT, C : SCHNITZLER, JY) - 1.0
- Grupo 169  
(GUILLEMOT, C : WASSMER, B) - 1.0
- Grupo 170  
(GUILLEMOT, C : HAGER, G) - 1.0
- Grupo 171  
(GUILLEMOT, C : TAYLOR, R) - 1.0
- Grupo 172  
(GUILLEMOT, C : THURET, G) - 1.0
- Grupo 173  
(GUILLEMOT, C : GAIN, P) - 1.0
- Grupo 174  
(Silvana Castano : Carlos Alberto Heuser) - 1.0
- Grupo 175  
(Silvana Castano : Ronaldo dos Santos Mello) - 1.0
- Grupo 176  
(Schroeder, Rebeca : Ronaldo dos Santos Mello) - 1.0
- Grupo 177  
(MORO, Mirella M. : Vanessa de Paula Braganholo) - 1.0

- Grupo 178  
(MORO, Mirella M. : Carina Friedrich Dorneles) - 1.0
- Grupo 179  
(MORO, Mirella M. : Denio Duarte) - 1.0
- Grupo 180  
(MORO, Mirella M. : Renata de Matos Galante) - 1.0
- Grupo 181  
(MORO, Mirella M. : Ronaldo dos Santos Mello) - 1.0
- Grupo 182  
(Ramesh Pinnamanemi : Juliana Freire) - 1.0
- Grupo 183  
(Ramesh Pinnamanemi : Ronaldo dos Santos Mello) - 1.0
- Grupo 184  
(Dorneles, Carina Friedrich : Rodrigo Gonçalves) - 1.0
- Grupo 185  
(Dorneles, Carina Friedrich : Ronaldo dos Santos Mello) - 1.0
- Grupo 186  
(Arthur Pereira Frantz : Helo Petry) - 1.0
- Grupo 187  
(Arthur Pereira Frantz : Ronaldo dos Santos Mello) - 1.0
- Grupo 188  
(Renato Fileto : Carina Friedrich Dorneles) - 1.0
- Grupo 189  
(Renato Fileto : Vânia Bogorny) - 1.0
- Grupo 190  
(Renato Fileto : Ronaldo dos Santos Mello) - 1.0
- Grupo 191  
(Rebeca Schroeder : Denio Duarte) - 1.0
- Grupo 192  
(Rebeca Schroeder : Ronaldo dos Santos Mello) - 1.0
- Grupo 193  
(C Celso Brasil Camargo : Silvia Modesto Nassar) - 1.0
- Grupo 194  
(C Celso Brasil Camargo : Cristiano Almeida Cunha) - 1.0
- Grupo 195  
(José Gonçalo dos Santos : Silvia Modesto Nassar) - 1.0
- Grupo 196  
(Lucimar F Carvalho : Cristiane Koehler) - 1.0
- Grupo 197  
(Lucimar F Carvalho : Silvia Modesto Nassar) - 1.0
- Grupo 198

- (Lucimar F Carvalho : Fernando M Azevedo) - 1.0  
 Grupo 199  
 (Rosemary Andrade Lentz : Regina G Santini Costenaro) - 1.0  
 Grupo 200  
 (Rosemary Andrade Lentz : Lucia H T Gonçalves) - 1.0  
 Grupo 201  
 (Rosemary Andrade Lentz : Silvia Modesto Nassar) - 1.0  
 Grupo 202  
 (Lucimar F de Carvalho : Cristiane Koeller) - 1.0  
 Grupo 203  
 (Lucimar F de Carvalho : Silvia Modesto Nassar) - 2.0  
 Grupo 204  
 (Lucimar F de Carvalho : Fernando M de Azevedo) - 2.0  
 Grupo 205  
 (Lucimar F de Carvalho : Hugo T de Carvalho) - 1.0  
 Grupo 206  
 (Lucimar Maria Fossatti de Carvalho : Fernando Mendes de Azevedo) - 2.0  
 Grupo 207  
 (Maria Marlene de Souza Pires : Silvia Modesto Nassar) - 1.0  
 Grupo 208  
 (Maria Marlene de Souza Pires : Mauricio José Pereima) - 1.0  
 Grupo 209  
 (Denise Bousfield da Silva : Maria Marlene de Souza Pires) - 1.0  
 Grupo 210  
 (Denise Bousfield da Silva : Silvia Modesto Nassar) - 1.0  
 Grupo 211  
 (Nilza Medeiros Perin : Maria Marlene de Souza Pires) - 1.0  
 Grupo 212  
 (Nilza Medeiros Perin : Silvia Modesto Nassar) - 1.0  
 Grupo 213  
 (Silvia Modesto Nassar : Lucia H T Gonçalves) - 1.0  
 Grupo 214  
 (L M F Carvalho : Silvia Modesto Nassar) - 2.0  
 Grupo 215  
 (L M F Carvalho : Fernando Mendes de Azevedo) - 2.0  
 Grupo 216  
 (L M F Carvalho : C A Dani) - 2.0  
 Grupo 217  
 (L M F Carvalho : H J T Carvalho) - 1.0  
 Grupo 218

- (Lucimar Fossatti de Carvalho : Silvia Modesto Nassar) - 2.0  
Grupo 219
- (Lucimar Fossatti de Carvalho : Fernando M Azevedo) - 1.0  
Grupo 220
- (Lucimar Fossatti de Carvalho : Hugo Jose T. Carvalho) - 1.0  
Grupo 221
- (Lucimar Fossatti de Carvalho : Lucas L. Monteiro) - 1.0  
Grupo 222
- (Lucimar Fossatti de Carvalho : Ciciliana Maila Zílio Rech) - 1.0  
Grupo 223
- (Lucimar F. Carvalho : Fernando M de Azevedo) - 1.0  
Grupo 224
- (Lucimar F. Carvalho : Silvia Modesto Nassar) - 1.0  
Grupo 225
- (Lucimar F. Carvalho : Hugo T de Carvalho) - 1.0  
Grupo 226
- (Lucimar F. Carvalho : Lucas L. Monteiro) - 1.0  
Grupo 227
- (Beatriz Wilges : Rogério Cid Bastos) - 2.0  
Grupo 228
- (Márcia Barros de Sales : Andréia Alves dos Santos) - 1.0  
Grupo 229
- (Márcia Barros de Sales : Silvia Modesto Nassar) - 1.0  
Grupo 230
- (LEONARDI, J. : Rogério Cid Bastos) - 1.0  
Grupo 231
- (LEONARDI, J. : Silvia Modesto Nassar) - 1.0  
Grupo 232
- (Beatriz Wilges : Juliana Leonardi) - 1.0  
Grupo 233
- (Beatriz Wilges : Araci Hack Catapan) - 1.0  
Grupo 234
- (Beatriz Wilges : Rogerio Cid Bastos) - 2.0  
Grupo 235
- (Penz, Cesar A. : Flesch, Carlos A.) - 1.0  
Grupo 236
- (Penz, Cesar A. : Silvia Modesto Nassar) - 1.0  
Grupo 237
- (Penz, Cesar A. : Flesch, Rodolfo C.C.) - 1.0  
Grupo 238
- (Penz, Cesar A. : de Oliveira, Marco A.) - 1.0

- Grupo 239  
(Felipe Schneider Costa : Maria Marlene de Souza Pires) - 1.0
- Grupo 240  
(Felipe Schneider Costa : Silvia Modesto Nassar) - 1.0
- Grupo 241  
(LSM da Silva : M Fisberg) - 1.0
- Grupo 242  
(LSM da Silva : MM de Souza Pires) - 1.0
- Grupo 243  
(LSM da Silva : Silvia Modesto Nassar) - 1.0
- Grupo 244  
(LSM da Silva : CB Sottovia) - 1.0
- Grupo 245  
(Mariana D.C. Lima : Silvia Modesto Nassar) - 1.0
- Grupo 246  
(Mariana D.C. Lima : Pedro Ivo R.B.G. Rodrigues) - 1.0
- Grupo 247  
(Mariana D.C. Lima : Paulo J. Freitas Filho) - 1.0
- Grupo 248  
(Mariana D.C. Lima : Carlos M.C. Jacinto) - 1.0
- Grupo 249  
(Fábio Bif Goularte : Beatriz Wilges) - 1.0
- Grupo 250  
(Fábio Bif Goularte : Silvia Modesto Nassar) - 1.0
- Grupo 251  
(Cleverson Alessandro Veronez : Carlos Becker Westphall) - 1.0
- Grupo 252  
(Cleverson Alessandro Veronez : Silvia Modesto Nassar) - 1.0
- Grupo 253  
(Lucimar Fossatti de Carvalho : Fernando Mendes Azevedo) -
- 1.0
- Grupo 254  
(Lucimar Fossatti de Carvalho : Lucas Lese Monteiro) - 1.0
- Grupo 255  
(Lucimar Fossatti de Carvalho : Hugo José Teixeira de Carvalho)
- 1.0
- Grupo 256  
(Lucimar Fossatti de Carvalho : Ciciliana Maíla Zilio Rech) - 1.0
- Grupo 257  
(Lucia Hisako Takase Gonçalves : POLARO, SANDRA HE-  
LENA ISSE) - 1.0

- Grupo 258  
 (Lucia Hisako Takase Gonçalves : LOPES, MÁRCIA MARIA BRAGANÇA) - 1.0
- Grupo 259  
 (Lucia Hisako Takase Gonçalves : FERREIRA, VIVIANE FER- RAZ) - 1.0
- Grupo 260  
 (Lucia Hisako Takase Gonçalves : MONTEIRO, HELLEN KA- RINNA) - 1.0
- Grupo 261  
 (Lucia Hisako Takase Gonçalves : Maria Francisca Santos Daussy) - 1.0
- Grupo 262  
 (Lucia Hisako Takase Gonçalves : Silvia Maria Azevedo dos San- tos) - 1.0
- Grupo 263  
 (Lucia Hisako Takase Gonçalves : Angela Maria Alvarez) - 1.0
- Grupo 264  
 (Lucia Hisako Takase Gonçalves : Maria Arminda da Silva Men- des Carneiro da Costa) - 1.0
- Grupo 265  
 (Lucia Hisako Takase Gonçalves : Maria Manuela F.P. da S. Martins) - 1.0
- Grupo 266  
 (Lucia Hisako Takase Gonçalves : Roberta Zunino) - 1.0
- Grupo 267  
 (Lucimar Maria Fossatti de Carvalho : Cândice Abella Dani) - 2.0
- Grupo 268  
 (Lucimar Maria Fossatti de Carvalho : Lucas Lese Monteiro) - 1.0
- Grupo 269  
 (Lucimar Maria Fossatti de Carvalho : Cristiane Koehler) - 1.0
- Grupo 270  
 (Lucimar Maria Fossatti de Carvalho : Fernanda Ramos) - 1.0
- Grupo 271  
 (Clarisse Araújo Cassol : Christine Prim de Pellegrin) - 1.0
- Grupo 272  
 (Clarisse Araújo Cassol : Mônica Lisboa Chang Wayhs) - 1.0
- Grupo 273  
 (Clarisse Araújo Cassol : Maria Marlene de Souza Pires) - 1.0

- Grupo 274  
(Clarisse Araújo Cassol : Silvia Modesto Nassar) - 1.0
- Grupo 275  
(Marilza Leal Nascimento : Maria Marlene de Souza Pires) - 1.0
- Grupo 276  
(Marilza Leal Nascimento : Silvia Modesto Nassar) - 1.0
- Grupo 277  
(MAURICIO L. SILVA : Silvia Modesto Nassar) - 1.0
- Grupo 278  
(MAURICIO L. SILVA : ANDRÉ P. SILVA) - 1.0
- Grupo 279  
(MAURICIO L. SILVA : LEANDRO L. PONCE) - 1.0
- Grupo 280  
(MAURICIO L. SILVA : MARIA M. DE S. PIRES) - 1.0
- Grupo 281  
(SILVANA MARIA DE MIRANDA : Maria Marlene de Souza Pires) - 1.0
- Grupo 282  
(SILVANA MARIA DE MIRANDA : Silvia Modesto Nassar) - 1.0
- Grupo 283  
(SILVANA MARIA DE MIRANDA : CARLOS ALBERTO JUSTO DA SILVA) - 1.0
- Grupo 284  
(C. M. LORENZO : XIKOTA, JC) - 1.0
- Grupo 285  
(C. M. LORENZO : WAYHS, M. L. C.) - 1.0
- Grupo 286  
(C. M. LORENZO : Silvia Modesto Nassar) - 1.0
- Grupo 287  
(C. M. LORENZO : Maria Marlene de Souza Pires) - 1.0
- Grupo 288  
(Juliana Cristina Lessmann : Denise Maria Guerreiro Vieira da Silva) - 1.0
- Grupo 289  
(Juliana Cristina Lessmann : Silvia Modesto Nassar) - 1.0
- Grupo 290  
(Vania Bogorny : Bart Kuijpers) - 2.0
- Grupo 291  
(Bart Kuijpers : Luis Otavio Alvares) - 1.0



- Grupo 292  
(Valiati, Joao F. : Alvares, Luis O.) - 1.0
- Grupo 293  
(Valiati, Joao F. : Vania Bogorny) - 1.0
- Grupo 294  
(Vania Bogorny : Hercules Matheus Avancini) - 1.0
- Grupo 295  
(Vania Bogorny : Bruno Cesar de Lima de Paula) - 1.0
- Grupo 296  
(Vania Bogorny : Cassiano Kuplich) - 1.0
- Grupo 297  
(Luis Otavio Alvares : Loy, Alisson Moscato) - 1.0
- Grupo 298  
(Luis Otavio Alvares : Renso, Chiara) - 1.0
- Grupo 299  
(Anna Monreale : Roberto Trasarti) - 1.0
- Grupo 300  
(Anna Monreale : Chiara Renso) - 1.0
- Grupo 301  
(Anna Monreale : Dino Pedreschi) - 1.0
- Grupo 302  
(Anna Monreale : Vania Bogorny) - 1.0
- Grupo 303  
(de Lucca Siqueira, Fernando : Vania Bogorny) - 1.0
- Grupo 304  
(Christine Parent : Stefano Spaccapietra) - 1.0
- Grupo 305  
(Christine Parent : Chiara Renso) - 1.0
- Grupo 306  
(Christine Parent : Gennady Adrienko) - 1.0
- Grupo 307  
(Christine Parent : Natalia Adrienko) - 1.0
- Grupo 308  
(Christine Parent : Vania Bogorny) - 1.0
- Grupo 309  
(Christine Parent : Maria Luisa Damiani) - 1.0
- Grupo 310  
(Christine Parent : Aris Gkoulalas-Divanis) - 1.0
- Grupo 311  
(Christine Parent : Jose Antonio Fernandes de Macedo) - 1.0
- Grupo 312

- (Christine Parent : Nikos Pelekis) - 1.0  
Grupo 313
- (Christine Parent : Yannis Theodoridis) - 1.0  
Grupo 314
- (Christine Parent : ZHIXIAN YAN) - 1.0  
Grupo 315
- (Andres Pineda : Francisco Moreno) - 1.0  
Grupo 316
- (Andres Pineda : Renato Fileto) - 1.0  
Grupo 317
- (Andres Pineda : Vania Bogorny) - 1.0  
Grupo 318
- (Chiara Renso : Artur Ribeiro de Aquino) - 1.0  
Grupo 319
- (Chiara Renso : Fernando de Lucca Siqueira) - 1.0  
Grupo 320
- (Chiara Renso : ALVARES, LUIS OTAVIO) - 1.0  
Grupo 321
- (Chiara Renso : Vania Bogorny) - 1.0  
Grupo 322
- (F. L. KOCH : Carlos Becker Westphall) - 1.0  
Grupo 323
- (A M Barotto : A Souza) - 1.0  
Grupo 324
- (A M Barotto : Carlos Becker Westphall) - 1.0  
Grupo 325
- (Luiz Fernando Kormann : Adriano COSER) - 1.0  
Grupo 326
- (Luiz Fernando Kormann : Carlos Becker Westphall) - 2.0  
Grupo 327
- (Luiz Fernando Kormann : M S M A Notare) - 1.0  
Grupo 328
- (Luiz Fernando Kormann : Bernardo Gonçalves Riso) - 1.0  
Grupo 329
- (Luiz Fernando Kormann : F. A. CRUZ) - 1.0  
Grupo 330
- (Brigitte Hitce : A Latreche) - 1.0  
Grupo 331
- (Brigitte Hitce : Carlos Becker Westphall) - 1.0  
Grupo 332
- (Abderrahim Sekkaki : Luis Marco Caceres Alvarez) - 2.0

- Grupo 333  
(Abderrahim Sekkaki : Wagner T Watanabe) - 2.0
- Grupo 334  
(Saliha Assoul : Carlos Becker Westphall) - 1.0
- Grupo 335  
(Marcos Dias de Assunção : FERNANDO LUIZ KOCH) - 1.0
- Grupo 336  
(Marcos Dias de Assunção : Carlos Becker Westphall) - 1.0
- Grupo 337  
(Désiré Nguessan : Carlos Becker Westphall) - 1.0
- Grupo 338  
(Morvan Daniel Muller : Carla Merkle Westphall) - 1.0
- Grupo 339  
(Morvan Daniel Muller : Carlos Becker Westphall) - 1.0
- Grupo 340  
(Alexandre Parra Carneiro da Silva : Carla Merkle Westphall) -  
1.0
- Grupo 341  
(Alexandre Parra Carneiro da Silva : Carlos Becker Westphall)  
- 1.0
- Grupo 342  
(F. L. KOCK : Marcos Dias de Assunção) - 1.0
- Grupo 343  
(F. L. KOCK : XAVIER, E.) - 1.0
- Grupo 344  
(F. L. KOCK : Carlos Becker Westphall) - 1.0
- Grupo 345  
(Aujor Tadeu Andrade : Carlos Becker Westphall) - 1.0
- Grupo 346  
(Luis Marco Caceres Alvarez : Carla Merkle Westphall) - 1.0
- Grupo 347  
(Luis Marco Caceres Alvarez : Carlos Becker Westphall) - 1.0
- Grupo 348  
(Alexandre Schulter : Kleber Magno Maciel Vieira) - 2.0
- Grupo 349  
(Alexandre Schulter : Carla Merkle Westphall) - 1.0
- Grupo 350  
(Alexandre Schulter : Carlos Becker Westphall) - 2.0
- Grupo 351  
(Nilo Acir Hirt : Micheli Jurczik Hirt) - 1.0
- Grupo 352

- (Nilo Acir Hirt : Carlos Becker Westphall) - 1.0  
Grupo 353
- (Alexandre Gava Menezes : Carlos Becker Westphall) - 1.0  
Grupo 354
- (Abderrahim Sekkaki : Mustapha El Hamzaoui) - 2.0  
Grupo 355
- (Abderrahim Sekkaki : Bahloul Bensassi) - 1.0  
Grupo 356
- (Abderrahim Sekkaki : Carla Merkle Westphall) - 1.0  
Grupo 357
- (Mustapha El Hamzaoui : Bahloul Bensassi) - 1.0  
Grupo 358
- (Mustapha El Hamzaoui : Carla Merkle Westphall) - 1.0  
Grupo 359
- (Mustapha El Hamzaoui : Carlos Becker Westphall) - 1.0  
Grupo 360
- (Brunner, Marcus : Nogueira, José Marcos) - 1.0  
Grupo 361
- (Brunner, Marcus : Ulema, Mehmet) - 1.0  
Grupo 362
- (Brunner, Marcus : Carlos Becker Westphall) - 1.0  
Grupo 363
- (NOGUEIRA, José Marcos : ULEMA, Mehmet) - 1.0  
Grupo 364
- (NOGUEIRA, José Marcos : Brunner, Marcus) - 1.0  
Grupo 365
- (NOGUEIRA, José Marcos : Carlos Becker Westphall) - 1.0  
Grupo 366
- (Kleber Magno Maciel Vieira : Carlos Becker Westphall) - 2.0  
Grupo 367
- (Kleber Magno Maciel Vieira : Carla Merkle Westphall) - 1.0  
Grupo 368
- (Arantes, Juliana Amaral : Carlos Becker Westphall) - 1.0  
Grupo 369
- (Arantes, Juliana Amaral : Custódio, Ricardo Felipe) - 1.0  
Grupo 370
- (Arantes, Juliana Amaral : Chaves, Shirlei Aparecida) - 1.0  
Grupo 371
- (Vieira, Kleber : Schulter, Alexandre) - 1.0  
Grupo 372
- (Vieira, Kleber : Carlos Becker Westphall) - 1.0

- Grupo 373  
(Vieira, Kleber : Westphall, Carla Merkle) - 1.0
- Grupo 374  
(Carlos Becker Westphall : Mueller, Peter) - 1.0
- Grupo 375  
(T. M. Bonetti : Carla Merkle Westphall) - 1.0
- Grupo 376  
(T. M. Bonetti : A. S. Córdova) - 1.0
- Grupo 377  
(T. M. Bonetti : Carlos Becker Westphall) - 1.0
- Grupo 378  
(Shirlei Aparecida de Chaves : Rafael Brundo Uriarte) - 1.0
- Grupo 379  
(Shirlei Aparecida de Chaves : Carla Merkle Westphall) - 1.0
- Grupo 380  
(Carlos Becker Westphall : Carla Merkle Westphall) - 2.0
- Grupo 381  
(Carlos Becker Westphall : FERNANDO LUIZ KOCH) - 2.0
- Grupo 382  
(Carlos Becker Westphall : Carlos Oberdan Rolim) - 1.0
- Grupo 383  
(Carlos Becker Westphall : Rafael de Souza Mendes) - 2.0
- Grupo 384  
(Carlos Becker Westphall : Rafael B. BRINHOSA;) - 1.0
- Grupo 385  
(Carlos Becker Westphall : Rafael Rodrigues de Freitas) - 2.0
- Grupo 386  
(Jorge Werner : Guilherme Arthur Geronimo) - 2.0
- Grupo 387  
(Jorge Werner : FERNANDO LUIZ KOCH) - 1.0
- Grupo 388  
(Jorge Werner : Rafael Rodrigues de Freitas) - 1.0
- Grupo 389  
(Jorge Werner : Carla Merkle Westphall) - 1.0
- Grupo 390  
(Ricardo Ferraro de Souza : Carlos Becker Westphall) - 2.0
- Grupo 391  
(Ricardo Ferraro de Souza : Daniel Ricardo dos Santos) - 1.0
- Grupo 392  
(Ricardo Ferraro de Souza : Carla Merkle Westphall) - 1.0
- Grupo 393

- (Daniel Ricardo dos Santos : Tiago Jaime Nascimento) - 1.0  
Grupo 394
- (Daniel Ricardo dos Santos : Carla Merkle Westphall) - 1.0  
Grupo 395
- (Daniel Ricardo dos Santos : Marcos Aurelio Pedroso  
Leandro) - 1.0  
Grupo 396
- (Daniel Ricardo dos Santos : Carlos Becker Westphall) - 2.0  
Grupo 397
- (Carlos Becker Westphall : Mauro Marcelo Mattos) - 1.0  
Grupo 398
- (Carlos Becker Westphall : Rafael Weingartner) - 2.0  
Grupo 399
- (Carlos Becker Westphall : Leonardo Defenti) - 1.0  
Grupo 400
- (Carlos Becker Westphall : Alexandre Augusto Flores) - 1.0  
Grupo 401
- (Carlos Becker Westphall : Gabriel Beim Bräscher) - 1.0  
Grupo 402
- ( Evandro Alencar Rigon : Carla Merkle Westphall) - 1.0  
Grupo 403
- ( Evandro Alencar Rigon : Daniel Ricardo dos Santos) - 1.0  
Grupo 404
- ( Evandro Alencar Rigon : Carlos Becker Westphall) - 1.0  
Grupo 405
- (Guilherme Arthur Geronimo : Rafael Weingartner) - 1.0  
Grupo 406
- (Guilherme Arthur Geronimo : Carla Merkle Westphall) - 1.0  
Grupo 407
- ( Rafael Weingartner : Gabriel Beim Bräscher) - 1.0  
Grupo 408
- (Sergio Roberto Villarreal : Mariana Elena Villarreal)- 1.0  
Grupo 409
- (Sergio Roberto Villarreal : Westphall, Carla Merkle) - 1.0  
Grupo 410
- (Luciano Paschoal Gaspary : Maria Janilce Almeida) - 1.0  
Grupo 411
- (Luciano Paschoal Gaspary : Roberto Willrich) - 1.0  
Grupo 412
- (Pierre de Saqui Sannes : Patrick Sénac) - 1.0  
Grupo 413

- (Pierre de Saqui Sannes : Michel Diaz) - 1.0  
 Grupo 414  
 (Pierre de Saqui Sannes : Roberto Willrich) - 1.0  
 Grupo 415  
 (Guillermo de Jesús Hoyos-Rivera : Roberta Lima-Gomes) - 2.0  
 Grupo 416  
 (Guillermo de Jesús Hoyos-Rivera : Jean Pierre Courtiat) - 1.0  
 Grupo 417  
 (Guillermo de Jesús Hoyos-Rivera : Roberto Willrich) - 2.0  
 Grupo 418  
 (Guillermo de Jesús Hoyos-Rivera : Giner Alor Hernandez) - 1.0  
 Grupo 419  
 (Christopher Viana Lima : Roberta Lima-Gomes) - 1.0  
 Grupo 420  
 (Christopher Viana Lima : Guillermo de Jesús Hoyos-Rivera) -  
 1.0  
 Grupo 421  
 (Christopher Viana Lima : Jean Pierre Courtiat) - 1.0  
 Grupo 422  
 (Christopher Viana Lima : Roberto Willrich) - 1.0  
 Grupo 423  
 (Madalena Pereira da Silva : Lia Caetano Bastos) - 1.0  
 Grupo 424  
 (Madalena Pereira da Silva : José Cé Júnior) - 1.0  
 Grupo 425  
 (Madalena Pereira da Silva : Roberto Willrich) - 1.0  
 Grupo 426  
 (Gomes, Roberta Lima : Guillermo de Jesús Hoyos-Rivera) - 1.0  
 Grupo 427  
 (Gomes, Roberta Lima : Roberto Willrich) - 1.0  
 Grupo 428  
 (Gomes, Roberta Lima : Christopher Viana Lima) - 1.0  
 Grupo 429  
 (Gomes, Roberta Lima : Jean Pierre Courtiat) - 1.0  
 Grupo 430  
 (CÉ JÚNIOR, JOSÉ : PRUDÊNCIO, ACHILLES C.) - 1.0  
 Grupo 431  
 (CÉ JÚNIOR, JOSÉ : Roberto Willrich) - 1.0  
 Grupo 432  
 (CÉ JÚNIOR, JOSÉ : SILVA, MADALENA P.) - 1.0  
 Grupo 433

(André Lins Gonzalez : Diego Izidoro) - 1.0

Grupo 434

(André Lins Gonzalez : Roberto Willrich) - 1.0

Grupo 435

(André Lins Gonzalez : Celso Alberto Saibel Santos) - 1.0

## A.2 *CLUSTERS* GERADOS NA EXECUÇÃO DO *BEST-STAR* COM O MÉTODO JACCARD - *THRESHOLD* 0.2

### Cluster 1

<Jean Carlo Rossa Hauck>

<Rogerio de Almeida Richa>

<Adriano Ferreti Borgatto>

<Mario Antonio Ribeiro Dantas>

<Roberto Willrich>

Áreas de conhecimento do cluster [Amostragem, Análise Bayesiana, Análise de Dados, Arquitetura de Sistemas de Computação, Automação Eletrônica de Processos Elétricos e Industriais, Avaliação da Aprendizagem, Banco de Dados, Computação Paralela e Distribuída, Engenharia de Software, Estatística Básica, Estatística Experimental, Hardware, Inferência Estatística, Linguagem Formais e Autômatos, Linguagens de Programação, Lógicas e Semântica de Programas, Mercadologia, Modelos Analíticos e de Simulação, Modelos Lineares Generalizados, Métodos e Técnicas de Ensino, Planejamento de Experimentos, Planejamento, Projeto e Controle de Sistemas de Produção, Processamento de Imagens, Sistemas Eletrônicos de Medida e de Controle, Sistemas de Informação, Software Básico, Tecnologia Educacional, Teleinformática, Teoria dos Jogos]

Especializações do cluster [Administração, Ciência da Computação, Educação, Engenharia Elétrica, Engenharia de Produção, Probabilidade e Estatística, Robótica, Mecatrônica e Automação]

### Cluster 2

<Carlos Becker Westphall>

<Lau Cheuk Lung>

<Renato Cislighi>

<Ricardo Azambuja Silveira>

<Paulo Jose de Freitas Filho>

Áreas de conhecimento do cluster [Acompanhamento e avaliação de disciplinas e cursos, Administração de Recursos Humanos, Administração



de Unidades Educativas, Administração e Gerência de Redes, Algoritmos Distribuídos, Análise de Custos, Análise de Risco em Gerência da Saúde, Aprendizagem Colaborativa, Arquitetura de Sistemas de Computação, Avaliação de Sistemas, Instituições, Planos e Programas Educacionais, Clusters, Computação Paralela e Distribuída, Custos da Construção de Sistemas Para a Perfuração e Completação de Poços de Petróleo, Custos de Construção Naval, Desempenho de Sistemas de Computação, Desempenho na Web, Engenharia de Software, Evasão discente, Gerência de Redes de Telecomunicações, Gerência Operacional, Gerência de Redes de Computadores, Gerência de Redes de Telecomunicações, Gerência de Serviços Computacionais, Grid, Linguagens de Programação, Lógicas e Semântica de Programas, Modelagem e Simulação, Modelos Analíticos e de Simulação, Modelos Probabilísticos, Organizações Públicas, Planejamento de Instalações Industriais, Planejamento, Projeto e Controle de Sistemas de Produção, Processos Estocásticos e Teoria das Filas, Redes Neurais, Redes de Computadores, Redes de Telecomunicações, Segurança Em Sistemas Distribuídos, Simulação Discreta, Simulação de Sistemas, Sistemas Cliente Servidor, Sistemas Distribuídos, Sistemas Especialistas, Tecnologia Educacional, Tecnologias de Perfuração e Completação de Poços de Petróleo, Teleinformática, Tolerância a Intrusões, Tratamento da Incerteza]

Especializações do cluster [Administração, Ciência da Computação, Educação, Engenharia Naval e Oceânica, Engenharia de Produção, Matemática, Probabilidade e Estatística]

### Cluster 3

<Aldo von Wangenheim>

<Christiane Anneliese Gresse von Wangenheim>

Áreas de conhecimento do cluster [Administração da Produção, Análise de Custos, Arquitetura de Sistemas de Computação, Banco de Dados, Cancerologia, Cardiologia, Cirurgia Cardiovascular, Cirurgia Otorrinolaringológica, Climatologia, Educação Permanente, Educação de Adultos, Engenharia de Software, Ensino e Aprendizagem na Sala de Aula, GERÊNCIA DE PROJETOS, Garantia de Controle de Qualidade, Gerência do Projeto e do Produto, Gestão de conhecimento, Instrumentação Meteorológica, Inteligencia Artificial, Interação Humano Computador, Linguagem Formais e Autômatos, Linguagens de Programação, MELHORIA DE PROCESSO DE SOFTWARE, Matemática Discreta e Combinatória, Meteorologia Sinótica, Modelos Analíticos e de Simulação, Métodos e Técnicas de Ensino, Neurologia, Neuropatologia,

Oftalmologia, Ortopedia, Otorrinolaringologia, Pediatria, Petrologia, Processamento Gráfico (Graphics), Processamento de Imagens, Processamento de Sinais Biológicos, Representação da Informação, Sensoriamento Remoto da Atmosfera, Sistemas de Informação, Sistemas de Telecomunicações, Software Básico, Tecnologia Educacional, Teleinformática, Telerradiologia, Videodifusão, prototipagem rápida, Ótica]

Especializações do cluster [Administração, Ciência da Computação, Ciência da Informação, Comunicação, Educação, Engenharia Biomédica, Engenharia Elétrica, Engenharia Mecânica, Engenharia de Produção, Física, Geociências, Matemática, Medicina, Psicologia]

#### Cluster 4

<Andrea Cristina Konrath>

<Antonio Carlos Mariani>

Áreas de conhecimento do cluster [Software Básico] Especializações do cluster [Ciência da Computação]

#### Cluster 5

<Luciana de Oliveira Rech>

<Ricardo Felipe Custodio>

<Frank Augusto Siqueira>

Áreas de conhecimento do cluster [Análise Numérica, Análise de Algoritmos e Complexidade de Computação, Automação Eletrônica de Processos Elétricos e Industriais, Computabilidade e Modelos de Computação, Criptografia, GOVERNO ELETRÔNICO, Hardware, Instrumentação Odontológica e Médico-Hospitalar, Matemática Aplicada, Pneumologia, Processamento de Sinais Biológicos, Segurança, Segurança em Informática, Sistemas de Informação, Sistemas de Telecomunicações, Software Básico, Teleinformática, Transmissão da Energia Elétrica, Distribuição da Energia Elétrica]

Especializações do cluster [Administração, Ciência da Computação, Engenharia Biomédica, Engenharia Elétrica, Física, Matemática, Medicina]

#### Cluster 6

<Silvia Modesto Nassar>

<Carla Merkle Westphall>

<João Bosco Mangueira Sobral>

<Mauro Roisenberg>

Áreas de conhecimento do cluster [Agentes Móveis, Análise de Dados, Aplicativos Web, Arquitetura de Sistemas de Computação, Automação Eletrônica de Processos Elétricos e Industriais, Banco de Dados, Computação Com Objetos Distribuídos, Comércio Eletrônico, Frameworks, Geofísica Aplicada, Gerência de Redes de Computadores, Gerência de Redes de Telecomunicações, Gerência de Serviços, Hardware, Informática Médica, Inteligência Artificial, J2me, Linguagem Formais e Autômatos, Linguagens de Programação, Medidas Elétricas, Pediatria, Processos Estocásticos e Teoria das Filas, Programação Para Web, Redes de Computadores, Referência Ambígua, Reumatologia, Segurança de Sistemas de Computação, Sistemas de Informação, Software Aplicativo, Software Básico, Telecomunicações, Teleinformática, Web Services]

Especializações do cluster [Ciência da Computação, Engenharia Elétrica, Engenharia de Produção, Geociências, Matemática, Medicina, Probabilidade e Estatística]

#### Cluster 7

<Olinto Jose Varela Furtado>

<José Luís Almada Güntzel>

<Ricardo Pereira e Silva>

Áreas de conhecimento do cluster [Adl Archc, Análise de Algoritmos e Complexidade de Computação, Arquitetura de Sistemas de Computação, Automação Eletrônica de Processos Elétricos e Industriais, Banco de Dados, Circuitos Eletrônicos, Compiladores e Linguagens de Programação, Engenharia de Software, Frameworks orientados a objetos, Geração Automática de Montadores, Gerência de Telecomunicações, Hardware, Linguagem Formais e Autômatos, Linguagens de Programação, Microeletrônica, Plano de Continuidade de Negócios, Processamento Gráfico (Graphics), Redes Neurais, Robótica, Sistemas Distribuídos e Tempo Real, Sistemas Embarcados, Sistemas Especialistas, Sistemas Operacionais, Sistemas de Informação, Software Básico, Teleinformática, desenvolvimento baseado em componentes, metodologias de desenvolvimento de software, teste de software]

Especializações do cluster [Ciência da Computação, Engenharia Elétrica]

#### Cluster 8

<Renato Fileto>

<Carina Friedrich Dorneles>

<Ronaldo dos Santos Mello>

<Vania Bogorny>

Áreas de conhecimento do cluster [Análise de Dados, Aprendizado de Máquina, Arquitetura de Sistemas de Computação, Banco de Dados, Bancos de Dados Geográficos, Bioinformática Estrutural, Consulta Por Similaridade, Dados Semi Estruturados e Web, Data Mining, Documentos, Engenharia de Software, Extração de Informações, Information Retrieval, Integração Semântica, Interfaces, Linguagens de Programação, Lógicas e Semântica de Programas, Mineração de Dados, Métodos e Modelos Matemáticos, Econométricos e Estatísticos, Ontologias, Regras de Associação, Sistemas de Informação, Spatial Data Mining, Xml]

Especializações do cluster [Ciência da Computação, Economia, Probabilidade e Estatística]

#### Cluster 9

<Raul Sidnei Wazlawick>

Áreas de conhecimento do cluster [Administração de Recursos Humanos, Banco de Dados, Educação a Distância, Engenharia de Software, Inteligência Artificial, Linguagens de Programação, Organizações Públicas, Realidade Virtual, Sistemas de Informação, Software Básico, Transmissão da Energia Elétrica, Distribuição da Energia Elétrica]

Especializações do cluster [Administração, Ciência da Computação, Educação, Engenharia Elétrica]

#### Cluster 10

<Patrícia Vilain>

<Vitório Bruno Mazzola>

Áreas de conhecimento do cluster [Banco de Dados, Controle de Processos Eletrônicos, Retroalimentação, Engenharia de Software, Linguagem Formais e Autômatos, Lógicas e Semântica de Programas, Redes de Computadores, Software Básico, Transmissão da Energia Elétrica, Distribuição da Energia Elétrica]

Especializações do cluster [Ciência da Computação, Engenharia Elétrica]

#### Cluster 11

<Jose Eduardo De Lucca>

<Luiz Cláudio Villar dos Santos>

Áreas de conhecimento do cluster [Arquitetura de Sistemas de Computação, Banco de Dados, Direito Comercial, Educação Especial, Empreendedorismo, Engenharia de Software, Hardware, Inovação, Linguagem Formais e Autômatos, Linguagens de Programação, Localização, Lógicas e Semântica de Programas, Modelagem de Fenômenos Biológicos, Modelagem de Sistemas Biológicos, Mudança Tecnológica, Métodos e Técnicas de Ensino, Negócios Internacionais, Organização Industrial e Estudos Industriais, Planejamento em Ciência e Tecnologia, Política e Planejamento Governamentais, Processamento Gráfico, Processamento Gráfico (Graphics), Sistemas de Informação, Software Básico, Tecnologia Educacional, Teleinformática, Teoria e Política de Planejamento Econômico, Tradução]

Especializações do cluster [Administração, Ciência da Computação, Direito, Economia, Educação, Engenharia Biomédica, Letras]

#### Cluster 12

<Antonio Augusto Medeiros Fröhlich>

Áreas de conhecimento do cluster [Arquitetura de Sistemas de Computação, Computação Paralela, Engenharia de Software, Hardware, Software Básico, Teleinformática]

Especializações do cluster [Ciência da Computação]

#### Cluster 13

<Rafael Luiz Cancian>

<Jean Everson Martina>

Áreas de conhecimento do cluster [Arquitetura de Sistemas de Computação, Astronomia Espacial, Astronomia Fundamental, Automação Eletrônica de Processos Elétricos e Industriais, Avaliação de Sistemas, Instituições, Planos e Programas Educacionais, Banco de Dados, Bioinformática, Cerimônias Criptográficas, Circuitos Eletrônicos, Clusters de Processamento Distribuído, Computabilidade e Modelos de Computação, Controle de Processos Eletrônicos, Retroalimentação, Criptoanálise, Dinâmica dos Fluidos, Economia Matemática, Emissão de Documentos Digitais, Engenharia de Software, Geomorfologia Submarina, Gerenciamento de Certificados Digitais, Hardware, Inferência Paramétrica, Inteligência Artificial, Interação do Oceano com a Atmosfera, Interação entre os Organismos Marinhos e os Parâmetros Ambientais, Linguagens de Programação, Lógicas e Semântica de Programas, Manejo e Conservação de Recursos Pesqueiros Marinhos, Maricultura, Matemática Simbólica, Materiais e Componentes Semicondutores, Me-

canismos de Assinatura Digital, Modelagem de Sistemas Biológicos, Modelos Analíticos e de Simulação, Métodos Formais para Segurança, Métodos e Modelos Matemáticos, Econométricos e Estatísticos, Métodos e Técnicas de Ensino, Operação de Sistemas de Transporte, Planejamento de Experimentos, Planejamento e Organização do Sistema de Transporte, Processos de Colisão e Interações de Átomos e Moléculas, Proteínas, Protocolos Criptográficos, Práticas de Segurança, Redes Privadas Virtuais, Redes Troncalizadas, Redes de Computadores, SISTEMAS OPERACIONAIS, Sedimentologia Marinha, Segurança de Redes, Sistema Planetário, Sistema de Tempo Real, Sistemas Dinâmicos, Sistemas Distribuídos, Sistemas Eletrônicos de Medida e de Controle, Sistemas de Informação, Sistemas de Tempo Real, Software Básico, Séries Temporais, Teleinformática, Tempo real, Teoria Geral de Planejamento e Desenvolvimento Curricular, Transporte e Tráfego Urbano e Regional]

Especializações do cluster [Astronomia, Bioquímica, Ciência da Computação, Economia, Educação, Engenharia Biomédica, Engenharia Elétrica, Engenharia de Produção, Engenharia de Transportes, Física, Matemática, Oceanografia, Planejamento Urbano e Regional, Probabilidade e Estatística, Recursos Pesqueiros e Engenharia de Pesca]

Clusters não unitários: 11

### A.3 CLUSTERS GERADOS NA EXECUÇÃO DO *BEST-STAR* PELO MÉTODO COSINE - *THRESHOLD* 0.3

#### Cluster 1

<Jean Carlo Rossa Hauck>  
 <Rogerio de Almeida Richa>  
 <Adriano Ferreti Borgatto>  
 <Mario Antonio Ribeiro Dantas>  
 <Roberto Willrich>

Áreas de conhecimento do cluster [Amostragem, Análise Bayesiana, Análise de Dados, Arquitetura de Sistemas de Computação, Automação Eletrônica de Processos Elétricos e Industriais, Avaliação da Aprendizagem, Banco de Dados, Computação Paralela e Distribuída, Engenharia de Software, Estatística Básica, Estatística Experimental, Hardware, Inferência Estatística, Linguagem Formais e Autômatos, Linguagens

de Programação, Lógicas e Semântica de Programas, Mercadologia, Modelos Analíticos e de Simulação, Modelos Lineares Generalizados, Métodos e Técnicas de Ensino, Planejamento de Experimentos, Planejamento, Projeto e Controle de Sistemas de Produção, Processamento de Imagens, Sistemas Eletrônicos de Medida e de Controle, Sistemas de Informação, Software Básico, Tecnologia Educacional, Teleinformática, Teoria dos Jogos]

Especializações do cluster [Administração, Ciência da Computação, Educação, Engenharia Elétrica, Engenharia de Produção, Probabilidade e Estatística, Robótica, Mecatrônica e Automação]

#### Cluster 2

<Carlos Becker Westphall>

<Lau Cheuk Lung>

<Renato Cislighi>

<Ricardo Azambuja Silveira>

<Paulo Jose de Freitas Filho>

Áreas de conhecimento do cluster [Acompanhamento e avaliação de disciplinas e cursos, Administração de Recursos Humanos, Administração de Unidades Educativas, Administração e Gerência de Redes, Algoritmos Distribuídos, Análise de Custos, Análise de Risco em Gerência da Saúde, Aprendizagem Colaborativa, Arquitetura de Sistemas de Computação, Avaliação de Sistemas, Instituições, Planos e Programas Educacionais, Clusters, Computação Paralela e Distribuída, Custos da Construção de Sistemas Para a Perfuração e Completação de Poços de Petróleo, Custos de Construção Naval, Desempenho de Sistemas de Computação, Desempenho na Web, Engenharia de Software, Evasão discente, Gerencia de Redes de Telecomunicações, Gerência Operacional, Gerência de Redes de Computadores, Gerência de Redes de Telecomunicações, Gerência de Serviços Computacionais, Grid, Linguagens de Programação, Lógicas e Semântica de Programas, Modelagem e Simulação, Modelos Analíticos e de Simulação, Modelos Probabilísticos, Organizações Públicas, Planejamento de Instalações Industriais, Planejamento, Projeto e Controle de Sistemas de Produção, Processos Estocásticos e Teoria das Filas, Redes Neurais, Redes de Computadores, Redes de Telecomunicações, Segurança Em Sistemas Distribuídos, Simulação Discreta, Simulação de Sistemas, Sistemas Cliente Servidor, Sistemas Distribuídos, Sistemas Especialistas, Tecnologia Educacional, Tecnologias de Perfuração e Completação de Poços de Petróleo, Teleinformática, Tolerância a Intrusões, Tratamento da Incerteza]

Especializações do cluster [Administração, Ciência da Computação, Educação, Engenharia Naval e Oceânica, Engenharia de Produção, Matemática, Probabilidade e Estatística]

### Cluster 3

<Aldo von Wangenheim>

<Patrícia Vilain>

<Christiane Anneliese Gresse von Wangenheim>

<Vitório Bruno Mazzola>

Áreas de conhecimento do cluster [Administração da Produção, Análise de Custos, Arquitetura de Sistemas de Computação, Banco de Dados, Cancerologia, Cardiologia, Cirurgia Cardiovascular, Cirurgia Otorrinolaringológica, Climatologia, Controle de Processos Eletrônicos, Retroalimentação, Educação Permanente, Educação de Adultos, Engenharia de Software, Ensino e Aprendizagem na Sala de Aula, GERÊNCIA DE PROJETOS, Garantia de Controle de Qualidade, Gerência do Projeto e do Produto, Gestão de conhecimento, Instrumentação Meteorológica, Inteligencia Artificial, Interação Humano Computador, Linguagem Formais e Autômatos, Linguagens de Programação, Lógicas e Semântica de Programas, MELHORIA DE PROCESSO DE SOFTWARE, Matemática Discreta e Combinatória, Meteorologia Sinótica, Modelos Analíticos e de Simulação, Métodos e Técnicas de Ensino, Neurologia, Neuropatologia, Oftalmologia, Ortopedia, Otorrinolaringologia, Pediatria, Petrologia, Processamento Gráfico (Graphics), Processamento de Imagens, Processamento de Sinais Biológicos, Redes de Computadores, Representação da Informação, Sensoriamento Remoto da Atmosfera, Sistemas de Informação, Sistemas de Telecomunicações, Software Básico, Tecnologia Educacional, Teleinformática, Telerradiologia, Transmissão da Energia Elétrica, Distribuição da Energia Elétrica, Videodifusão, prototipagem rápida, Ótica]

Especializações do cluster [Administração, Ciência da Computação, Ciência da Informação, Comunicação, Educação, Engenharia Biomédica, Engenharia Elétrica, Engenharia Mecânica, Engenharia de Produção, Física, Geociências, Matemática, Medicina, Psicologia]

### Cluster 4

<Andrea Cristina Konrath>

<Antonio Carlos Mariani>

Áreas de conhecimento do cluster [Software Básico]



## Especializações do cluster [Ciência da Computação]

### Cluster 5

<Luciana de Oliveira Rech>

<Ricardo Felipe Custodio>

<Carla Merkle Westphall>

<Frank Augusto Siqueira>

Áreas de conhecimento do cluster [Análise Numérica, Análise de Algoritmos e Complexidade de Computação, Automação Eletrônica de Processos Elétricos e Industriais, Computabilidade e Modelos de Computação, Criptografia, GOVERNO ELETRÔNICO, Hardware, Instrumentação Odontológica e Médico-Hospitalar, Matemática Aplicada, Pneumologia, Processamento de Sinais Biológicos, Segurança, Segurança em Informática, Sistemas de Informação, Sistemas de Telecomunicações, Software Básico, Telecomunicações, Teleinformática, Transmissão da Energia Elétrica, Distribuição da Energia Elétrica]

Especializações do cluster [Administração, Ciência da Computação, Engenharia Biomédica, Engenharia Elétrica, Física, Matemática, Medicina]

### Cluster 6

<Silvia Modesto Nassar>

<João Bosco Mangueira Sobral>

<Mauro Roisenberg>

Áreas de conhecimento do cluster [Agentes Móveis, Análise de Dados, Aplicativos Web, Arquitetura de Sistemas de Computação, Automação Eletrônica de Processos Elétricos e Industriais, Banco de Dados, Computação Com Objetos Distribuídos, Comércio Eletrônico, Frameworks, Geofísica Aplicada, Gerência de Redes de Computadores, Gerência de Redes de Telecomunicações, Gerência de Serviços, Hardware, Informática Médica, Inteligência Artificial, J2me, Linguagem Formais e Autômatos, Linguagens de Programação, Medidas Elétricas, Pediatria, Processos Estocásticos e Teoria das Filas, Programação Para Web, Redes de Computadores, Referência Ambígua, Reumatologia, Segurança de Sistemas de Computação, Sistemas de Informação, Software Aplicativo, Software Básico, Teleinformática, Web Services]

Especializações do cluster [Ciência da Computação, Engenharia Elétrica, Engenharia de Produção, Geociências, Matemática, Medicina, Probabilidade e Estatística]

## Cluster 7

&lt;Olinto Jose Varela Furtado&gt;

&lt;José Luís Almada Güntzel&gt;

&lt;Ricardo Pereira e Silva&gt;

Áreas de conhecimento do cluster [Adl Archc, Análise de Algoritmos e Complexidade de Computação, Arquitetura de Sistemas de Computação, Automação Eletrônica de Processos Elétricos e Industriais, Banco de Dados, Circuitos Eletrônicos, Compiladores e Linguagens de Programação, Engenharia de Software, Frameworks orientados a objetos, Geração Automática de Montadores, Gerência de Telecomunicações, Hardware, Linguagem Formais e Autômatos, Linguagens de Programação, Microeletrônica, Plano de Continuidade de Negócios, Processamento Gráfico (Graphics), Redes Neurais, Robótica, Sistemas Distribuídos e Tempo Real, Sistemas Embarcados, Sistemas Especialistas, Sistemas Operacionais, Sistemas de Informação, Software Básico, Teleinformática, desenvolvimento baseado em componentes, metodologias de desenvolvimento de software, teste de software]

Especializações do cluster [Ciência da Computação, Engenharia Elétrica]

## Cluster 8

&lt;Renato Fileto&gt;

&lt;Carina Friedrich Dorneles&gt;

&lt;Ronaldo dos Santos Mello&gt;

&lt;Vania Bogorny&gt;

Áreas de conhecimento do cluster [Análise de Dados, Aprendizado de Máquina, Arquitetura de Sistemas de Computação, Banco de Dados, Bancos de Dados Geográficos, Bioinformática Estrutural, Consulta Por Similaridade, Dados Semi Estruturados e Web, Data Mining, Documentos, Engenharia de Software, Extração de Informações, Information Retrieval, Integração Semântica, Interfaces, Linguagens de Programação, Lógicas e Semântica de Programas, Mineração de Dados, Métodos e Modelos Matemáticos, Econométricos e Estatísticos, Ontologias, Regras de Associação, Sistemas de Informação, Spatial Data Mining, Xml]

Especializações do cluster [Ciência da Computação, Economia, Probabilidade e Estatística]

## Cluster 9

<Raul Sidnei Wazlawick>

Áreas de conhecimento do cluster [Administração de Recursos Humanos, Banco de Dados, Educação a Distância, Engenharia de Software, Inteligência Artificial, Linguagens de Programação, Organizações Públicas, Realidade Virtual, Sistemas de Informação, Software Básico, Transmissão da Energia Elétrica, Distribuição da Energia Elétrica]

Especializações do cluster [Administração, Ciência da Computação, Educação, Engenharia Elétrica]

#### Cluster 10

<Jose Eduardo De Lucca>

<Luiz Cláudio Villar dos Santos>

Áreas de conhecimento do cluster [Arquitetura de Sistemas de Computação, Banco de Dados, Direito Comercial, Educação Especial, Empreendedorismo, Engenharia de Software, Hardware, Inovação, Linguagem Formais e Autômatos, Linguagens de Programação, Localização, Lógicas e Semântica de Programas, Modelagem de Fenômenos Biológicos, Modelagem de Sistemas Biológicos, Mudança Tecnológica, Métodos e Técnicas de Ensino, Negócios Internacionais, Organização Industrial e Estudos Industriais, Planejamento em Ciência e Tecnologia, Política e Planejamento Governamentais, Processamento Gráfico, Processamento Gráfico (Graphics), Sistemas de Informação, Software Básico, Tecnologia Educacional, Teleinformática, Teoria e Política de Planejamento Econômico, Tradução]

Especializações do cluster [Administração, Ciência da Computação, Direito, Economia, Educação, Engenharia Biomédica, Letras]

#### Cluster 11

<Antonio Augusto Medeiros Fröhlich>

Áreas de conhecimento do cluster [Arquitetura de Sistemas de Computação, Computação Paralela, Engenharia de Software, Hardware, Software Básico, Teleinformática]

Especializações do cluster [Ciência da Computação]

#### Cluster 12

<Rafael Luiz Cancian>

<Jean Everson Martina>

Áreas de conhecimento do cluster [Arquitetura de Sistemas de Computação, Astronomia Espacial, Astronomia Fundamental, Automação Eletrônica de Processos Elétricos e Industriais, Avaliação de Sistemas, Instituições, Planos e Programas Educacionais, Banco de Dados, Bioin-

formática, Cerimônias Criptográficas, Circuitos Eletrônicos, Clusters de Processamento Distribuído, Computabilidade e Modelos de Computação, Controle de Processos Eletrônicos, Retroalimentação, Criptoanálise, Dinâmica dos Fluidos, Economia Matemática, Emissão de Documentos Digitais, Engenharia de Software, Geomorfologia Submarina, Gerenciamento de Certificados Digitais, Hardware, Inferência Paramétrica, Inteligência Artificial, Interação do Oceano com a Atmosfera, Interação entre os Organismos Marinhos e os Parâmetros Ambientais, Linguagens de Programação, Lógicas e Semântica de Programas, Manejo e Conservação de Recursos Pesqueiros Marinhos, Maricultura, Matemática Simbólica, Materiais e Componentes Semicondutores, Mecanismos de Assinatura Digital, Modelagem de Sistemas Biológicos, Modelos Analíticos e de Simulação, Métodos Formais para Segurança, Métodos e Modelos Matemáticos, Econométricos e Estatísticos, Métodos e Técnicas de Ensino, Operação de Sistemas de Transporte, Planejamento de Experimentos, Planejamento e Organização do Sistema de Transporte, Processos de Colisão e Interações de Átomos e Moléculas, Proteínas, Protocolos Criptográficos, Práticas de Segurança, Redes Privadas Virtuais, Redes Troncalizadas, Redes de Computadores, SISTEMAS OPERACIONAIS, Sedimentologia Marinha, Segurança de Redes, Sistema Planetário, Sistema de Tempo Real, Sistemas Dinâmicos, Sistemas Distribuídos, Sistemas Eletrônicos de Medida e de Controle, Sistemas de Informação, Sistemas de Tempo Real, Software Básico, Séries Temporais, Teleinformática, Tempo real, Teoria Geral de Planejamento e Desenvolvimento Curricular, Transporte e Tráfego Urbano e Regional]

Especializações do cluster [Astronomia, Bioquímica, Ciência da Computação, Economia, Educação, Engenharia Biomédica, Engenharia Elétrica, Engenharia de Produção, Engenharia de Transportes, Física, Matemática, Oceanografia, Planejamento Urbano e Regional, Probabilidade e Estatística, Recursos Pesqueiros e Engenharia de Pesca]

Clusters não unitários: 10

#### A.4 CLUSTERS GERADOS NA EXECUÇÃO DO *BEST-STAR* COM O MÉTODO OVERLAP - *THRESHOLD* 0.5

Cluster 1

<Jean Carlo Rossa Hauck>

<Rogerio de Almeida Richa>

<Adriano Ferreti Borgatto>

<Mario Antonio Ribeiro Dantas>

<Christiane Anneliese Gresse von Wangenheim>

<Roberto Willrich>

Áreas de conhecimento do cluster [Administração da Produção, Amostragem, Análise Bayesiana, Análise de Custos, Análise de Dados, Arquitetura de Sistemas de Computação, Automação Eletrônica de Processos Elétricos e Industriais, Avaliação da Aprendizagem, Banco de Dados, Computação Paralela e Distribuída, Educação de Adultos, Engenharia de Software, Ensino e Aprendizagem na Sala de Aula, Estatística Básica, Estatística Experimental, GERÊNCIA DE PROJETOS, Gestão de conhecimento, Hardware, Inferência Estatística, Inteligencia Artificial, Interação Humano Computador, Linguagem Formais e Autômatos, Linguagens de Programação, Lógicas e Semântica de Programas, MELHORIA DE PROCESSO DE SOFTWARE, Mercadologia, Modelos Analíticos e de Simulação, Modelos Lineares Generalizados, Métodos e Técnicas de Ensino, Planejamento de Experimentos, Planejamento, Projeto e Controle de Sistemas de Produção, Processamento de Imagens, Sistemas Eletrônicos de Medida e de Controle, Sistemas de Informação, Software Básico, Tecnologia Educacional, Teleinformática, Teoria dos Jogos]

Especializações do cluster [Administração, Ciência da Computação, Educação, Engenharia Elétrica, Engenharia de Produção, Probabilidade e Estatística, Psicologia, Robótica, Mecatrônica e Automação]

## Cluster 2

<Carlos Becker Westphall>

<Lau Cheuk Lung>

<Renato Cislaghi>

<Ricardo Azambuja Silveira>

<Paulo Jose de Freitas Filho>

Áreas de conhecimento do cluster [Acompanhamento e avaliação de disciplinas e cursos, Administração de Recursos Humanos, Administração de Unidades Educativas, Administração e Gerência de Redes, Algoritmos Distribuídos, Análise de Custos, Análise de Risco em Gerência da Saúde, Aprendizagem Colaborativa, Arquitetura de Sistemas de Computação, Avaliação de Sistemas, Instituições, Planos e Programas Educacionais, Clusters, Computação Paralela e Distribuída, Custos da Construção de Sistemas Para a Perfuração e Completação de Poços]

de Petróleo, Custos de Construção Naval, Desempenho de Sistemas de Computação, Desempenho na Web, Engenharia de Software, Evasão discente, Gerencia de Redes de Telecomunicações, Gerência Operacional, Gerência de Redes de Computadores, Gerência de Redes de Telecomunicações, Gerência de Serviços Computacionais, Grid, Linguagens de Programação, Lógicas e Semântica de Programas, Modelagem e Simulação, Modelos Analíticos e de Simulação, Modelos Probabilísticos, Organizações Públicas, Planejamento de Instalações Industriais, Planejamento, Projeto e Controle de Sistemas de Produção, Processos Estocásticos e Teoria das Filas, Redes Neurais, Redes de Computadores, Redes de Telecomunicações, Segurança Em Sistemas Distribuídos, Simulação Discreta, Simulação de Sistemas, Sistemas Cliente Servidor, Sistemas Distribuídos, Sistemas Especialistas, Tecnologia Educacional, Tecnologias de Perfuração e Completação de Poços de Petróleo, Teleinformática, Tolerância a Intrusões, Tratamento da Incerteza]

Especializações do cluster [Administração, Ciência da Computação, Educação, Engenharia Naval e Oceânica, Engenharia de Produção, Matemática, Probabilidade e Estatística]

#### Cluster 3

<Andrea Cristina Konrath>

<Antonio Carlos Mariani>

Áreas de conhecimento do cluster [Software Básico]

Especializações do cluster [Ciência da Computação]

#### Cluster 4

<Luciana de Oliveira Rech>

<Ricardo Felipe Custodio>

<Carla Merkle Westphall>

<Frank Augusto Siqueira>

Áreas de conhecimento do cluster [Análise Numérica, Análise de Algoritmos e Complexidade de Computação, Automação Eletrônica de Processos Elétricos e Industriais, Computabilidade e Modelos de Computação, Criptografia, GOVERNO ELETRÔNICO, Hardware, Instrumentação Odontológica e Médico-Hospitalar, Matemática Aplicada, Pneumologia, Processamento de Sinais Biológicos, Segurança, Segurança em Informática, Sistemas de Informação, Sistemas de Telecomunicações, Software Básico, Telecomunicações, Teleinformática, Transmissão da Energia Elétrica, Distribuição da Energia Elétrica]

Especializações do cluster [Administração, Ciência da Computação, Engenharia Biomédica, Engenharia Elétrica, Física, Matemática, Medicina]

#### Cluster 5

<Silvia Modesto Nassar>

<João Bosco Mangueira Sobral>

<Mauro Roisenberg>

Áreas de conhecimento do cluster [Agentes Móveis, Análise de Dados, Aplicativos Web, Arquitetura de Sistemas de Computação, Automação Eletrônica de Processos Elétricos e Industriais, Banco de Dados, Computação Com Objetos Distribuídos, Comércio Eletrônico, Frameworks, Geofísica Aplicada, Gerência de Redes de Computadores, Gerência de Redes de Telecomunicações, Gerência de Serviços, Hardware, Informática Médica, Inteligência Artificial, J2me, Linguagem Formais e Autômatos, Linguagens de Programação, Medidas Elétricas, Pediatria, Processos Estocásticos e Teoria das Filas, Programação Para Web, Redes de Computadores, Referência Ambígua, Reumatologia, Segurança de Sistemas de Computação, Sistemas de Informação, Software Aplicativo, Software Básico, Teleinformática, Web Services]

Especializações do cluster [Ciência da Computação, Engenharia Elétrica, Engenharia de Produção, Geociências, Matemática, Medicina, Probabilidade e Estatística]

#### Cluster 6

<Olinto Jose Varela Furtado>

<José Luís Almada Güntzel>

<Ricardo Pereira e Silva>

Áreas de conhecimento do cluster [Adl Archc, Análise de Algoritmos e Complexidade de Computação, Arquitetura de Sistemas de Computação, Automação Eletrônica de Processos Elétricos e Industriais, Banco de Dados, Circuitos Eletrônicos, Compiladores e Linguagens de Programação, Engenharia de Software, Frameworks orientados a objetos, Geração Automática de Montadores, Gerência de Telecomunicações, Hardware, Linguagem Formais e Autômatos, Linguagens de Programação, Microeletrônica, Plano de Continuidade de Negócios, Processamento Gráfico (Graphics), Redes Neurais, Robótica, Sistemas Distribuídos e Tempo Real, Sistemas Embarcados, Sistemas Especialistas, Sistemas Operacionais, Sistemas de Informação, Software Básico, Teleinformática, desenvolvimento baseado em componentes, metodologias

de desenvolvimento de software, teste de software]

Especializações do cluster [Ciência da Computação, Engenharia Elétrica]

#### Cluster 7

<Raul Sidnei Wazlawick>

Áreas de conhecimento do cluster [Administração de Recursos Humanos, Banco de Dados, Educação a Distância, Engenharia de Software, Inteligência Artificial, Linguagens de Programação, Organizações Públicas, Realidade Virtual, Sistemas de Informação, Software Básico, Transmissão da Energia Elétrica, Distribuição da Energia Elétrica]

Especializações do cluster [Administração, Ciência da Computação, Educação, Engenharia Elétrica]

#### Cluster 8

<Carina Friedrich Dorneles>

<Ronaldo dos Santos Mello>

<Vania Bogorny>

Áreas de conhecimento do cluster [Aprendizado de Máquina, Arquitetura de Sistemas de Computação, Banco de Dados, Bancos de Dados Geográficos, Consulta Por Similaridade, Dados Semi Estruturados e Web, Data Mining, Documentos, Engenharia de Software, Extração de Informações, Information Retrieval, Integração Semântica, Interfaces, Linguagens de Programação, Mineração de Dados, Métodos e Modelos Matemáticos, Econométricos e Estatísticos, Ontologias, Regras de Associação, Sistemas de Informação, Spatial Data Mining, Xml]

Especializações do cluster [Ciência da Computação, Economia]

#### Cluster 9

<Patrícia Vilain>

<Aldo von Wangenheim>

<Renato Fileto>

<Vitório Bruno Mazzola>

Áreas de conhecimento do cluster [Análise de Dados, Arquitetura de Sistemas de Computação, Banco de Dados, Bioinformática Estrutural, Cancerologia, Cardiologia, Cirurgia Cardiovascular, Cirurgia Otorrinolaringológica, Climatologia, Controle de Processos Eletrônicos, Retroalimentação, Educação Permanente, Educação de Adultos, Engenharia



de Software, Garantia de Controle de Qualidade, Gerência do Projeto e do Produto, Instrumentação Meteorológica, Linguagem Formais e Autômatos, Linguagens de Programação, Lógicas e Semântica de Programas, Matemática Discreta e Combinatória, Meteorologia Sinótica, Modelos Analíticos e de Simulação, Métodos e Técnicas de Ensino, Neurologia, Neuropatologia, Oftalmologia, Ortopedia, Otorrinolaringologia, Pediatria, Petrologia, Processamento Gráfico (Graphics), Processamento de Imagens, Processamento de Sinais Biológicos, Redes de Computadores, Representação da Informação, Sensoriamento Remoto da Atmosfera, Sistemas de Informação, Sistemas de Telecomunicações, Software Básico, Tecnologia Educacional, Teleinformática, Telerradiologia, Transmissão da Energia Elétrica, Distribuição da Energia Elétrica, Videodifusão, prototipagem rápida, Ótica]

Especializações do cluster [Ciência da Computação, Ciência da Informação, Comunicação, Educação, Engenharia Biomédica, Engenharia Elétrica, Engenharia Mecânica, Engenharia de Produção, Física, Geociências, Matemática, Medicina, Probabilidade e Estatística]

#### Cluster 10

<Jose Eduardo De Lucca>

<Luiz Cláudio Villar dos Santos>

Áreas de conhecimento do cluster [Arquitetura de Sistemas de Computação, Banco de Dados, Direito Comercial, Educação Especial, Empreendedorismo, Engenharia de Software, Hardware, Inovação, Linguagem Formais e Autômatos, Linguagens de Programação, Localização, Lógicas e Semântica de Programas, Modelagem de Fenômenos Biológicos, Modelagem de Sistemas Biológicos, Mudança Tecnológica, Métodos e Técnicas de Ensino, Negócios Internacionais, Organização Industrial e Estudos Industriais, Planejamento em Ciência e Tecnologia, Política e Planejamento Governamentais, Processamento Gráfico, Processamento Gráfico (Graphics), Sistemas de Informação, Software Básico, Tecnologia Educacional, Teleinformática, Teoria e Política de Planejamento Econômico, Tradução]

Especializações do cluster [Administração, Ciência da Computação, Direito, Economia, Educação, Engenharia Biomédica, Letras]

#### Cluster 11

<Antonio Augusto Medeiros Fröhlich>

Áreas de conhecimento do cluster [Arquitetura de Sistemas de Compu-

tação, Computação Paralela, Engenharia de Software, Hardware, Software Básico, Teleinformática]  
Especializações do cluster [Ciência da Computação]

### Cluster 12

<Rafael Luiz Cancian>

<Jean Everson Martina>

Áreas de conhecimento do cluster [Arquitetura de Sistemas de Computação, Astronomia Espacial, Astronomia Fundamental, Automação Eletrônica de Processos Elétricos e Industriais, Avaliação de Sistemas, Instituições, Planos e Programas Educacionais, Banco de Dados, Bioinformática, Cerimônias Criptográficas, Circuitos Eletrônicos, Clusters de Processamento Distribuído, Computabilidade e Modelos de Computação, Controle de Processos Eletrônicos, Retroalimentação, Criptoanálise, Dinâmica dos Fluidos, Economia Matemática, Emissão de Documentos Digitais, Engenharia de Software, Geomorfologia Submarina, Gerenciamento de Certificados Digitais, Hardware, Inferência Paramétrica, Inteligência Artificial, Interação do Oceano com a Atmosfera, Interação entre os Organismos Marinhos e os Parâmetros Ambientais, Linguagens de Programação, Lógicas e Semântica de Programas, Manejo e Conservação de Recursos Pesqueiros Marinhos, Maricultura, Matemática Simbólica, Materiais e Componentes Semicondutores, Mecanismos de Assinatura Digital, Modelagem de Sistemas Biológicos, Modelos Analíticos e de Simulação, Métodos Formais para Segurança, Métodos e Modelos Matemáticos, Econométricos e Estatísticos, Métodos e Técnicas de Ensino, Operação de Sistemas de Transporte, Planejamento de Experimentos, Planejamento e Organização do Sistema de Transporte, Processos de Colisão e Interações de Átomos e Moléculas, Proteínas, Protocolos Criptográficos, Práticas de Segurança, Redes Privadas Virtuais, Redes Troncalizadas, Redes de Computadores, SISTEMAS OPERACIONAIS, Sedimentologia Marinha, Segurança de Redes, Sistema Planetário, Sistema de Tempo Real, Sistemas Dinâmicos, Sistemas Distribuídos, Sistemas Eletrônicos de Medida e de Controle, Sistemas de Informação, Sistemas de Tempo Real, Software Básico, Séries Temporais, Teleinformática, Tempo real, Teoria Geral de Planejamento e Desenvolvimento Curricular, Transporte e Tráfego Urbano e Regional]

Especializações do cluster [Astronomia, Bioquímica, Ciência da Computação, Economia, Educação, Engenharia Biomédica, Engenharia Elétrica, Engenharia de Produção, Engenharia de Transportes, Fí-

sica, Matemática, Oceanografia, Planejamento Urbano e Regional, Probabilidade e Estatística, Recursos Pesqueiros e Engenharia de Pesca]

Clusters não unitários: 10



## ANEXO A - Artigo



# Agrupando pesquisadores por coautoria de publicações segundo currículo Lattes

Gabriela Bussolo Colonetti

9 de dezembro de 2016

## Resumo

Técnicas de agrupamento de dados são responsáveis por criar grupos, ou *clusters*, de objetos que são semelhantes para que se possa extrair informações desses grupos com facilidade. Para comparar pesquisadores, diferentes abordagens são possíveis. Uma delas é com base em competências extraídas de suas produções via palavras-chave, títulos etc. Outra possibilidade é considerar a coautoria, ou seja, se um profissional X publicou junto com outro profissional Z e Y também publicou com Z, há uma possível similaridade de interesse entre X e Y, sendo interessante sugerir uma colaboração entre eles. Este trabalho aborda a clusterização de pesquisadores por coautoria sobre o Currículo Lattes, comparando e agrupando (*clusterizando*) profissionais e utiliza-se trabalhos existentes como apoio e baseline.

**Palavras-chaves:** Conhecimento. Agrupamento de dados. Lattes. Clusterização. Coautoria.

## Introdução

O crescimento da comunidade acadêmica, tanto em membros quanto em áreas de pesquisa e publicações, torna complexo analisar e identificar profissionais com interesses semelhantes para sugerir colaborações. Desta forma, automatizar processos que façam essa análise é vantajoso, como por exemplo, extrair e analisar informações dos currículos dos pesquisadores.

No Brasil, o CNPQ implantou a plataforma Lattes para criação e gestão de currículos virtuais com intuito de integrar currículos profissionais de instituições e grupos de pesquisa (CNPQ, 2015).

Este trabalho utiliza a coautoria para clusterização de pesquisadores. Verificando as publicações dos profissionais e analisando os coautores. Busca-se verificar se pelas coautorias pode-se construir *clusters* de profissionais afins

em sua pesquisa. A capacidade de definir grupos de profissionais com características ou interesses acadêmicos em comum é uma forma de incentivar o compartilhamento de experiências e conhecimentos (LI et al., 2014).

Pesquisar colaborativas são cada vez mais populares e importantes no ciclo acadêmico, porém não existem plataformas para recomendação desses potenciais colaboradores (CHEN et al., 2011). Neste trabalho será explorado a possibilidade de descobrir colaboradores com base na estrutura da rede de coautores e interesses de pesquisa de um usuário.

## 1 Técnicas de agrupamento

Técnicas ou algoritmos de agrupamento permitem criar grupos de dados para determinar padrões a partir dos dados (ABERNETHY, 2010). Técnicas de agrupamento podem ser separadas em três grandes grupos: agrupamento por partição total, agrupamento hierárquico e agrupamento baseado em densidade (GROSS, 2014), que são exploradas a seguir:

- Agrupamento por partição total: Técnica que gera grupos distintos sem relação entre eles - os objetos são agrupados de forma que todos os elementos de um mesmo aglomerado tenham um grau de similaridade mínima (*threshold*) entre eles (GROSS, 2014). A técnica de agrupamento por partição total foi adotada no desenvolvimento deste trabalho;
- Agrupamento hierárquico: Um algoritmo hierárquico que inicia com um agrupamento disjunto, ou seja, todos os elementos pertencem a um único grupo, colocando cada objeto em um grupo individual. Definidos os grupos iniciais, eles são analisados aos pares e, ao encontrar o maior grau de similaridade com outro grupo, um novo grupo é criado contendo os dois grupos. O processo é feito recursivamente até que se forme apenas um único grupo raiz (GROSS, 2014);
- Agrupamento baseado em Densidade: A ideia desse tipo de agrupamento é agrupar objetos vizinhos de um grupo de objetos em um *cluster* baseado em condições de densidade (GROSS, 2014). Esses algoritmos tipicamente consideram como aglomerados regiões densas de objetos no espaço de objetos.

## 2 Agrupamento por coautoria

Nas próximas seções serão apresentadas as principais etapas para a definição e implementação do algoritmo de agrupamento de pesquisadores com base em suas coautorias. Como entrada para o agrupador temos uma série de currículos Lattes em formato XML, nos quais foram extraídos os dados necessários para o agrupador através da linguagem de consulta XML *Path*



*Language* (XPath). Técnicas implementadas no algoritmo de Chen et al. (2011) foram utilizadas para criação dos *clusters*.

## 2.1 Extração dos dados

O objetivo desta primeira etapa é a extração de nodos relacionados aos artigos publicados dos autores, além de informações referentes às áreas de atuação de cada currículo processado. Foram considerados elementos "ARTIGO-PUBLICADO", "TRABALHO-EM-EVENTOS" e "AREA-DO-CONHECIMENTO", sendo os atributos "ANO-DO-AR-TIGO", "ANO-DO-TRABALHO", "TITULO-DO-ARTIGO", "TITU-LO-DO-TRABALHO", "NOME-COMPLETO-DO-AUTOR", "NRO-ID-CNPQ", "NOME-DA-AREA-DO-CONHECIMENTO" e "NOME-DA-ESPECIALIDADE" extraídos de cada currículo de pesquisador analisado.

Os dados referentes aos trabalhos em eventos e artigos publicados são organizados em listas de *Publicacoes* compostas dos atributos: título, ano, autor principal e uma lista de coautores. O autor e os coautores de cada publicação são objetos do tipo *Pessoa* constituídos de nome, número identificador do CNPq e um *HashMap* que compila os nomes das áreas de conhecimento e nomes das especialidades retiradas de todo documento XML. Os atributos "nome-area-do-conhecimento" e "nome-especialidade" são extraídos em pares dos elementos "AREA-DO-CONHECIMENTO" encontrados ao longo de cada currículo Lattes e apenas os pares que possuem os dois atributos preenchidos são inseridos no mapa. Somente os autores que tem seus currículos processados, ou seja, os autores que são donos dos currículos, possuem a área de atuação preenchida.

Relativo aos pesquisadores primeiramente ficou decidido utilizar o número identificador do CNPQ<sup>1</sup> como chave primária para identificar cada autor e coautor, porém, após um estudo mais aprofundado, foi verificado a falta deste número em alguns profissionais. Por este motivo, o nome do pesquisador também foi extraído do XML e utilizado para sua identificação. No início do desenvolvimento não foi feito nenhum tratamento em relação aos nomes divergentes do mesmo profissional. No entanto, posteriormente foi verificado que as diferenças entre nomes referenciando a mesma pessoa poderia ocasionar alteração no resultado. Por isso, medidas foram implementadas e são explicadas na seção 2.5.

## 2.2 Criação do Grafo não Direcionado

Após a criação da lista de objetos *Tupla* com os atributos necessários, foi realizada a criação do grafo não direcionado, utilizando métodos da biblioteca JGraphT. O grafo criado dispõe dos nomes dos autores como vértices e as arestas são valoradas pelo número de coautorias. O cálculo feito na classe *Tupla* é realizado da seguinte forma: cada vez que o programa encontra uma

---

<sup>1</sup> <http://cnpq.br/>

publicação de ano e título diferente de uma publicação já inserida na lista de publicações da tupla analisada, esta publicação é adicionada na lista e o peso é incrementado.

### 2.3 Clusterização

Para a clusterização dos pesquisadores com base em suas coautorias em publicações, foi implementado, em um primeiro momento, um algoritmo simples denominado "Similaridade por Coautoria Direta - SCD", da seguinte forma:

1. Foi estabelecido um *threshold*, ou seja, um valor mínimo para o relacionamento entre os autores, de valor três;
2. O algoritmo faz uma busca em todo o grafo para encontrar o maior peso entre dois pesquisadores, ou seja, o maior peso de uma aresta do grafo. Após descobrir quais são esses vértices (pesquisadores), a aresta é adicionada em um objeto `Grupo`, classe que possui uma lista de atributos do tipo `DefaultWeightedEdge`, que é uma classe referente ao vértice na biblioteca `JGraphT`;
3. Verifica-se então as arestas dos dois vértices (da aresta maior), as arestas que dispõem de um valor maior que o *threshold* será adicionado também ao `Grupo`. O processo é repetido até não existir mais ligações entre membros desse grupo inicial com valores superiores ao *threshold*;
4. Feito isso, o algoritmo repete o processo, criando um novo grupo, até não existir mais nenhuma aresta com peso superior ao *threshold*;
5. Os pesquisadores que não estiveram em nenhum *cluster*, são inseridos um por um em grupos individuais;
6. Durante o processo, cada vez que alguma aresta é adicionada em um grupo, seu peso é alterado para zero. Assim em uma nova busca pelo grafo, aquela aresta não será incluída em nenhum outro grupo.

### 2.4 Análise da Similaridade entre os Vértices

*CollabSeer: A Search Engine for Collaboration Discovery* (CHEN et al., 2011) é um trabalho que consiste em uma plataforma de recomendação de colaboradores em pesquisas científicas, realizando clusterizações com base em coautoria, no qual também é possível encontrar um profissional de uma determinada expertise através de buscas por palavras-chaves referentes a área de pesquisa desejada.

Neste trabalho, a clusterização é feita com base no *dataset* do CiteSeerX, repositório para trabalhos científicos e acadêmicos com foco em ciência

da computação e da informação (CITeseerX, 2015). O artigo realizou cálculos de similaridade entre os autores com base em coautorias utilizando três frações de similaridade: *Jaccard*, *Cosine* e *Overlap*, elas foram implementadas neste trabalho e são através das fórmulas à seguir:

1. Coeficiente de *Jaccard*: A equação define que a relação entre dois autores, ou seja, dois vértices de um grafo, é medida de acordo com os conjuntos vizinhos, onde a quantidade da intersecção dos vizinhos entre dois vértices é dividida pela quantidade da união dos vizinhos entre esses dois autores. A Equação 1 ilustra a fórmula da equação, em que  $m_i$  e  $m_j$  são os conjuntos de vizinhos dos vértices  $v_i$  e  $v_j$ ,  $\tau$  representa a quantidade,  $\cap$  intersecção e  $\cup$  a união.

$$S_{Jaccard}(v_i, v_j) = \frac{\tau(m_i \cap m_j)}{\tau(m_i \cup m_j)} \quad (1)$$

2. Coeficiente de *Cosine*: O cálculo do coeficiente de *Cosine* é similar ao de *Jaccard*, com a diferença no denominador, conforme mostra a Equação 2, a quantidade ( $\tau$ ) da intersecção ( $\cap$ ) de vizinhos  $m_i$  e  $m_j$  dos vértices  $v_i$  e  $v_j$  é dividida pela multiplicação da quantidade de vizinhos que cada vértice possui e, então, é aplicado a raiz quadrada sobre o valor.

$$S_{Cosine}(v_i, v_j) = \frac{\tau(m_i \cap m_j)}{\sqrt{\tau(m_i)\tau(m_j)}} \quad (2)$$

3. Coeficiente de *Overlap*: O coeficiente utiliza também os vizinhos para calcular a similaridade entre dois vértices, porém a quantidade ( $\tau$ ) da intersecção de vizinhos entre dois vértices é dividido pelo valor mínimo de vizinhos de um dos dois vértices analisados, ou seja, se o vértice  $v_i$  possuir 10 vizinhos e o vértice  $v_j$  possuir 5 vizinhos, o denominador da equação é o valor 5. A Equação 3 representa a equação.

$$S_{Overlap}(v_i, v_j) = \frac{\tau(m_i \cap m_j)}{\min(\tau(m_i)\tau(m_j))} \quad (3)$$

Devido ao fato dos testes realizados com o algoritmo SCD não apresentarem resultados satisfatórios foi implementado, em um segundo momento, o algoritmo baseado no artigo *CollabSeer: A Search Engine for Collaboration Discovery* (CHEN et al., 2011).

## 2.5 Criação dos Grupos

Após calculados todos os coeficientes de similaridade utilizando as três equações, foi aplicado o *framework* URSA. O algoritmo utilizado é implementado pela biblioteca, o *Best-Star*, tem como abordagem atribuir um objeto ao

grupo de objetos que possui maior similaridade e realocar objetos não centróides, ou seja, que não são pontos iniciais da criação do grupo, para grupos com similaridade superior ao que está inserido.

O algoritmo obteve um resultado não muito satisfatório, em virtude de muitos autores colaboradores em artigos aparecerem com nomes divergentes em currículos distintos e dos autores sem currículo sendo processados gerarem muitos ruídos nos dados. Com isso, o algoritmo gerou uma série de grupos individuais.

Com o intuito de buscar melhores resultados para clusterização, foi realizada uma nova análise nos currículos e observado que os nomes que haviam muitas divergências em função de abreviações e inconsistências eram normalmente dos pesquisadores que não estavam no *dataset* de currículos Lattes utilizados para o trabalho, ou seja, quando o pesquisador insere uma nova pesquisa em seu currículo, seu nome é inserido de forma completa e correta, porém os de seus parceiros pesquisadores eram abreviados ou inseridos em formato de citação.

Desta forma, foi realizado um tratamento nos nomes que estavam em formato de citação, isto é, os nomes que estavam no formato "Sobrenome, Nome". No entanto, ainda assim, a clusterização não teve uma melhora considerável, em razão de que muitas vezes o autor possui dois sobrenomes e seu nome era inserido no currículo de diversas formas.

Muitos autores/coautores que não possuem currículos sendo processados geraram ruídos nos dados através de ligações unitárias, foi possível observar que os pesquisadores diversas vezes publicavam com um autor apenas uma vez o fato ocorre em decorrência de orientações de graduação ou mestrado, por exemplo, no qual a publicação realiza-se em virtude de um desenvolvimento de trabalho de conclusão de curso.

Para minimizar o problema dos ruídos os autores/coautores que não possuíam currículos sendo processados geravam, foram alteradas as consultas do XPath na classe `ManipuladorXML` para, assim, de cada elemento `ARTIGO-PUBLICADO` e `TRABALHO-EM-EVENTOS` ser retirado os títulos das respectivas publicações e apenas o nome do autor/coautor que corresponde ao pesquisador que o currículo pertence. As informações foram inseridas em um *HashMap* no qual o título da publicação é a chave do mapa e os autores são inseridos em uma lista nos valores. Dessa forma, toda vez que for encontrado em um currículo um título de trabalho ou publicação já disponível no mapa, o autor é inserido na lista de valores do *HashMap*.

Na análise dos currículos também foi observado que alguns títulos de mesmo artigo eram escritos com palavras diferentes. Testes foram feitos para calcular a similaridade entre títulos, porém o custo para processamento era alto e o resultado não trouxe melhoras para o problema, já que alguns títulos eram muito abreviados e a similaridade calculada era muito próxima a zero.

### 3 Experimentos

Neste capítulo são descritos os experimentos realizados. Eles foram feitos com um conjunto de cinquenta currículos Lattes de professores do INE/UFSC, cinco testes foram feitos:

1. Aplicando o algoritmo *Best-Star* apenas nos pesquisadores que possuem currículos no *dataset* de currículos e calculando a similaridade através do método de *Jaccard*;
2. Aplicando o algoritmo *Best-Star* apenas nos pesquisadores que possuem currículos no *dataset* de currículos e calculando a similaridade através do método de *Cosine*;
3. Aplicando o algoritmo *Best-Star* apenas nos pesquisadores que possuem currículos no *dataset* de currículos e calculando a similaridade através do método de *Overlap*;
4. Aplicando o algoritmo *Best-Star* sobre os currículos dos autores de três grupos gerados por método (*Cosine*, *Overlap* e *Jaccard*), considerando todos os autores presentes no currículo;
5. Aplicando o algoritmo SCD apenas sobre os autores que possuem currículos no *dataset* de currículos.

#### 3.1 Resultados *Best-Star*

Com o objetivo de verificar o melhor método para cálculo do coeficiente de similaridade na geração dos *clusters*, foi executado o algoritmo *Best-Star* utilizando as três equações referentes ao artigo de Chen et al. (2011) com grau de similaridade de 0.1 até 0.9, para, assim, examinar qual o *threshold* de similaridade entre os autores seria mais adequado na geração de um bom resultado, ou seja, resultado no qual as relações entre os autores fossem coerentes e não houvesse uma grande geração de grupos unitários.

Para apurar a qualidade do resultado, os pesquisadores foram analisados antes da geração dos grupos com a finalidade de associar alguns autores do *dataset* utilizando como critério sua área de conhecimento e especialidade. As seguintes associações foram feitas:

1. Carla Merkle Westphall e Carlos Becker Westphall;
2. Ronaldo dos Santos Mello e Carina Friedrich Dorneles;
3. Christiane Anneliese Gresse von Wangenheim e Jean Carlo Rossa Hauck;
4. Jean Everson Martina e Ricardo Felipe Custodio;
5. Silvia Modesto Nassar e Renato Cislighi.

Dentre os resultados das três equações, a clusterização pelo método *Overlap* com o *threshold* de similaridade 0.5 gerou o melhor resultado. Ele gerou doze grupos nos quais apenas dois autores ficaram isolados. Utilizando este mesmo valor de similaridade, o coeficiente de *Jaccard* e *Cosine* geraram vinte e três e sete *clusters* com apenas um membro, respectivamente.

Utilizando o coeficiente de *Cosine*, o *threshold* de melhor resultado foi de valor 0.3 e no método de *Jaccard* o valor 0.2 apresentou o melhor ganho.

Um resumo da execução com as três equações pode ser visto nas Tabelas 1, 2, 3.

Tabela 1 – Resultado Coeficiente *Overlap*

Threshold	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Número de Clusters	12	12	12	12	12	19	22	22	22
Clusters com apenas um membro	2	2	2	2	2	9	14	14	14

Tabela 2 – Resultado Coeficiente *Jaccard*

Threshold	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Número de Clusters	12	13	20	26	29	32	33	33	33
Clusters com apenas um membro	2	2	9	18	23	28	30	30	30

Tabela 3 – Resultado Coeficiente *Cosine*

Threshold	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Número de Clusters	12	12	12	16	18	29	30	32	33
Clusters com apenas um membro	2	2	2	4	7	23	25	28	30

Pelo método *Overlap*, dos autores que foram associados anteriormente, apenas os pesquisadores Ronaldo dos Santos Mello com Carina Dorneles e Christiane Anneliese Gresse von Wangenheim com Jean Carlo Rossa Hauck ficaram no mesmo *cluster*. Isso ocorreu em virtude de que alguns autores que eram da mesma especialidade e publicaram juntos não possuíram membros em comum no *dataset*. No método de *Jaccard* e *Cosine* apenas os autores Ronaldo dos Santos Mello com Carina Dorneles ficaram no mesmo grupo.

Outro exemplo a ser citado são os pesquisadores Carla Merkle Westphall e Carlos Becker Westphall. Ambos possuem juntos quarenta e uma pu-

blições em comum, porém, ao realizar o cálculo de similaridade entre os dois, o valor resultou em zero devido ao fato dos autores não possuírem outros pesquisadores em comum no *dataset* além deles mesmos. Com isso, ao aplicar os métodos de similaridade, o numerador, ou seja, a intersecção entre eles, era zero e o cálculo resultava em zero. As publicações do Carlos com outros autores INE/UFSC presentes em seu currículo Lattes não incluía Carla Merkle Westphall como coautora e vice-versa. Sendo assim, não existe intersecção de coautores entre eles.

### 3.2 *Best-star* Utilizando Todos Autores

Para este teste, foram utilizados vinte e sete currículos de professores do INE/UFSC. A escolha dos pesquisadores foi feita através dos resultados dos testes com o algoritmo *Best-Star* da subseção 3.1, no qual foram extraídos os autores de três grupos não unitários de cada método.

O resultado de todos os métodos, no entanto, não foi superior aos resultados da subseção 3.1, já que centenas de grupos unitários foram criados utilizando *threshold* de 0.1 à 0.9. Portanto, é vantajoso considerar apenas os autores dos currículos Lattes do *dataset*.

### 3.3 Resultados SCD

Nesta seção são expostos os resultados referentes à execução utilizando o algoritmo SCD implementado na primeira parte deste trabalho. Os testes foram executados considerando apenas os autores dos currículos Lattes e com pesos de três à dez. Isso significa que o algoritmo agrupou apenas arestas (pares de pesquisadores) que possuem no mínimo três publicações em conjunto.

Os resultados dos testes também não foram superiores ao algoritmo *Best-Star*. Apesar de alguns grupos gerados nos dois métodos serem similares, muitos grupos unitários foram gerados no teste, o que deixa o resultado da Seção 3.1 ainda como o melhor resultado encontrado no trabalho.

Os resultados obtidos neste método podem ser considerados mais adequado para achar autores que publicam juntos, no entanto, não para agrupar autores com "coautorias semelhantes".

## Conclusões

Este trabalho teve como objetivo implementar uma ferramenta que, ao receber uma série de currículos Lattes, em formato XML, gere grupos de pesquisadores através de suas coautorias. No primeiro momento do trabalho, os *clusters* foram gerados de acordo com o algoritmo SCD. Os resultados não foram satisfatórios devido ao fato do método ser mais adequado para localizar coautores que publicam juntos. Com isso, outros algoritmos de agrupamento foram executados visando uma aprimoração dos resultados.

Conforme pode-se concluir, foi necessário uma análise mais detalhada nos currículos, já que o conteúdo do documento diversas vezes não condiz com o esperado, uma vez que alguns pesquisadores não preenchem todos os campos de forma correta e completa em seu currículo Lattes. Com a utilização do XPath, foi possível retirar os dados necessários dos documentos em XML e aplicar métodos de similaridade com base em trabalhos relacionados, bem como, executar algoritmos de *clustering* implementados por bibliotecas já existentes.

Com isso, os grupos gerados continham os nomes dos pesquisadores, agrupados com a união da área de conhecimento e especialização de cada membro do grupo, o que permite concluir que os resultados da análise realizada na subseção 3.1 é convincente, pois os grupos gerados através de cada método de similaridade resulta em *clusters* coerentes. Os testes das subseção 3.2 e 3.3 não demonstraram um melhor ganho para o trabalho, em virtude do excesso de grupos unitários não compensa considerar todos os autores presentes nos currículos para clusterização com base em coautorias, assim como o algoritmo SCD implementado na primeira parte do trabalho.

Contudo, a capacidade de definir grupos de pesquisadores com base em suas coautorias em comum é uma forma de estimular o compartilhamento de interesses e conhecimento de profissionais. Como trabalho futuro, os experimentos podem ser mais amplos e considerando as expertises dos autores, visando verificar se comumente relacionam-se. Para isso o problema de escalabilidade precisa ser aprimorado, já que com muitos currículos o processo de clusterização tende a ser lento na implementação atual.

## Referências

ABERNETHY, M. *Mineração de dados com o WEKA*. [S.l.], 2010. Disponível em: <<http://www.ibm.com/developerworks/br/opensource/library/os-weka2/>>. Acesso em: 22/10/2016. Citado na página 2.

CHEN, H.-H. et al. Collabseer: A search engine for collaboration discovery. In: *Proceedings of the 11th Annual International ACM/IEEE Joint Conference on Digital Libraries*. New York, NY, USA: ACM, 2011. (JCDL '11), p. 231–240. ISBN 978-1-4503-0744-4. Disponível em: <<http://doi.acm.org/10.1145/1998076.1998121>>. Citado 5 vezes nas páginas 2, 3, 4, 5 e 7.

CITeseerX. *About CiteSeerX*. [S.l.], 2015. Disponível em: <<http://csxstatic.ist.psu.edu/about>>. Acesso em: 03/10/2016. Citado na página 5.



CNPQ. *Sobre o CNPQ*. [S.l.], 2015. Disponível em: <<http://www.cnpq.br/web/portal-lattes/sobre-a-plataforma>>. Acesso em: 28/06/2015. Citado na página 1.

GROSS, J. *Um framework para agrupamento de dados e validação de resultados*. Porto Alegre, 2014. Citado na página 2.

LI, J. et al. Acrec: A co-authorship based random walk model for academic collaboration recommendation. In: *Proceedings of the 23rd International Conference on World Wide Web*. New York, NY, USA: ACM, 2014. (WWW '14 Companion), p. 1209–1214. ISBN 978-1-4503-2745-9. Disponível em: <<http://doi.acm.org/10.1145/2567948.2579034>>. Citado na página 2.



## REFERÊNCIAS

- ABERNETHY, M. *Mineração de dados com o WEKA*. [S.l.], 2010. <<http://www.ibm.com/developerworks/br/opensource/library/os-weka2/>>. Acessado em 22/10/2016.
- ALMEIDA, M. Uma introdução ao xml, sua utilização na internet e alguns conceitos complementares. *Ciência da Informação*, v. 31, n. 2, 2002. ISSN 1518-8353. <<http://revista.ibict.br/ciinf/article/view/955>>.
- ALMEIDA, N. Z. C. O. *Projeto de Algoritmos com Implementações em Pascal e C*. São Paulo - SP, Brasil: Ed. Cengage Learning, 2004.
- CHEN, H.-H. et al. Collabseer: A search engine for collaboration discovery. In: *Proceedings of the 11th Annual International ACM/IEEE Joint Conference on Digital Libraries*. New York, NY, USA: ACM, 2011. (JCDL '11), p. 231–240. ISBN 978-1-4503-0744-4. <<http://doi.acm.org/10.1145/1998076.1998121>>.
- CITeseerX. *About CiteSeerX*. [S.l.], 2015. <<http://csxstatic.ist.psu.edu/about>>. Acessado em 03/10/2016.
- CNPQ. *Sobre o CNPQ*. [S.l.], outubro 2015. <<http://www.cnpq.br/web/portal-lattes/sobre-a-plataforma>>. Acessado em 28/06/2015.
- CNPQ, P. L. *Histórico CNPq*. [S.l.], 2016. <<http://lattes.cnpq.br>>. Acessado em 27/06/2016.
- GROSS, J. *Um framework para agrupamento de dados e validação de resultados*. Porto Alegre, Novembro 2014.
- JGRAPHT. *About JGraphT*. [S.l.], 2016. <<http://jgrapht.org/>>. Acessado em 01/07/2016.
- LI, J. et al. Acree: A co-authorship based random walk model for academic collaboration recommendation. In: *Proceedings of the 23rd International Conference on World Wide Web*. New York, NY, USA: ACM, 2014. (WWW '14 Companion), p. 1209–1214. ISBN 978-1-4503-2745-9. <<http://doi.acm.org/10.1145/2567948.2579034>>.

LINDEN, R. *Técnicas de Agrupamento*. [S.l.], 2009. 32 p.  
<[http://www.fsma.edu.br/si/edicao4/FSMA\\_SI\\_2009\\_2\\_Tutorial.pdf](http://www.fsma.edu.br/si/edicao4/FSMA_SI_2009_2_Tutorial.pdf)>.  
Acessado em 01/07/2016.

RIBACKI, G. *Um framework para agrupamento de dados*. Porto Alegre, Janeiro 2013.

SILVEIRA, F. V. Fragmentação e decomposição de consultas em xml. p. 33, 2006.

STELA, I. *Histórico*. [S.l.], 2016. <<http://www.stela.org.br>>.  
Acessado em 27/06/2016.

VAZQUEZ, F. *XPath sem Traumas*. [S.l.], 2005.  
<<https://www.microsoft.com/brasil/msdn/Tecnologias/visualc/XPath.msp>>.  
Acessado em 22/06/2016.

W3C. *Extensible Markup Language (XML) 1.0 (Fifth Edition)*. [S.l.], 2008. <<http://www.w3.org/TR/xml/>>. Acessado em 01/06/2016.