



# **CENTRO UNIVERSITÁRIO LUTERANO DE PALMAS**

*Recredenciado pela Portaria Ministerial nº 1.162, de 13/10/16, D.O.U. nº 198, de 14/10/2016*  
AELBRA EDUCAÇÃO SUPERIOR - GRADUAÇÃO E PÓS-GRADUAÇÃO S.A.

Caio Vinicius Davantel

CARACTERIZAÇÃO DA REDE DE COAUTORIA DOS CURSOS DE CIÊNCIA DA  
COMPUTAÇÃO, SISTEMAS DE INFORMAÇÃO E ENGENHARIA DE SOFTWARE DO  
CEULP/ULBRA A PARTIR DAS PUBLICAÇÕES DO ENCOINFO E PLATAFORMA  
LATTES

Palmas – TO

2021

Caio Vinicius Davantel

CARACTERIZAÇÃO DA REDE DE COAUTORIA DOS CURSOS DE CIÊNCIA DA  
COMPUTAÇÃO, SISTEMAS DE INFORMAÇÃO E ENGENHARIA DE SOFTWARE DO  
CEULP/ULBRA A PARTIR DAS PUBLICAÇÕES DO ENCOINFO E PLATAFORMA  
LATTES

Monografia apresentada como requisito parcial para aprovação na disciplina de Laboratório de Criação do curso de bacharel em Engenharia de Software pelo Centro Universitário Luterano de Palmas (CEULP/ULBRA).

Orientador: Prof. M.e Fabiano Fagundes.

Palmas – TO

2021

Caio Vinicius Davantel

CARACTERIZAÇÃO DA REDE DE COAUTORIA DOS CURSOS DE CIÊNCIAS DA  
COMPUTAÇÃO, SISTEMAS DA INFORMAÇÃO E ENGENHARIA DE SOFTWARE DO  
CEULP/ULBRA A PARTIR DAS PUBLICAÇÕES DO ENCOINFO E PLATAFORMA  
LATTES

Monografia apresentada como requisito parcial para  
aprovação na disciplina de Laboratório de Criação do  
curso de bacharel em Engenharia de Software pelo  
Centro Universitário Luterano de Palmas  
(CEULP/ULBRA).

Orientador: Prof. M.e Fabiano Fagundes.

Aprovado em: \_\_\_\_/\_\_\_\_/\_\_\_\_

BANCA EXAMINADORA

---

Prof. M.e Fabiano Fagundes

Orientador

Centro Universitário Luterano de Palmas – CEULP

---

Prof. M.e Jackson Gomes de Souza

Centro Universitário Luterano de Palmas – CEULP

---

Prof. Esp. Fábio Castro Araújo

Centro Universitário Luterano de Palmas – CEULP

Palmas – TO

2021

## RESUMO

DAVANTEL, Caio V. **Caracterização da rede de coautoria dos cursos de Ciências da computação, Sistemas da Informação e Engenharia de Software do CEULP/ULBRA a partir das publicações do ENCOINFO e Plataforma Lattes**. 2021. Projeto Tecnológico (Graduação) – Curso de Engenharia de Software, Centro Universitário Luterano de Palmas, Palmas/TO, 2021.

A caracterização das redes de coautorias tem sido abordada com grande frequência nos últimos anos. O seu estudo proporciona uma ampla visão de como está se desenvolvendo o campo de pesquisa de determinada instituição de ensino, programa ou curso. O presente trabalho teve como propósito caracterizar a rede de coautoria dos cursos de Ciência da Computação, Sistemas de Informação e Engenharia de Software do CEULP/ULBRA, utilizando os dados disponibilizados pelas publicações do ENCOINFO e Plataforma Lattes. Desta forma, foi possível extrair indicadores que revelam como o campo de pesquisa evolui e como ele está sendo gerenciado, fornecendo *insights* e um agregado de informações para os gestores da instituição e cursos citados, contribuindo na tomada de decisão em relação ao campo científico. Para obter os resultados foram utilizados grafos, que facilitaram a visualização da rede, e por conseguinte, geraram alguns parâmetros que foram mensurados e estruturados em tabelas e gráficos, com as ferramentas *Gephi* e *Excel*. Ao analisar estes parâmetros foram obtidas informações sobre as coautorias estabelecidas nessas redes, podendo definir a densidade, a centralização de autores nas produções e as diferentes comunidades científicas. Como conclusão, pode-se perceber que as redes são caracterizadas predominantemente por fracas ligações entre os autores, número elevado de grupos que produzem somente com autores do mesmo campo científico e um número de comunidades médio para o tamanho da rede. Também percebeu-se a importância de uma visualização dos dados dessas redes em grafos, para que contribuam para o entendimento e direcionamento das pesquisas nos cursos citados, além de possibilitar medir a existência de uma relação científica real e como ela ocorre.

Palavras chave: Redes de coautoria, coautoria científica, grafos.

## LISTA DE FIGURAS

<b>Figura 1</b> - Grafo utilizado na representação de uma rede social.	23
<b>Figura 2</b> - Fluxograma dos procedimentos metodológicos do trabalho.	27
<b>Figura 3</b> - Coleta de dados	30
<b>Figura 4</b> - Template de inserção dos currículos em formato .xml	31
<b>Figura 5</b> - Visualização dos currículos inseridos na ferramenta <i>Coletaprod</i> .	31
<b>Figura 6</b> - Planilha dos nós	32
<b>Figura 7</b> - Planilha das arestas	32
<b>Figura 8</b> - Código para ler o arquivo html.	33
<b>Figura 9</b> - Segunda etapa do código.	33
<b>Figura 10</b> - Arquivo .csv com autores e artigos coletados.	34
<b>Figura 11</b> - Tratamento de dados	35
<b>Figura 12</b> - Planilha de nós com nome dos autores e ID.	35
<b>Figura 13</b> - Planilha de arestas com seus autores de Origem e Destino.	37
<b>Figura 15</b> - Importação dos nós na ferramenta <i>Gephi</i> .	38
<b>Figura 16</b> - Seleção de grafo dirigido e não dirigido.	38
<b>Figura 17</b> - Inserção das arestas na ferramenta <i>Gephi</i> .	39
<b>Figura 18</b> - Peso das ligações entre os nós.	39
<b>Figura 19</b> - Grafo inicial da rede de coautoria.	40
<b>Figura 20</b> - Algoritmos de layout do <i>Gephi</i> .	41
<b>Figura 21</b> - Algoritmo Force Atlas 2 na rede.	41
<b>Figura 22</b> - Tela de configuração do algoritmo Force Atlas 2 no <i>Gephi</i> .	42
<b>Figura 23</b> - Algoritmo Fruchterman-Reingold na rede.	43
<b>Figura 24</b> - Tela de configuração do algoritmo Fruchterman-Reingold no <i>Gephi</i> .	44
<b>Figura 25</b> - Grafo da rede de coautoria segundo os artigos publicados pelos professores na Plataforma Lattes nos anos de 2015 a 2020.	45
<b>Figura 26</b> - Grafo A e B representando respectivamente as redes de coautorias dos anos de 2015/16/17 e 2018/19/20 segundo os artigos publicados pelos professores na Plataforma Lattes.	46
<b>Figura 27</b> - Grafo da rede de coautoria segundo os artigos publicados no ENCOINFO nos anos de 2015 a 2020.	47
<b>Figura 28</b> - Grafo A e B representando respectivamente as redes de coautorias dos anos de 2015/16/17 e 2018/19/20 segundo os artigos publicados no ENCOINFO.	49
<b>Figura 29</b> - Números totais de artigos, coautores e ligações.	50
<b>Figura 30</b> - Gráfico do número de artigos e número de coautores por ano.	50
<b>Figura 31</b> - Gráfico do número de ligações dos dez primeiros coautores da rede de coautoria.	51
<b>Figura 32</b> - Números totais de artigos, coautores e ligações.	52
<b>Figura 33</b> - Gráfico do número de artigos e número de coautores por ano.	52
<b>Figura 34</b> - Gráfico do número de ligações dos dez primeiros coautores da rede de coautoria.	52

## **LISTA DE TABELAS**

**Tabela 1** - Descrição dos problemas encontrados na mensuração de uma rede de coautoria. **19**

**Tabela 2** - Cronograma de etapas do projeto. **29**

## SUMÁRIO

<b>1 INTRODUÇÃO</b>	<b>10</b>
<b>2 REFERENCIAL TEÓRICO</b>	<b>13</b>
2.1 COLABORAÇÃO CIENTÍFICA	13
2.2 REDES DE COAUTORIAS	15
2.2.1 Coleta e tratamento de dados	16
2.2.2 Limites da rede	17
2.2.3 Desambiguação do nome do autor	18
2.2.4 Estruturação das redes de coautorias	20
2.2.5 Parâmetros	21
2.2.6 Visualização da rede	22
<b>3 METODOLOGIA</b>	<b>24</b>
3.1 MATERIAIS	24
3.1.1 Coletaprod	24
3.1.2 Gephi	24
3.2 MÉTODOS	25
<b>4 RESULTADOS E DISCUSSÃO</b>	<b>28</b>
4.1 COLETA DOS DADOS	28
4.1.1 Coleta dos dados da Plataforma Lattes	29
4.1.2 Coleta dos dados do ENCOINFO	30
4.1.3 Definindo o limite da rede	32
4.2 TRATAMENTO DOS DADOS	32
4.3 GRAFOS COM O GEPHI	36
4.3.1 Inserção dos dados	36
4.3.2 Algoritmos de visualização	38
4.3.2.1 Force Atlas 2	39
4.3.2.2 Fruchterman-Reingold	41
4.4 GRAFOS E PARÂMETROS DAS REDES DE COAUTORIAS DO PROJETO	42
4.4.3 Grafos e parâmetros da rede de coautoria a partir da Plataforma Lattes	42
4.4.4 Grafos e parâmetros da rede de coautoria a partir do ENCOINFO	45
4.5 GRÁFICOS E TABELAS	48
4.5.1 Gráficos e tabelas da rede de coautoria a partir da Plataforma Lattes	48
4.5.2 Gráficos e tabelas da rede de coautoria a partir do ENCOINFO	50
<b>5 CONSIDERAÇÕES FINAIS</b>	<b>52</b>
<b>REFERÊNCIAS</b>	<b>54</b>
<b>APÊNDICES</b>	<b>58</b>
APÊNDICE A - Código Python utilizado na coleta dos dados do ENCOINFO.	59
<b>ANEXOS</b>	<b>60</b>

ANEXO A - Artigo do XXIII Encoinfo – Congresso de Computação e Tecnologias da Informação.	61
ANEXO B - Artigo da XXI JORNADA DE INICIAÇÃO CIENTÍFICA DO CEULP/ULBRA.	72



## 1 INTRODUÇÃO

A colaboração científica é uma prática exercida há muito tempo, antes mesmo do século XX. De acordo com Beaver e Rosen (1978), o primeiro artigo publicado em colaboração científica tem a data definida em 1665. Além disso, os autores apresentam mais 21 produções colaborativas, datadas entre os anos de 1665 e 1760, com destaque para o tema de astronomia, que mais envolveu os pesquisadores nessa época. Dessa forma, os estudos relacionados a coautoria buscam entender a produção científica por meio da análise da colaboração entre os pesquisadores (PARREIRAS *et al.*, 2006).

Segundo Vanz e Stumpf (2010), a coautoria científica é reconhecida a partir do trabalho conjunto de dois ou mais cientistas em projetos de pesquisas, que por consequência compartilham recursos intelectuais, econômicos e físicos. Pode ser definida como uma interação ocorrendo dentro de um contexto social entre dois ou mais cientistas que facilita o compartilhamento de significados e conclusão de tarefas, no que diz respeito a um mutuamente objetivo superordenado e compartilhado (SONNENWALD, 2007).

Devido ao grande número de artigos com coautorias, não é de se espantar que também haja inúmeras motivações para que ela ocorra. Segundo Balancieri (2004), o mais óbvio dos motivos, certamente, é a necessidade de contribuição especializada para alcançar os objetivos da pesquisa, incluindo o ganho de novas habilidades e conhecimento tácito.

A coautoria é uma forma de medir o quão desenvolvido está o campo de pesquisa de determinada instituição ou programa educacional. Além disso, possibilita verificar como estão sendo aplicadas as propostas de ensino e sua relação com a pesquisa, auxiliando a avaliação deste segmento. Um conjunto de colaborações forma uma rede de coautorias que, segundo Silva (2012), contém estratégias para compartilhar informações e conhecimentos através das comunidades científicas, sendo que, dentre os benefícios da colaboração, pode-se obter uma maior qualidade dos artigos e um incremento do número de publicações. Além disso, o estudo das redes de coautoria permite identificar e analisar porque os pesquisadores colaboram e quais as consequências dessas colaborações. Pesquisas apontam a melhoria da qualidade dos artigos e o aumento do número de publicações como benefícios da colaboração (ANDRADE, 2016).

Determinar como a rede de coautoria se enuncia em uma universidade pode contribuir amplamente para o desenvolvimento do ensino, já que grande parte da formação dos pesquisadores parte delas. Portanto, ao analisar essas coautorias, torna-se possível medir a existência de um feito científico real e como ele ocorre, além de influenciar as políticas de

pesquisas, estimular a coordenação e a cooperação, e identificar porque e como os pesquisadores colaboram e quais as consequências dessa ação.

Uma rede de coautoria pode ser estudada através de grafos, que têm sido amplamente utilizados para representar as redes sociais e quantificar a sua estrutura, vide Recuero (2017). Além disso, na opinião de Bufrem, Junior e Sorribas (2011), a visualização por meio desse modelo é considerada mais intuitiva do que a visualização na forma de matrizes, embora os próprios autores afirmem que os dados coletados sejam, normalmente, apresentados dessa forma. Como os grafos são baseados na matemática, a partir de sua análise, pode-se obter indicadores que possibilitam a caracterização das redes de coautorias. Dessa forma, Silva (2012), argumenta que as pesquisas que envolvem esse tema, em sua maioria analisam os seguintes parâmetros: grau de centralidade, a colaboração, a produtividade dos autores e o elitismo da área, a densidade, a transitividade e o tamanho da rede, a distribuição dos laços e cliques.

Para que seja possível chegar a esses indicadores, é preciso que se elabore um instrumento de coleta de dados que possibilite a busca de informações para a construção da rede e, posteriormente, o cálculo dos indicadores (ANDRADE; DAVID, 2015). Normalmente essa coleta é realizada a partir de bases de dados bibliográficas, disponibilizadas por instituições nacionais e internacionais, que possibilitam o desenvolvimento de redes sociais e uma análise estruturada.

Digiampietri et al. (2015), afirma que os currículos da Plataforma Lattes são uma grande fonte de informação da comunidade acadêmica no Brasil. No entanto, existem diversos desafios computacionais na utilização desta fonte de dados, devido à quantidade de dados, processos manuais e dados semi estruturados. Além disso, Lima e Gomes (2019), também mostram que o Centro Universitário Luterano de Palmas (CEULP) dispõe de uma biblioteca digital voltada para armazenar documentos e informações acadêmicas, que caso sejam analisadas e tratadas podem contribuir para a estratégia de negócio da instituição. Dessa forma, faz-se necessário o uso do processo mais adequado de tratamento e organização dos dados em estudos de redes de coautoria, tendo como foco a sistematização dos coautores por meio de matrizes, e por conseguinte, a sua apresentação em grafos estruturados.

Diante dos dados importantes relacionados a produções acadêmicas que são fornecidos pela Plataforma Lattes - projeto do Conselho Nacional de Desenvolvimento Científico e Tecnológico do Brasil (CNPq) e bem como as informações que são fornecidas no site do ENCOINFO – Congresso de Computação e Tecnologias da Informação, evento realizado pelo Departamento de Computação e Informática do CEULP/ULBRA, mostra-se

viável e interessante caracterizar as redes de coautorias dos cursos de Ciência da Computação, Sistemas de Informação e Engenharia de Software do CEULP/ULBRA. Identificar como as redes de colaboração científica estão se desenvolvendo dentro das Instituições de Ensino Superior auxilia no desenvolvimento pedagógico, na medida em que a formação de pesquisadores parte delas, sendo assim, todo o processo de produções acadêmicas devem ser monitorados e mensurados, para que os incentivos e as propostas de ensino se justifiquem (REICHERT *et al.*, 2016).

Assim, este trabalho teve como propósito coletar os dados dos pesquisadores dos cursos mencionados, disponibilizados pela Plataforma Lattes e ENCOINFO e criar grafos que apresentem visualmente as redes de coautorias, possibilitando a obtenção de métricas que contribuam para o entendimento e direcionamento das pesquisas nos cursos citados e obtendo *insights* e direcionamentos que auxiliem no desenvolvimento das pesquisas nos cursos de Ciência da Computação, Sistemas de Informação e Engenharia de Software do CEULP/ULBRA. Desta forma, foi objetivo deste trabalho caracterizar as redes de coautorias destes cursos e para isso, foram coletados os dados dos pesquisadores que estão disponíveis na Plataforma Lattes e anais do ENCOINFO, criados grafos para organizar as redes de coautorias e, por fim, identificadas as características do campo científico nos cursos citados, utilizando parâmetros como grau de centralidade, grau de intermediação, densidade e nível de agregação, o que permitiu apresentar as informações da rede de coautorias em gráficos e tabelas.

Este trabalho está estruturado em cinco capítulos. O segundo capítulo apresenta uma fundamentação teórica sobre duas partes, a primeira é a colaboração científica, e por conseguinte, a segunda é redes de coautorias. Esta segunda parte se divide em seis, sendo a coleta e tratamento dos dados, limites da rede, desambiguação do nome dos autores, estruturação das redes, parâmetros e visualização da rede. O terceiro capítulo descreve os materiais utilizados no desenvolvimento dos grafos, além da metodologia adotada. O quarto capítulo expõe os resultados obtidos e discussões sobre esses resultados. Por fim, o quinto capítulo apresenta as considerações finais e trabalhos futuros.

## 2 REFERENCIAL TEÓRICO

### 2.1 COLABORAÇÃO CIENTÍFICA

A colaboração científica é uma interação social estabelecida há bastante tempo, por conta disso, é utilizada frequentemente para estudos científicos. Segundo Meadows (1999), “na primeira metade do século XX surgiu a ideia de trabalho em equipe como atividade orientada, surgindo os primeiros grupos científicos formados por assistentes de pesquisa, estudantes de doutorado e técnicos, orientados por um pesquisador sênior”. Portanto, pode ser definida como uma interação entre dois ou mais pesquisadores que trabalham juntos em um projeto de pesquisa, onde irão compartilhar recursos intelectuais, econômicos ou físicos para atingirem um objetivo mútuo de produzirem novos conhecimentos científicos (SAMPAIO *et al.*, 2015; VANZ, 2009).

É possível evidenciar a colaboração científica em diferentes níveis e com distintas configurações, desta forma, é requerida uma pesquisa qualificada para cada partição, evitando que apresentem uma alta complexidade e ruídos devido a estudos generalizados. Segundo Savic *et al.* (2017), em todos esses níveis, existem uma busca para entender a estrutura da colaboração, como ela evolui e como se relaciona com a produtividade e o impacto das publicações estudadas. Ainda de acordo com o autor (2017), os vários níveis da colaboração podem ser divididos em: indivíduos, grupos, departamentos, instituições; setores ou áreas e nações.

Macedo (2014) afirma que para realizar uma pesquisa são necessários abrangentes conhecimentos especializados e inúmeros recursos como, os pessoais e financeiros, sendo assim, a colaboração científica se torna mais atrativa e abundante. E essa ampliação das colaborações acadêmicas, pode ser verificada a partir dos dados do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), que apresenta um acréscimo expressivo no total dos grupos de pesquisa, um crescimento de 15.158, em 2002, para 37.640 grupos em 2016, equivalente a 149%. Já o número de pesquisadores cresceu 251% no mesmo período e o de doutores 278%. Além disso, a tecnologia também possui uma parcela de créditos no aumento da colaboração científica, visto que, houve o desenvolvimento de softwares destinados à comunicação e a ampliação do acesso às redes sem fio, que permitiram um contato via computador, possibilitando a colaboração em produções acadêmicas de maneira remota e com um baixo custo (MACEDO, 2014).

Katz e Martin (1997) elencaram algumas vantagens da colaboração para os pesquisadores individuais:

- O compartilhamento de conhecimento, habilidades e técnicas, pois ninguém é capaz de deter todo o tipo de conhecimento;
- Habilidades sociais e de gestão, necessárias para trabalhar como parte de uma equipe;
- Troca de ideias e choque de pontos de vista, que podem gerar novas perspectivas, e ser uma fonte de estimulação e criatividade;
- Gerar novas relações e conexões com uma rede mais ampla de contatos na comunidade científica;
- Aumentar a visibilidade potencial do trabalho, sendo provável que seja citado mais frequentemente e tenha maior impacto.

Portanto, a colaboração se torna maior do que a soma de suas partes, pois todos os benefícios geram uma pesquisa que, em princípio, é realizada de forma mais eficaz e utiliza os conhecimentos de forma mais precisa. Além disso, o envolvimento de pessoas de diferentes origens científicas expande ainda mais os proveitos. Em geral, entende-se que a colaboração seja uma boa interação e que deva ser encorajada, com incentivo à formação de grupos interdisciplinares e entre setores, como universidade e indústria. No entanto, embora traga benefícios aos colaboradores, também implica em custos que podem eventualmente superar seus benefícios, e não devem ser negligenciados, exigindo um monitoramento regular de insumos e resultados (GLÄNZEL; SCHUBERT, 2004).

Na literatura é comum encontrar a coautoria relacionada a colaboração científica, pois esse termo tem sido utilizado com sucesso por muitos pesquisadores das áreas de bibliometria e cientometria para investigar a colaboração entre pessoas, instituições e países. A vantagem no uso dessa técnica é a possibilidade de verificação dos dados por outros autores, visto que, eles estão disponíveis publicamente e permitem o exame de grandes amostras, possibilitando resultados mais significativos do que estudos de caso (KATZ; MARTIN, 1997). Entretanto, é necessário distinguir estes dois termos que não são sinônimos, pois a coautoria é apenas um aspecto da colaboração científica, já que, ela não mede a colaboração na sua totalidade e complexidade, uma vez que há outras formas de colaboração que acontecem informalmente e são difíceis de serem analisadas e quantificadas. (VANZ, 2009).

## 2.2 REDES DE COAUTORIAS

Newman (2001) diz que uma definição razoável para as redes de coautoria é quando dois ou mais autores colaboram em pesquisas. Esta rede pode ser representada por um

conjunto de vértices que denota as pessoas, unidas em pares, e por arestas que denotam a familiaridade entre elas. Portanto, compreende-se por rede, neste contexto, um grupo de indivíduos que se relacionam uns com os outros, mediante determinados objetivos sendo, assim, a rede de coautoria formada por três elementos: autores ou nós, vínculos ou relações e os fluxos (ALEJANDRO; NORMAN, 2005).

Segundo Mueller e Passos (2000), os estudos sobre as comunidades científicas é um tema que tem crescido constantemente em todas as áreas que se interessam pelo assunto, dentre elas a filosofia, a sociologia e a ciência da informação. Essas pesquisas utilizam-se de metodologias quantitativas como, por exemplo, no estudo de citações, produções e mapeamento de ligações entre os pesquisadores. Além disso, recentemente também estão sendo empregados métodos qualitativos, que provêm da sociologia e comunicação, como forma de analisar estas redes de coautoria e conseguir compreender além dos indicadores, identificando fatores do contexto social e do próprio indivíduo.

Degenne e Forsé (1999) afirmam que uma rede não pode ser reduzida a uma simples soma de relações e que a sua forma exerce uma influência em cada relação. Desta forma, é necessária a análise das redes de coautorias, pois proporciona um melhor entendimento do seu funcionamento, dos hábitos e das estratégias dos pesquisadores, além de influenciar políticas de pesquisa e incentivo para a articulação e cooperação (OLIVEIRA, 2017). Entretanto, são poucos ou superficiais os estudos que analisam essas redes, o que torna interessante conhecê-las mais, mapeando seus atores e relações (SILVA, 2012).

De acordo com Stefano e Giordano (2011), a caracterização da rede de coautoria envolve um processo com várias fases sucessivas, desde a coleta e o tratamento de dados até a sua visualização.

### **2.2.1 Coleta e tratamento de dados**

Mesmo que a colaboração entre pesquisadores seja definida em coautoria, existe a possibilidade de considerar fontes de dados numerosas e heterogêneas para a coleta de dados bibliográficos. Normalmente, esses dados são coletados usando documentos de arquivos (fornecidos por bases de dados bibliográficas internacionais e nacionais) em vez de entrevistas ou questionários administrados diretamente aos autores dos artigos (DE STEFANO; GIORDANO, 2011).

Para Martins (2012), essas novas tecnologias, internet, desenvolvimento Web e redes sociais, acarretaram uma explosão informacional sem precedentes na história da humanidade.

O que marca esse novo tempo são duas características principais: a facilidade em coletar dados provenientes dessas múltiplas relações entre pessoas que se formam nesses ambientes digitais (lista de e-mails, blogs, fóruns, publicações, etc.) e o alto poder de processamento desses dados, permitindo variadas formas de análise, premissas e exames de novos modelos, apresentações e novos indicadores. Compreender o que surge desses grandes volumes de dados tem se tornado elemento central de pesquisas científicas de diferentes áreas. Não somente com objetivo de conhecer e monitorar, mas, sobretudo, como ambiente estratégico de mediação em movimentos sociais (MARTINS, 2012).

Sendo assim, os dados coletados para análise das redes de coautoria são sistematizados por meio de matrizes. Já que a análise estrutural da matriz possibilita o cruzamento dos dados, nó e aresta, além disso, utiliza operações matemáticas para mensurar os padrões da rede através de cálculos estruturais. As medidas de importância do pesquisador dentro da rede representam os principais cálculos utilizados na análise de redes de colaboração científica (CERVANTES, 2015; MARTINS, 2012).

São várias as ferramentas que estão ao dispor do investigador para que este proceda à sua investigação. No entanto, é necessário que se tenha o devido cuidado para seleccionar o método mais indicado no tratamento e análise de dados, face à estratégia desenhada, bem como ao paradigma assumido. Uma abordagem mais quantitativa usará técnicas relacionadas com o tratamento de um grande número de variáveis e de observações (RODRIGUES, 2011)

Para realizar o tratamento dos dados, onde estão incluídas as fases de definição do limite da rede e desambiguação dos nomes dos autores, que envolvem certo nível de complexidade, muitos estudos de redes de coautorias utilizaram-se de trabalho manual, ou de softwares específicos que facilitem esse processo, possibilitando uma inserção dos dados e estruturação da rede de maneira mais assertiva. As ferramentas *Vantage Point* e *Microsoft Excel* são as mais utilizadas para essa etapa. A primeira foi utilizada por Franco (2018), Silva (2018) e Oliveira (2017), já a segunda é material de estudo de Silva (2014), Silva e Martins (2016) e Giacheti *et al.* (2017).

### **2.2.2 Limites da rede**

O limite da rede vai determinar quais os dados necessários e adequados para a pesquisa se tornar coesa. Segundo Marsden (2005), definir os limites de uma rede é uma tarefa complicada devido à dificuldade em discernir qual o processo mais adequado. Sendo assim, alguns conceitos devem ser levados em consideração, pois se o interesse do estudo é descrever o padrão de coautoria entre pesquisadores de uma comunidade científica, uma

abordagem de rede inteira pode ser considerada. No entanto, uma rede de coautores quando é analisada de forma completa, se torna um conjunto de atores interrelacionados com um coletivo social bastante limitado. Desta forma, em estudos de rede, uma rede inteira ou separada pode ser usada, já que, a decisão é sobre qual o propósito da pesquisa e quais autores se enquadram na rede requerida.

Laumann, Marsden e Prensky (1989) definiram três estratégias de especificação de limite genéricas:

- Abordagem posicional baseada nas características dos atores ou critérios formais de filiação. Estruturada por autores que se filiam a instituições de pesquisas ou disciplinas, sendo assim, as produções são derivadas dessa relação intraorganizacional e auxiliam na análise de padrões de relacionamentos.
- Abordagem sobre eventos baseado na participação em alguns eventos relacionais. É formada por autores que estão inseridos em bancos de dados de eventos científicos.
- Abordagem relacional guiado por vínculos sociais entre os autores. É representado por uma estrutura de autores envolvidos em disciplinas e instituições específicas, sendo possível a adição de novos atores à rede com o decorrer do tempo.

As combinações dessas abordagens também podem ser extremamente úteis na prática. Por exemplo, os limites da rede de coautoria podem ser definidos usando uma combinação de abordagens relacionais e posicionais (DE STEFANO; GIORDANO, 2011). Portanto, esses métodos têm por objetivo facilitar a identificação de atores dentro de uma rede e a causalidade de suas relações.

### **2.2.3 Desambiguação do nome do autor**

O processo de desambiguação de nomes é antigo, sendo definido pela primeira vez em 1959 com a produção de Newcombe sobre a *Automatic Linkage of Vital Records* (NEWCOMBE, *et al.*, 1959). Desde então, este processo é objeto de estudo de várias publicações, como as de Phelan (1999), Houvardas e Stamatatos (2006), Smalheiser e Torvik (2009) e Kang (2009), que buscam entender como ele afeta os resultados das pesquisas e como solucioná-lo de forma assertiva. No entanto, a desambiguação de nomes na bibliometria continua sendo uma das questões sem solução (TANG; WALSH, 2010).



Nos estudos que analisam redes de colaboração, como a análise recai sobre a rede, a definição correta dos atores (nós) e dos laços (arestas) que os unem se torna fator crítico para a obtenção de resultados corretos e confiáveis. (BARBASTEFANO *et al.*, 2013). Além disso, De Choudhury *et al.* (2010) concluem que as redes que possuem grande quantidade de nós e arestas, quando são construídas com apoio de recursos computacionais podem acarretar vários problemas, gerando informações incorretas devido a abundância de dados. E as redes de coautoria se enquadram nesse caso, podendo até serem caracterizadas como redes complexas, pois envolvem milhares de nós, cuja obtenção e tratamento de dados vêm acontecendo de forma automatizada, dessa forma, devem receber atenção especial nesse processo para evitar conflitos nos resultados (BARBASTEFANO *et al.*, 2013).

Conforme apontado por Kang *et al.* (2009), uma pessoa pode ter vários nomes (sinônimos), decorrentes de abreviações, omissões, mudanças de nome, pseudônimos e erros ortográficos. Além disso, diferentes pessoas podem compartilhar o mesmo nome (homônimos). Casos como esses podem gerar erros na rede de autores erroneamente agregados e desagregados. Na Tabela 1 são elencados alguns problemas, de acordo com Wang *et al.* (2012), no processo de identificação correta dos autores na formação de uma rede de coautoria:

**Tabela 1 - Descrição dos problemas encontrados na mensuração de uma rede de coautoria.**

<b>Cenário de erro</b>	<b>Exemplo</b>
Falsos vértices negativos	Ausência de vértices que deveriam compor a rede
Falsos vértices positivos	Vértices erroneamente inseridos na rede
Falsas arestas negativas	Ausências de ligações entre vértices que deveriam estar presentes;
Falsas arestas positivas	Ligações erroneamente representadas na rede
Vértices falsamente agregados	Situação em que dois ou mais vértices são erroneamente considerados como um único vértice
Vértices falsamente desagregados	Situação em que um vértice é erroneamente considerado como dois ou mais vértices separados.

**Fonte:** WANG *et al.* (2012) adaptado.

Diante disso, alguns estudos voltados para o tratamento de ambiguidade de nomes de autores foram publicados. Smalheiser e Torvik (2009) afirmam que a aplicação de técnicas de mineração de dados e também a possibilidade da identificação única de autores poderiam resolver o problema de ambiguidade no longo prazo, caso uma solução universal fosse adotada e as técnicas fossem otimizadas. Strotmann *et al.* (2009) apresentaram um algoritmo heurístico para eliminar a ambigüidade de nomes de autores de publicações com base em medidas de similaridade bem definidas, comparando as publicações em que seus nomes aparecem como autores.

Cota, Gonçalves e Laender (2007) propuseram o agrupamento Hierárquico Heurístico (*heuristic-based hierarchical clustering* - HHC), um método específico para a desambiguação de nomes de autores, baseado nos atributos de citações presentes em bibliotecas digitais (nomes dos coautores, título, local de publicação). O objetivo de HHC é agrupar citações em grupos que correspondem ao mesmo autor. Para isso, ele utiliza diversas heurísticas e funções de similaridade sobre os atributos das citações. O agrupamento é dividido em fases, e em cada fase uma heurística diferente é utilizada, fundindo os grupos existentes e provendo mais informações para as fases subsequentes. Desta forma, este método conseguiu obter ótimos resultados em relação a outros que utilizam informações privilegiadas que este não possui.

Apesar das várias soluções propostas decorridas de estudos científicos que tentaram diminuir os erros decorrentes da identificação incorreta dos autores, Tang e Walsh (2010) citam os poucos artigos produzidos entre os anos de 2006 a 2009 na área de cientometria que abordam esse tema, diante disso, asseguram que essa questão não tem sido explorada de forma intensa nas áreas da bibliometria e análises de redes de coautoria. Os autores (2010) também afirmam que devido a problemas envolvidos na desambiguação de nomes de autores, alguns estudos evitam esse processo, enquanto outros não apresentam o método especificando claramente como esse problema foi tratado ou simplesmente mostram os resultados das análises mantendo a identificação dos autores como uma caixa preta.

## **2.2.4 Estruturação das redes de coautorias**

Uma rede de coautoria é composta por três elementos: nós (atores), vínculos (relações ou laços) e fluxos, ou seja, é um conjunto de atores conectados por um conjunto de laços, dessa forma, possuem uma estruturação básica que permite a extração de várias métricas de análise (ANDRADE; DAVID, 2015). Esses elementos podem variar dependendo do propósito da pesquisa, no entanto, os estudos de colaboração científica, em específico os de redes de coautorias, possuem um padrão de nós e arestas que facilitam as análises.

Diante dos vários tipos de redes que existem para estudos de análises e mensuração de informações, em que cada uma delas, os nós e as arestas alteram as informações que representam, é interessante mostrar qual informação esses elementos da rede de coautoria apresentam. Por exemplo, ao contrário das redes de citação, onde os nós são os papéis e as arestas representam as citações de produções científicas, em uma rede de coautoria os nós estão representando os autores, e as arestas entre eles implica em uma colaboração científica. (UDDIN et al., 2012). Bárbasi et al. (2002) também afirmam que os estudos de análise de coautoria estão sendo trabalhados conjuntamente com estudos de formação de redes, onde se interpretam as publicações como relações, e os autores como os nós.

A informação representada pelo nó de uma rede de coautoria pode ser bem flexível dependendo do que o estudo pretende analisar, dessa forma, é comumente associado o termo “ator” para generalizar essa variedade de títulos. Segundo Newman (2000), os nós de uma rede de coautoria podem ser pessoas, equipes, subgrupos, organizações, conceitos, laços, relações comerciais e outras coletividades. Ou seja, ele permite vários níveis de agregação, o que comporta uma adaptação ao problema estudado (OLIVEIRA, 2017).

As arestas estabelecem as relações entre os atores. Em redes de coautorias essa ligação é um trabalho científico produzido pelos pesquisadores Além disso, esses laços que os conectam podem ser direcionados (ou seja, como uma contribuição unidirecional) ou não direcionado (como uma contribuição mútua), que é o mais comum na colaboração científica, também pode ser dicotômico (presente ou ausente, como em se duas pessoas são amigas ou não) ou valorizadas (medido em peso, com uma relação de alta produtividade entre dois autores) (BORGATTI; FOSTER, 2003).

### **2.2.5 Parâmetros**

Para caracterizar uma rede de coautoria são necessários alguns parâmetros matemáticos. Segundo Tomaél (2005), as redes apresentam propriedades que devem fornecer algumas medidas que facilitem a sua análise e entendimento, quais sejam: coesão social, densidade da rede, transitividade, distância geodésica, fluxo máximo e centralidade. Bufrem, Junior e Sorribas (2011) também constataam que as propriedades estatísticas da rede destacam-se, com diferenças nos padrões de colaboração entre os indivíduos estudados, por meio de uma série de medidas de centralidade e conexão.

De acordo com Wasserman e Faust (1994) e Cherven (2015), para estudar e analisar uma rede de coautoria, algumas métricas dos indicadores de rede podem ser usadas, destacando-se:

- Nós: podem ser considerados como elementos individuais em uma rede e representar pessoas, lugares ou objetos que definem coletivamente uma rede;
- Ligações: atuam como as conexões que ligam os nós, podem ser representadas por produções científicas, trabalhos, proximidade e qualquer outro objeto que estabeleça uma relação;
- Grau de centralidade: tem relação com o coautor que mais estabeleceu ligações na rede. De acordo com o grafo, os maiores nós são os que têm maior número de ligações e consequentemente estabelecem um maior número de coautorias diferentes. O grau de centralidade é obtido através da seguinte fórmula:  $C_d(n_i) = d(n_i)/g-1$ . Na fórmula o  $d(n_i)$  representa a proporção de vértices que estão conectados ao nó que desejamos calcular o grau de centralidade, e  $g$  representa o número de vértices total da rede.
- Grau de intermediação: é referente ao coautor que mais coordena as ligações dentro da rede, dessa forma, ele é quem detém o maior contato com os outros coautores e também é o que mais se relaciona entre as diversas áreas de estudo. O grau de intermediação é calculada pela fórmula:  $C_b(n_i) = \sum_{j < k} g_{jk}(n_i) / g_{jk}$ . Para o cálculo da centralidade por intermediação  $g_{jk}(n_i)$  representa o número de caminhos mais curtos que interliga dois atores e que passam pelo ator  $i$ , para o qual se quer calcular o grau de intermediação;
- Modularidade: define o agrupamento de nós e os separa em comunidades (exemplificadas no grafo através das cores) e que determinam a taxa de representatividade dos coautores e quais os temas de estudo mais publicados dentro da rede de coautoria;
- Densidade: inicialmente a densidade é medida em percentual, sendo que um grafo completo tem percentual igual a 1, sendo assim, quando uma rede apresenta uma densidade muito abaixo desse número, é por não existir alta conectividade entre os nós e ligações fortes entre eles, ou seja, existe uma homofilia, pois muitos grupos estão afastados e pouco se relacionam dentro da rede. O cálculo para obtê-la é:  $D = 2 \cdot L / N \cdot (N-1)$ , de forma que  $L$  representa o número de ligações e  $N$  representa o número de nós.

Além disso, vale salientar que os indicadores de centralidade permitem analisar a rede tanto no seu conjunto como individualmente, encontrando diversos resultados como: grau de conectividade da rede, indivíduos com o maior e menor número de interações, intermediação de alguns atores nas relações entre indivíduos e a proximidade entre os indivíduos através das suas interações (ALEJANDRO; NORMAN, 2005).

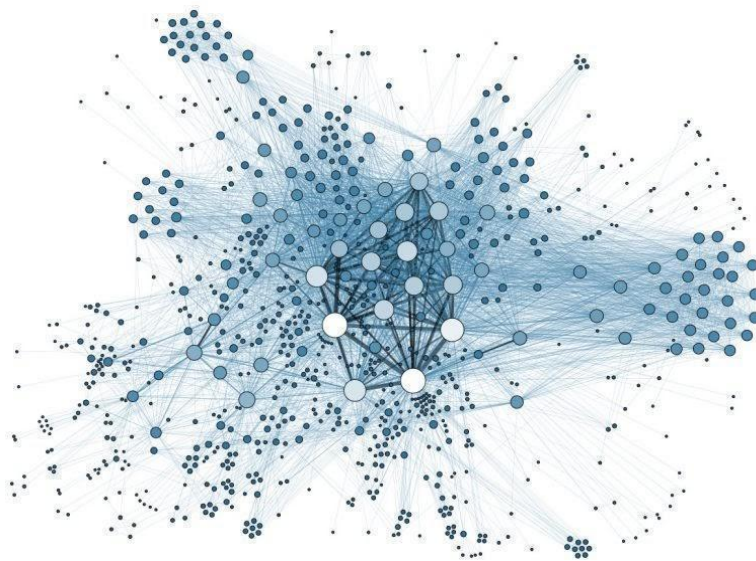
### **2.2.6 Visualização da rede**

Andrade (2016) afirma que as redes de coautoria são um tipo de rede social, logo, sua análise pode ser baseada na premissa de que as relações entre autores da rede podem ser

organizadas através de um grafo. Uma definição mais simples de grafo pode ser descrita da seguinte forma: um grafo  $G = (V, E)$  é um sistema formado por um conjunto  $V$  de elementos chamados vértices, pontos ou nodos, e um conjunto  $E$  de pares não ordenados de vértices chamados de arestas (SIMÕES-PEREIRA, 2013)

Wasserman e Faust (1994) constataam que os grafos têm sido amplamente utilizados na análise de redes de colaboração como uma forma de representar as relações sociais e quantificar as suas propriedades estruturais mais importantes. Também citam que este modelo de visualização beneficia-se da utilização da Teoria dos Grafos que permite a análise das interações sociais, dos padrões e das implicações. A Figura 1 apresenta um grafo utilizado para representar uma rede social.

**Figura 1 - Grafo utilizado na representação de uma rede social.**



**Fonte:** GRANDJEAN (2014)

Um grafo, neste contexto, resulta de uma matriz na qual se introduz previamente a informação de que se dispõe, no entanto, a representação das interações entre os diferentes atores de uma rede faz-se por meio deles, por serem mais compreensíveis e terem uma representação mais agradável para a interpretação (ALEJANDRO; NORMAN, 2005).

### **3 METODOLOGIA**

#### **3.1 MATERIAIS**

Os materiais que foram utilizados no desenvolvimento do trabalho proposto são o *Coletaprod* e *Gephi*, descritos a seguir.

##### **3.1.1 *Coletaprod***

O *Coletaprod* é um software de código aberto, disponibilizado na plataforma *GitHub*, e foi desenvolvido por Tiago Marçal Murakami, bibliotecário da Universidade de São Paulo (USP), com o intuito de facilitar a coleta de dados científicos, através de uma interface simples que fornece um catálogo e filtro para pesquisas de registros.

Segundo Mukami (2020), a ferramenta cataloga os registros de forma individual, como registros de bibliotecas e, com isso, é possível criar um sistema de busca para que o usuário trabalhe com registros delimitados, sendo assim, é possível estabelecer uma comparação com outros repositórios de maneira automática, caso seja necessário. Além disso, o software pode ser utilizado em estudos bibliométricos, já que, ele processa dados de diversas fontes, sendo o *Lattes* a principal, embora também processe *Web of Science*, *Scopus*, *Incites*, *Scival*.

O processo de extração do *Coletaprod* é feito de maneira pragmática. Primeiramente, é necessário baixar o XML do currículo do pesquisador diretamente da Plataforma *Lattes* e, em seguida, submetê-lo à ferramenta. Cada produção do pesquisador vira um registro no banco de dados, que segundo o desenvolvedor (2020), é uma das vantagens em relação a outras ferramentas que fazem a extração de currículos da Plataforma *Lattes*, pois em vez de gerar relatórios em que o pesquisador terá de analisar manualmente, o resultado é quase um catálogo de biblioteca, em que é possível pesquisar nos registros, aprimorar os resultados e realizar filtros.

##### **3.1.2 *Gephi***

A ferramenta *Gephi* é um software desenvolvido em código aberto, usado para exploração e compreensão de redes. Uma arquitetura flexível e multitarefa, que traz novas possibilidades para trabalhar com conjuntos de dados complexos e produzir resultados visuais valiosos. Fornece fácil e amplo acesso aos dados da rede e permite espacializar, filtrar, navegar, manipular e agrupar (BASTIAN; HEYMANN; JACOMY, 2009)

A criação de redes de coautorias ou sociais através de algumas ferramentas requerem um investimento de tempo significativo com uma curva de aprendizagem acentuada, o que o

torna um desafio para muitos criadores ou pesquisadores. No entanto, o *Gephi* ajuda a superar essas barreiras, fornecendo uma poderosa ferramenta, mas que seja de fácil uso, possibilitando ao pesquisador gastar mais tempo na criação e implantação de grafos, e menos tempo codificando. Além disso, através dos conjuntos de dados fornecidos, o software irá auxiliar rapidamente no desenvolvimento, personalização e implantação de interessantes grafos (CHERVEN, 2013).

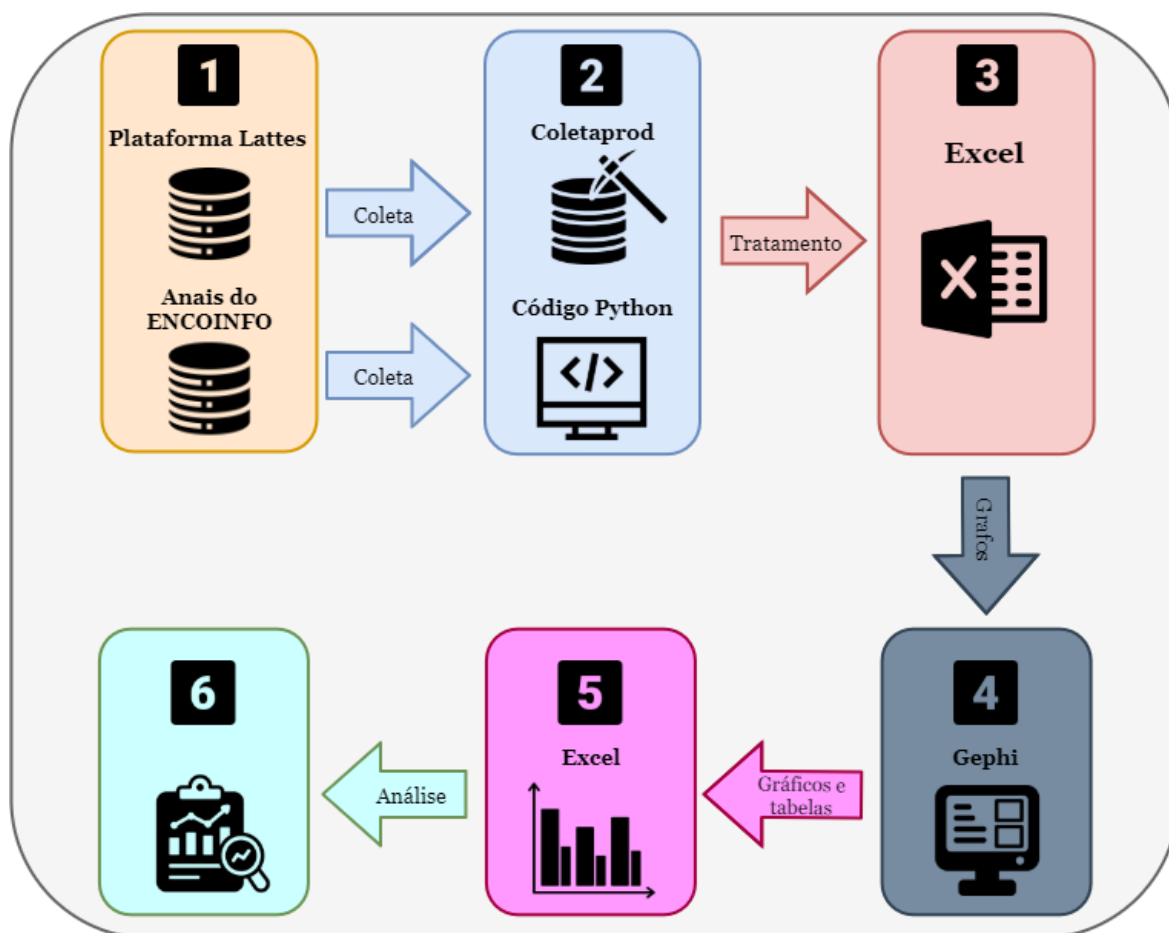
Segundo Khokhar (2015), a capacidade de manipular dados com uma boa experiência do usuário é um recurso fundamental para qualquer rede ferramenta de análise e visualização, pois o usuário tem de ter a flexibilidade de modificar a rede de dados facilmente, sejam nós e seus atributos, arestas e seus atributos. Dessa forma, a autora afirma que o *Gephi* fornece a seus usuários uma interface muito simples e fácil de usar com a qual podem manipular os dados em uma forma tabular. Por conseguinte, os dados podem ser inseridos diretamente na ferramenta, ou também podem ser importados em formato CSV, sempre seguindo um modelo de matriz.

O software usa um mecanismo de renderização 3D, que possibilita exibir grandes redes em tempo real e agilizar a exploração. Também oferece inúmeras métricas para análise de redes, como: grau, diâmetro, densidade, coeficiente de *cluster*, modularidade, componentes conectados, excentricidade e métricas de centralidade (ANDRADE, 2016).

### 3.2 MÉTODOS

No âmbito dos currículos dos atuais e ex-professores dos cursos já citados, disponibilizados na Plataforma *Lattes*, juntamente com os anais do ENCOINFO, foi definido como principal *corpus* desta pesquisa, as publicações que estão relacionadas a essas bases de dados e com o plano de ensino do CEULP/ULBRA, abrangendo o período de 2015 ao ano de 2020. Na Figura 2 é demonstrado como seguirá o fluxo de trabalho para o desenvolvimento deste projeto.

Figura 2 - Fluxograma dos procedimentos metodológicos do trabalho.



Fonte: Autor

Na etapa 1 têm-se as bases de dados de onde serão retirados os dados do trabalho. Em seguida, na etapa 2 ocorre a coleta dos dados dos pesquisadores diretamente dos anais que estão disponíveis de forma documental eletrônica no site do ENCOINFO, através de um código em linguagem *Python*, que extrai os nomes de todos os autores. Também de forma paralela, é realizada a extração de todos os currículos dos professores da Plataforma *Lattes*, com a ferramenta *Coletaprod*, sendo assim, deve ser feito o download de todos os currículos dos docentes em formato XML diretamente da Plataforma *Lattes*, em seguida, serão inseridos no software a partir de uma das suas funcionalidades e extraídos de acordo com a necessidade da pesquisa.

Na etapa 3 é realizado o tratamento dos dados coletados utilizando o software *Microsoft Excel*. Todo o processo deve ser executado de acordo com o modelo de importação



da ferramenta *Gephi*, portanto, os dados têm de estar formatados em matrizes e correlacionados em duas planilhas.

Na etapa 4 serão gerados os grafos e medidas com base nos resultados obtidos por meio do software *Gephi*. Dentre os parâmetros gerados serão utilizados: o Grau de centralização, Modularidade, Grau de intermediação e Densidade. Por conseguinte, na etapa 5 ocorrerá a criação dos gráficos e tabelas estatísticas com a utilização do software *Microsoft Excel*, tendo como base o grafo de rede coautoria gerado pela ferramenta *Gephi*. Dentre os possíveis dados estatísticos encontrados, será utilizado para análise os parâmetros: número de artigos, número de coautores, número de ligações e coautores com mais ligações. Por fim, na etapa 6 têm-se a caracterização e interpretação dos resultados obtidos através dos grafos da rede de coautoria, gráficos e tabelas estatísticas.

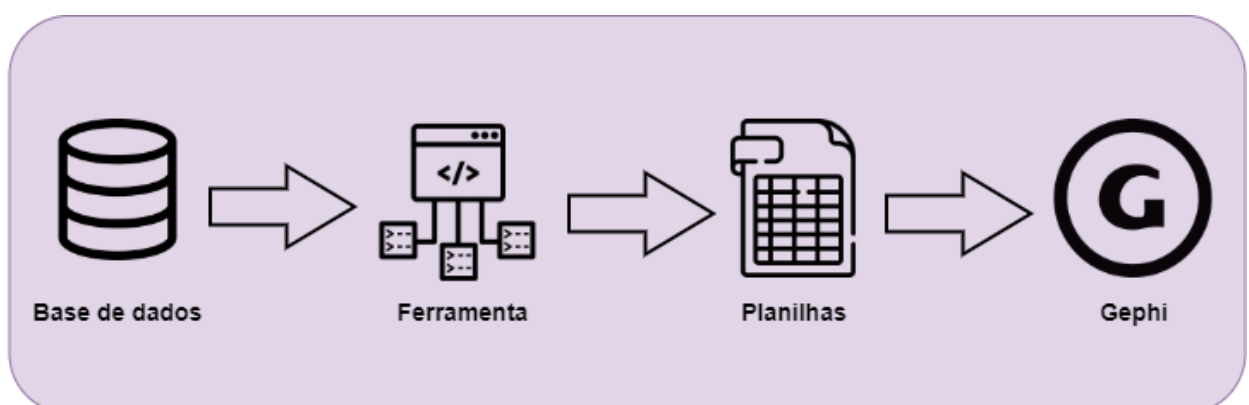
## 4 RESULTADOS E DISCUSSÃO

Nesta seção é descrito o processo de execução de cada etapa da solução proposta. Nesse contexto, ressalta-se que foram extraídos, tratados e analisados todos os dados que formam a rede de coautoria dos cursos de Engenharia de Software, Ciências da Computação e Sistemas da informação, do CEULP/ULBRA e foram utilizadas as tecnologias *Coletaprod*, *Gephi* e *Excel* para auxiliar no desenvolvimento do trabalho. Assim, são descritos os resultados obtidos durante a execução do projeto.

### 4.1 COLETA DOS DADOS

A coleta de dados foi realizada em dois momentos: um para a base de dados da Plataforma Lattes, e outra para o ENCOINFO. Foi realizado dessa maneira, pois existem algumas ferramentas *open-source* disponíveis para a extração dos dados da Plataforma Lattes, enquanto o ENCOINFO, por ser um repositório de publicações, não possui ferramenta de coleta já desenvolvida para este uso, sendo necessária a implementação de um código específico para realizar o processo. Visto que ambos os processos seguem um mesmo padrão, se diferenciando somente na ferramenta utilizada, na Figura 3, é mostrado de forma geral o fluxograma desta etapa. Seguindo da base de dados, para a ferramenta utilizada na extração das informações, a criação das planilhas contendo as matrizes, e por fim, a importação das planilhas para o *Gephi*.

**Figura 3 - Coleta de dados**



Importante destacar que a formatação das planilhas deve seguir o padrão de inserção da ferramenta *Gephi*, pois é a partir das matrizes criadas que serão produzidos os grafos. Sendo assim, a planilha de nós, deve conter os nomes, sem repetições, e colocados na primeira coluna Label, e na segunda coluna foi associado um ID a cada autor. Já a segunda planilha é relacionada às arestas, por conseguinte, traz algumas diferenças se comparada a de nós em decorrência das relações em si.

#### 4.1.1 Coleta dos dados da Plataforma Lattes

O *Coletaprod* foi a ferramenta escolhida para extrair os dados da Plataforma Lattes. Primeiramente, para poder utilizá-lo é necessário baixar os currículos dos professores e ex-professores dos cursos citados, diretamente da Plataforma Lattes, em formato .xml e inserir no *Coletaprod*, através de seu template, conforme mostrado na Figura 4.

**Figura 4 - Template de inserção dos currículos em formato .xml**

XML Lattes	Escolha o ar	Browse	TAG	Identificador	Unidade	Tipo de vínculo	Incluir

Em seguida, é possível visualizar os currículos inseridos na ferramenta, como mostra a Figura 5. Dessa forma, é necessário acessar cada documento separadamente, onde têm-se os artigos científicos de cada professor. A partir disso, a ferramenta disponibiliza alguns filtros que facilitam a busca por periódicos específicos, autores e ano de publicação. No entanto, para ser extraído o título de cada artigo e seus respectivos autores, foi necessário realizar um processo manual.

**Figura 5 - Visualização dos currículos inseridos na ferramenta *Coletaprod*.**

Unidade	Tipo de material	Fonte	Alguns números
<b>Tags</b> (293) Fabiano (182) Jackson (88) Edeilson (66) Cristina (57) Madianita (53) Heloise (52) Fernanda (17) Fabio (1)	Trabalhos em eventos (621) Textos em jornais de notícias/revistas (80) Software (48) Livro publicado ou organizado (23) Capítulo de livro publicado (20) Artigo publicado (17)	Base Lattes (809)	809 registros 10 currículos 0 registros na fonte

A partir do momento em que os nomes vão sendo extraídos, foram criadas manualmente duas planilhas no *Excel*, sendo uma para os nós, e outra para as arestas, conforme a Figura 6 e Figura 7, respectivamente.

**Figura 6 - Planilha dos nós**

Label	ID
Fabiano Fagundes	1
Kennedy Santos Torres	2

A planilha de nós contém o nome do autor na primeira coluna e na segunda o seu respectivo ID.

**Figura 7 - Planilha das arestas**

Source	Target
1	2
1	3
1	4

A planilha de arestas utiliza dois IDs para fazer as relações entre os coautores. Na primeira coluna é o ID do coautor de origem e na segunda coluna é o ID do coautor de destino.

#### **4.1.2 Coleta dos dados do ENCOINFO**

Para a coleta dos artigos do ENCOINFO foi desenvolvido um código Python. Primeiramente, foi necessário fazer o *download* da página html de cada ano do evento, disponível no endereço <https://ulbra-to.br/encoinfo/edicoes/>. A partir disso, o código na Figura 8 inicia com a leitura do arquivo html e faz a procura da string ‘titulo-link’, para que ache o início dos títulos dos artigos. Em seguida tem um condicional que fará a verificação dos autores deste artigo.

**Figura 8 - Código para ler o arquivo html.**

```

1  import io
2  arq = open('18.html','r', encoding='utf8')
3  conteudo = arq.read()
4  arq.close()
5
6  inicioTitulo = conteudo.find('titulo-link')+30
7  lista = []
8
9  if inicioTitulo != '': temAutores = True

```

A segunda etapa do código irá fazer uma verificação em *loop* que pegará o nome dos autores e armazenará em uma lista, concatenando-os posteriormente ao nome do artigo a que eles estão relacionados, como mostrado na Figura 9. Após, será criado um arquivo .csv com estas relações de autores e seus respectivos artigos.

**Figura 9 - Segunda etapa do código.**

```

11 while temAutores:
12
13     fimTitulo = conteudo.find('/h6',inicioTitulo)-27
14     linha = conteudo[inicioTitulo:fimTitulo]+';'
15
16     inicioAutor = conteudo.find('line-content',fimTitulo)
17     inicioAutor = conteudo.find('/span',inicioAutor)+6
18     fimAutor = conteudo.find('span',inicioAutor)-1
19
20     autores = conteudo[inicioAutor:fimAutor]
21     autores=autores.replace(' ',';')
22     listaDeAutores = autores.split(',')
23
24     for i in listaDeAutores:
25         linha += i+';'
26     lista.append(linha)
27
28     arq = open('autores2.csv','w')
29     for i in lista:
30         lista2 = i.split(',')
31         for j in lista2:
32             arq.write(j+';')
33         arq.write('\n')
34
35     temAutores = False
36     inicioTitulo = conteudo.find('titulo-link',fimTitulo)+30
37     if inicioTitulo != '' and inicioTitulo>fimTitulo: temAutores = True
38
39     arq.close()

```

A planilha com os autores e artigos coletados podem ser vistos na Figura 10 que apresenta o arquivo .csv que foi gerado.

**Figura 10 - Arquivo .csv com autores e artigos coletados.**

1	Adequação e testagem do IMP	Leomar Camargo de Souza	Pierre Soares Brandão	Fernando Luiz de Oliveira	Fabiano Fagundes			
2	Aplicação das Boas Práticas da U.S. S. Barros		P. F. Brito					
3	Aplicação de ferramenta OLAP	Ranyelson Neres Carvalho	Madianita Bogo Marioti	Walkiria Regis de Medeiros				
4	Avaliação dos Processos de Dest	Felipe Oliveira Simões	Tathinay de Souza Siqueira	Brunna Caroline do Carmo	Thatiane de Oliveira Rosa	André Mesquita Rincon		
5	CODE LIVE: Gamificação de um	Djonathas C. Cardoso	Jackson G. de Sousa	Fabiano Fagundes	Parcilene F. de Brito			
6	Estudo da segmentação do ICM	Leomar Camargo de Souza	Haroldo Fernando Fritsch	Fernando Luiz de Oliveira				
7	Modelo de inferência de locais	Pedro Henrique G. Camargo	Fernando L. de Oliveira	Felipe Dias da S. Cabral	João B. B. Junior	Macon J. Fristh		
8	Modelo de Inferência de Preferê	Douglas J. Ramalho Araujo	Jackson G. Souza	Parcilene F. de Brito	Edeilson M. da Silva			
9	Modelo de visualização de Infor	Lucas Henrique Roese	Parcilene Fernandes de Brito					
10	ORDENA – um Jogo Educacional	Haroldo F. Fritsch	Edeilson M. da Silva	Jackson G. de Souza				
11	Proposta de Framework para Us	Márcia M. Savoine	Joyce Karoline M. Matias	Vanessa F. Rocha	André M. Rincon	André Magno C. Araújo		
12	Sistema de Avaliação de Usabilic	Fernando Lúcio	Guilherme Reinaldo	Matheus Gregorio	Miguel Lemos	Rodrigo Melo		
13	Utilização do Algoritmo de Máq	Kevin M. Araújo	Edeilson M. da Silva					
14	A Importância da Auditoria de TI	Fabricio da Silva Rodrigues	Brunna Caroline do Carmo	Thatiane de Oliveira Rosa				
15	Aplicação da técnica de Text	Mir Felipe E. B. Schmitz	Parcilene F. de Brito					

Assim como na coleta dos dados da Plataforma Lattes, a partir do momento em que os nomes vão sendo extraídos, são criadas duas planilhas no *Excel*, sendo uma para os nós, e outra para as arestas, seguindo o padrão de inserção da ferramenta *Gephi* conforme descrito anteriormente.

#### 4.1.3 Definindo o limite da rede

O limite da rede tem como propósito determinar quais os dados necessários e adequados para a pesquisa se tornar coesa. Diante disso, a definição dos limites da rede foi realizada de forma que acrescente mais relevância aos possíveis resultados encontrados. Foi delimitado que a rede de coautoria, a partir dos dados do ENCOINFO, abrangeria as publicações científicas vinculadas ao evento no intervalo que vai do ano de 2015 ao ano de 2020.

Além disso, também foi definido que o limite da rede de coautoria, proveniente dos dados da Plataforma Lattes, se daria a partir dos dados extraídos dos currículos dos atuais e ex-professores dos cursos já citados tendo, como foco, as publicações encontradas que estejam dentro do intervalo que vai do ano de 2015 ao ano de 2020 e que estão relacionadas aos eventos científicos que fazem parte do calendário científico do CEULP/ULBRA, sendo assim, contendo os eventos como o ENCOINFO e Jornada de Iniciação Científica.

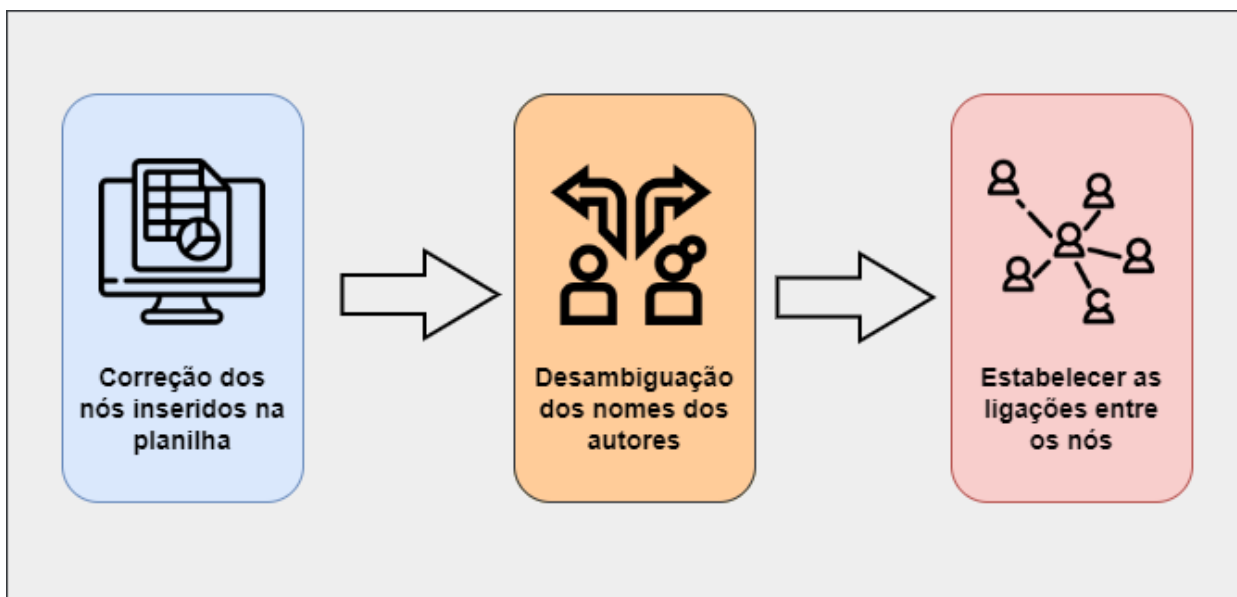
Essas limitações das redes de coautoria visam uma maior agilidade no processo de caracterização da rede, já que, é possível diminuir o escopo e também tirar um maior proveito para as estatísticas de possíveis resultados encontrados, pois os dados se tornarão menos abstratos, ou seja, serão menos abrangentes e genéricos, sendo que a rede de coautoria das bases de dados escolhidas contém poucas ligações entre os nós, então quanto maior o escopo, mais espalhados estarão os nós e menores serão as métricas de caracterização da rede.

#### 4.2 TRATAMENTO DOS DADOS

O tratamento dos dados ocorre a partir da inserção dos dados nas planilhas anteriormente citadas, criadas na ferramenta *Excel*. Na Figura 11 é apresentado o fluxograma desta etapa de

tratamento dos dados, que se inicia com a correção dos nomes dos autores e seus respectivos ID na planilha de nós, segue para a desambiguação dos nomes dos autores ainda na planilha de nós e, a partir disto, são estabelecidas as ligações entre os nós de maneira corretiva, onde é verificado se algum nó está ligado a outro erroneamente, a fim de evitar erros.

**Figura 11 - Tratamento de dados**



Primeiramente, todos os nomes dos autores têm um ID relacionado, conforme exemplificado na Figura 12. Esse processo serve para relacionar os autores entre si e criar as matrizes da rede de coautoria.

**Figura 12 - Planilha de nós com nome dos autores e ID.**

Label	ID
Fabiano Fagundes	1
Kennedy Santos Torres	2
Dennis Dyodi Kawakami	3
Heloise Acco Tives Leão	4
Madianita Bogo Marioti	5
Fernanda Pereira Gomes	6
Ana Clara da C. M. da Silva	8
Fernando L. de Oliveira	9
Lucas Ribeiro Reis de Sousa	10
Aurélio Miguel Machado da Silva	11
Jackson Gomes de Souza	12
Rodrigo V. F. de Souza	13
Anilton A. F. Maciel	14
Sergio A. A. de Freitas	15
Edna D. Canedo	16
Alexandre Moraes Matos	17
Dionnys Santos Marinho	18
Parcilene Fernandes Brito	19
Luan Gomes de Almeida Araújo	20
Matheus R. Leal	21
Taylor Santos Oliveira	22

Nessa etapa também ocorre um processo de desambiguação do nome dos autores, visto que, ao realizar a coleta das bases de dados, é comum que tenha um mesmo autor referenciado de várias maneiras, em vários artigos diferentes. Por exemplo, o professor Fabiano Fagundes, é referido como “F. Fagundes”, “Fabiano F.” e “Fabiano Fagundes”. Desta forma, para minimizar possíveis problemas como esse, é realizado o procedimento manual de desambiguação.

Assim, para realização deste processo, é selecionado somente um nome que o professor possa receber, para que ele seja inserido na planilha de maneira correta. Seguindo o exemplo anterior, o nome Fabiano Fagundes é escolhido para preencher as planilhas e, a partir disso, é realizada uma revisão dos nomes inseridos na planilha de nós com a busca de outros nomes com que, porventura, o professor possa ter sido referido. Caso seja encontrada alguma incongruência, é retirado esse nome, e seu ID é passado para o nome anterior ao dele, desta forma evitando alterações nos resultados e trabalhos repetitivos, visto que, será necessário refazer as matrizes de arestas, pois o mesmo autor estará presente na rede com dois nomes diferentes.

Na segunda planilha, a de arestas, exemplificada na Figura 13, também foi necessária uma atenção especial, pois é nela que os relacionamentos entre os autores serão definidos. Todo o procedimento é feito de forma manual, sendo assim, foi preciso uma cautela ao inserir os dados, para que se evite erros. No entanto, caso algum erro de relacionamento com alguma situação mais explícita seja realizado, a ferramenta *Gephi* mostrará no momento em que os grafos são



criados, como, por exemplo, ao deixar um autor sem produção ou o um autor ficar sem ligações.

Na Figura 13 é possível visualizar as duas colunas que fazem as matrizes da rede de coautoria: colunas Origem (*source*) e Destino (*target*). Para esse preenchimento, é preciso seguir os IDs da planilha de nós, portanto, cada autor se interliga na planilha de arestas, através de sua identificação.

**Figura 13 - Planilha de arestas com seus autores de Origem e Destino.**

Source	Target
1	2
1	3
1	4
1	5
1	17
1	11
1	25
1	4
4	3
4	5
4	13
4	14
4	15
4	16
4	17
4	10
4	25

Por conseguinte, o autor que tem dois coautores em um artigo tem estabelecidas duas ligações diretas com seus colaboradores. Por exemplo, o autor Fabiano Fagundes que é o ID 1 tem artigo com a Heloíse Acco Tives Leão que é o ID 4, e com o Kennedy Santos Torres que é o ID 2, logo, são realizadas duas ligações entre os três autores, sendo a Origem (Fabiano) e Destino os outros dois, Heloíse e Kennedy. No entanto, é necessário ressaltar que este relacionamento por artigo é feito somente uma vez, como mostrado anteriormente, na Figura 13, por exemplo: o autor de ID 1 está ligado ao autor de ID 4, não sendo necessário então estabelecer o vínculo do autor de ID 4 com o autor de ID 1.

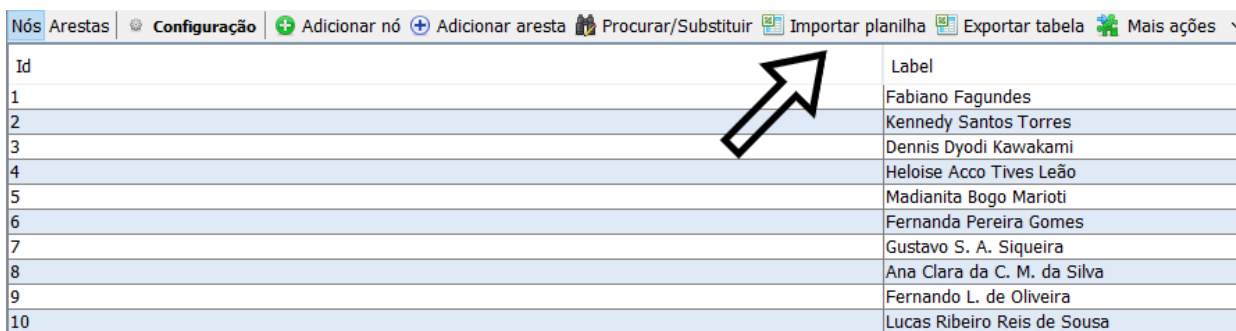
### 4.3 GRAFOS COM O GEPHI

Nesta seção são apresentados os procedimentos necessários para a produção dos grafos que representam as redes de coautorias das bases de dados coletadas. É mostrada a inserção dos dados na ferramenta *Gephi*, seguindo toda a padronização e processo. Além disso, são evidenciados os algoritmos de visualização aos quais os grafos foram submetidos, para que eles pudessem ter uma disposição mais compreensível para a extração de métricas.

#### 4.3.1 Inserção dos dados

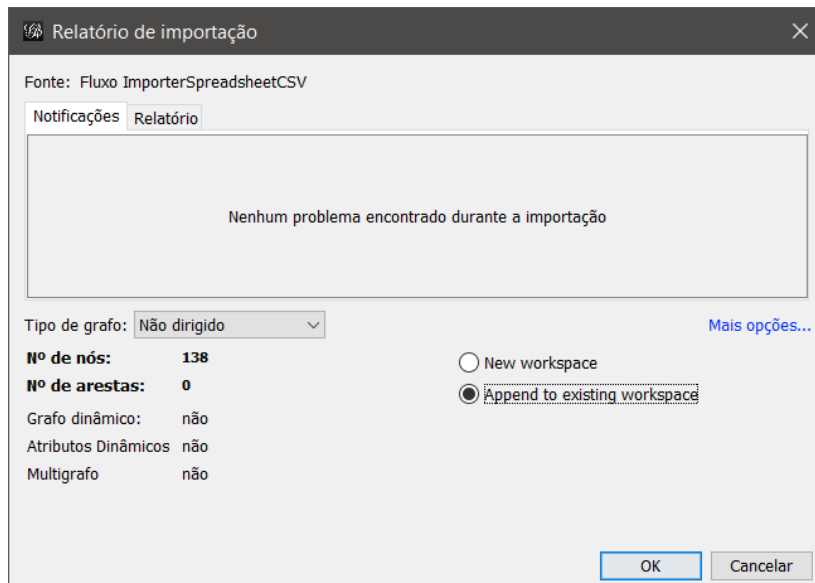
Após serem preenchidas as planilhas de nós e arestas na ferramenta *Excel*, foi iniciada a importação delas no *Gephi*. Dentro da ferramenta, foi criado um novo projeto, em seguida, no template de inserção, as planilhas são importadas. Na Figura 15, é possível visualizar o local de importação da planilha na barra superior, além disso, também são mostrados todos os nós e arestas, igualmente como no *Excel*. Também é válido destacar que o grafo a ser criado é não-dirigido, isso é apresentado no momento da importação da planilha de nós, como na Figura 16, pois, todas as arestas são bidirecionais.

**Figura 15 - Importação dos nós na ferramenta *Gephi*.**



Id	Label
1	Fabiano Fagundes
2	Kennedy Santos Torres
3	Dennis Dyodi Kawakami
4	Heloise Acco Tives Leão
5	Madianita Bogo Marioti
6	Fernanda Pereira Gomes
7	Gustavo S. A. Siqueira
8	Ana Clara da C. M. da Silva
9	Fernando L. de Oliveira
10	Lucas Ribeiro Reis de Sousa

**Figura 16 - Seleção de grafo dirigido e não dirigido.**



Em seguida, as arestas serão inseridas da mesma forma que o nós, como mostra na Figura 17. Nesta tela, pode-se visualizar o ID de origem e o ID de destino das ligações, dessa forma, são facilmente identificados os possíveis erros com o filtro disponível. Este filtro torna possível detectar e mesclar nós duplicados, pesquisa por grau de centralidade, por ID, por nome e ligações.

**Figura 17 - Inserção das arestas na ferramenta Gephi.**

Origem	Destino	Tipo	Id
15	17	Não dirigido	1
15	1	Não dirigido	2
15	3	Não dirigido	3
15	11	Não dirigido	4
24	1	Não dirigido	11
24	3	Não dirigido	12
24	11	Não dirigido	13
19	20	Não dirigido	14
19	1	Não dirigido	15
19	2	Não dirigido	16

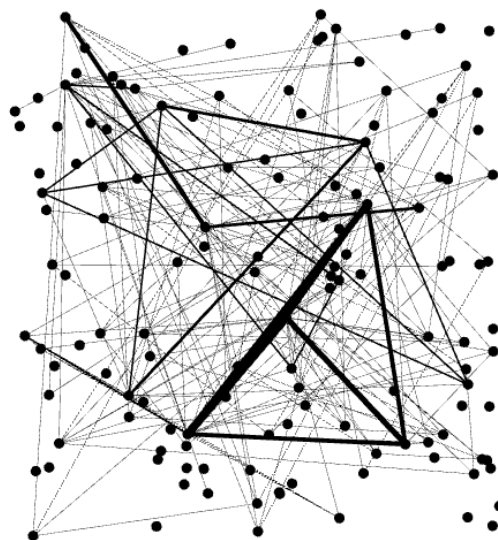
Também é apresentado o peso das ligações entre os vértices, na Figura 18, sendo assim, é possível verificar os autores que mais contribuem para uma rede homofílica.

**Figura 18 - Peso das ligações entre os nós.**

Weight
4.0
5.0
4.0
7.0
6.0
1.0
3.0
3.0
2.0
1.0
1.0

Por conseguinte, o grafo inicial da rede de coautoria já pode ser visualizado, como mostra a Figura 19.

**Figura 19 - Grafo inicial da rede de coautoria.**

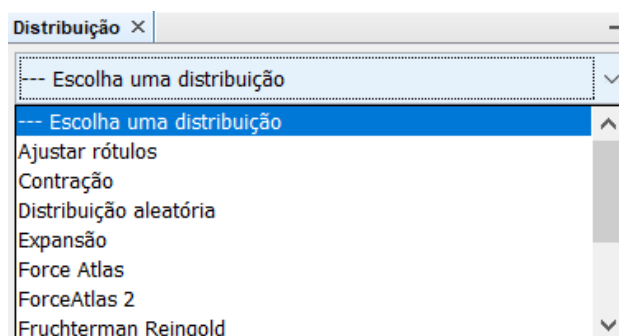


No entanto, devido ao volume de dados, a falta de visualização adequada e também uma disposição desfavorável para análise, o grafo necessita de algumas configurações que permitam uma representação mais legível. Portanto, são adicionados os nomes dos nós, cores por modularidade, aumento de nós e arestas de acordo com a produção e peso de arestas, e por fim, algoritmos de visualização

#### **4.3.2 Algoritmos de visualização**

Um *layout* adequado é de extrema importância para determinar se o grafo será decifrável e contará uma história ou se somente mostrará um emaranhado de dados sem definição. Com o *Gephi* é possível testar alguns algoritmos de visualização, como mostra na Figura 20, que irão modelar a rede para que ela tenha uma estrutura mais estética e limpa, sendo assim, demonstrando os grupos da rede e suas ligações.

**Figura 20 - Algoritmos de layout do Gephi.**

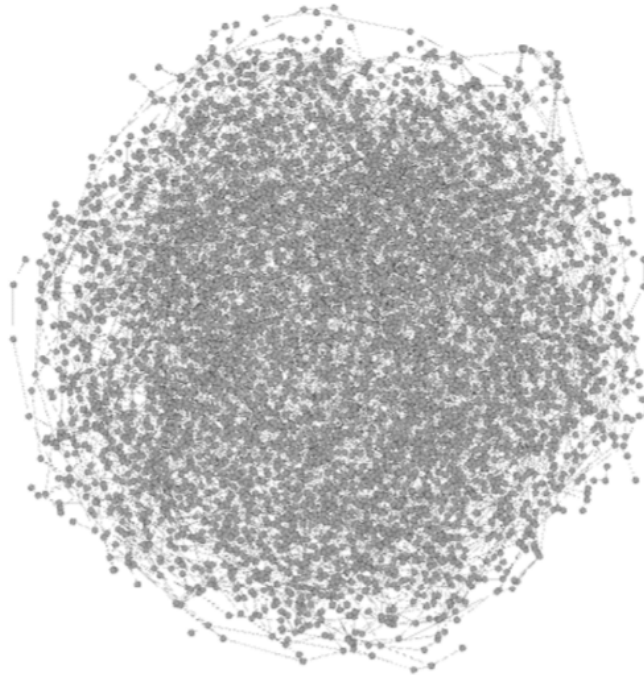


Muitos dos algoritmos de *layout* fornecem opções que permitem ao usuário determinar o espaçamento ideal dentro de um grafo, podendo mexer com atração, repulsão, gravidade e outras configurações disponíveis (CHERVEN, 2015). Dessa forma, é importante selecionar os algoritmos que melhor se encaixam na rede, por conseguinte, os algoritmos Force Atlas e Fruchterman-Reingold.

#### **4.3.2.1 Force Atlas 2**

O Force Atlas 2 é uma versão mais rápida e atualizada de outro algoritmo, o Force Atlas. Ele substitui a atração e configurações de repulsão com uma única configuração de escala, que permite aos usuários definir um nível de repulsão que espalha o grafo para melhor legibilidade (CHERVEN, 2015). Ele tende a deixar os nós em uma disposição que favoreça a visualização dos grupos científicos da rede, podendo alterar o espaçamento entre eles e aproximação conforme a ligação dos coautores.

**Figura 21 - Algoritmo Force Atlas 2 na rede.**



Dentro do *Gephi* é possível fazer a configuração do algoritmo em algumas etapas, desta forma, os grafos desse projeto seguiram essa padronização, conforme apresentado na Figura 22.

**Figura 22 - Tela de configuração do algoritmo Force Atlas 2 no *Gephi*.**

ForceAtlas 2	
▶ Executar	
Threads	
Número de threads	3
Alternativas de comportamento	
Dissuadir hubs	<input type="checkbox"/>
Modo LinLog	<input type="checkbox"/>
Evitar sobreposição	<input type="checkbox"/>
Influência do peso das arestas	1.0
Afinação	
Dimensionamento	10.0
Gravidade mais forte	<input type="checkbox"/>
Gravidade	1.0
Desempenho	
Tolerância (velocidade)	1.0
Aproximar repulsão	<input type="checkbox"/>
Aproximação	1.2

Primeiramente, as *threads* vem no padrão 2, no entanto, podem ser modificadas de acordo com os recursos da máquina, sendo assim, irá acelerar o processo de executar o *layout*. Por conseguinte, as Alternativas de Comportamento determinam como o grafo se manterá em relação aos nós, podendo alterar a aproximação dos mesmos com a opção de “Dissuadir” “*hubs*”, e também evitar sobreposições de nós maiores, sobre os menores, com a opção de

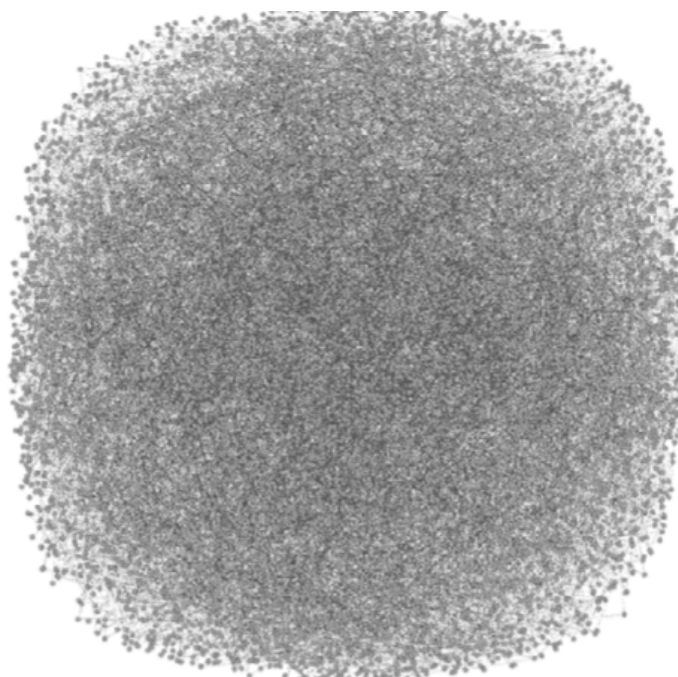
“Evitar sobreposição”.

Por fim, as configurações de “Escala e Gravidade” foram escolhidas para que os grafos se mantenham espaçados de maneira legível. Já que, quanto maior a gravidade, mais próximos os nós ficarão. Como isso vai depender também da densidade da rede, é necessário ir realizando testes até chegar ao ideal.

#### 4.3.2.2 Fruchterman-Reingold

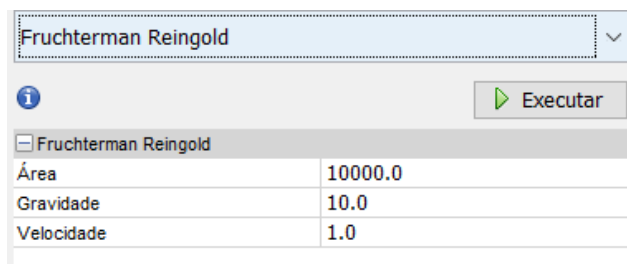
Em seguida tem-se o *layout* Fruchterman-Reingold, que também é outra abordagem baseada em força. É possível ajustar as configurações para este algoritmo, embora não com o mesmo grau de opções como no modelo Force Atlas 2. Os principais ajustes que se pode fazer aqui envolvem a área do tamanho do grafo e a gravidade. Assim, uma rede densa pode ser forçada a espalhar manipulando a área do grafo, em vez de ajustar a repulsão ou configurações de atração (CHERVEN, 2015).

**Figura 23 - Algoritmo Fruchterman-Reingold na rede.**



Esse *layout* foi escolhido para estabelecer uma maior estética ao grafo, já que ele agrupa os nós sem deixar eles soltos nas pontas dos grupos que vão se formando, dessa forma, ele foi aplicado após o Force Atlas 2. Na Figura 24, são apresentadas as configurações padrões do algoritmo, da forma como foram submetidas a todos os grafos desse projeto

**Figura 24 - Tela de configuração do algoritmo Fruchterman-Reingold no Gephi.**



Esse modelo é mais simples com as configurações, colocando mais responsabilidade no próprio algoritmo do que no usuário, tendo a função de “Área” que irá fazer a repulsa e aproximação da rede, tornando-a mais espalhada ou conjunta. Também é possível modificar a “Gravidade”, da mesma forma como no Force Atlas 2, e por fim, tem-se a “Velocidade” que irá acelerar o processamento do algoritmo, de acordo com os recursos da máquina local.

#### 4. 4 GRAFOS E PARÂMETROS DAS REDES DE COAUTORIAS DO PROJETO

Nesta seção os grafos e parâmetros das redes de coautorias foram separados de acordo com a sua base de dados, sendo a primeira da Plataforma Lattes e a segunda do ENCOINFO.

##### 4.4.3 Grafos e parâmetros da rede de coautoria a partir da Plataforma Lattes

A Figura 25 representa o grafo da rede de coautoria do *corpus* deste projeto. Ele foi analisado segundo os indicadores de densidade, grau de centralidade, grau de intermediação e modularidade.

Foi obtida uma densidade de 0.057, a partir da fórmula inicial  $D = 2 \cdot L/N$ . Devido a essa baixa numeração, é verificável uma rede com pouca conectividade e muitas ligações fracas entre os nós. Essa condição é evidenciada no grafo através do grande número de nós pequenos e afastados do centro da rede. Além disso, com exceção dos coautores mais ao centro da rede, as ligações são predominantemente tênues, caracterizando essas coautorias com poucas produções científicas e escassas ligações entre autores.

Os professores dos cursos de Ciência da Computação, Sistemas de Informação e Engenharia de Software do CEULP/ULBRA destacam-se na rede como coautores assíduos por estarem presentes em quase todos os artigos, sendo assim, eles detêm os maiores graus de centralização, obtido a partir do cálculo  $Cd(n_i) = d(n_i)/g-1$ . A exemplo, os 4 primeiros professores possuem grau acima de 0.25, como: Fabiano Fagundes com 0.44, Heloise Acco com 0.34, Jackson Gomes com 0.29 e Fernando Luiz com 0.28. Além disso, alguns professores externos aos cursos, como o Pierre Brandão, também recebem bastante destaque na rede, com um grau de 0.21, acima da média geral de menos de 0.02.

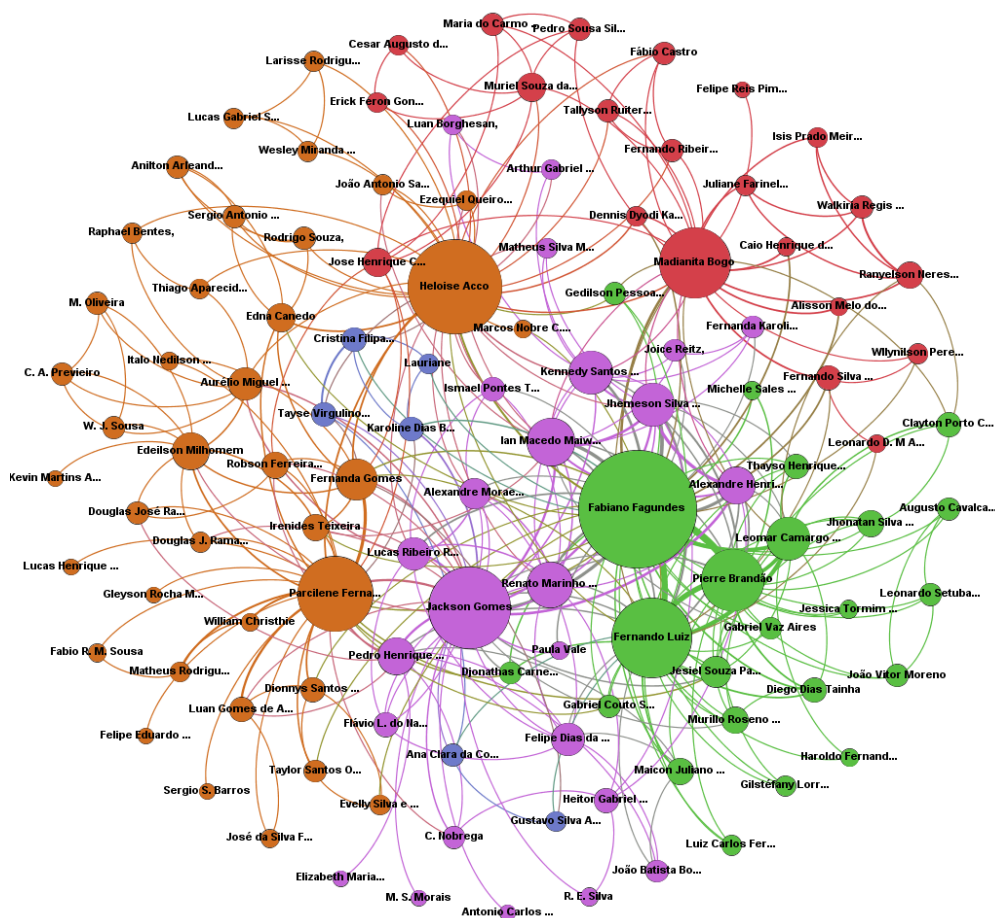


Consequentemente, os nós intermediários também são os docentes, pois além de estarem mais ao centro, também possuem maior nível de diversificação do campo de estudo, o que gera coautorias distintas e mais ligações na rede. Já os nós mais afastados, tendem a formar conectividades menores, pois permanecem dentro de grupos da mesma intenção de pesquisa científica. O grau foi obtido a partir do cálculo demonstrado anteriormente,  $C_b(n_i) = \sum_{j < k} g_{jk}(n_i) / g_{jk}$ , e com isso normalizou-se os valores máximos como 1. Portanto, os 3 primeiros coautores com mais importância para a rede, são: Fabiano Fagundes com 0.29, Heloise Acco com 0.28 e Parcilene Fernandes com 0.21.

Foi identificado, de acordo com a modularidade, um total de 5 comunidades, identificadas pelas cores no grafo e definidas pela proximidade de seus vizinhos e a densidade de ligação entre eles, desta forma, quanto mais espessa a aresta no grafo, maior também a ligação entre os nós, ou seja, os coautores possuem mais artigos publicados juntos. A partir disso, pode-se evidenciar o nó centralizador de cada grupo, responsável pelos maiores números de coautorias e diversificação de estudos científicos, além disso, tanto a comunidade que ele se insere, como também a rede, são amplamente influenciadas pelas suas produções, visto que, resulta no aumento de ligações e conectividade.

Vale destacar que a rede é dominada, com mais de 75% de tamanho, pelas comunidades laranja, roxo e verde, sendo assim, existe um grande fluxo de produção científica e coautoria entre elas. Por outro lado, a pequena porcentagem restante é composta pela vermelha e azul, sendo essa última extremamente pequena em relação às outras comunidades, contendo apenas 6 nós e poucas ligações entre eles.

**Figura 25 - Grafo da rede de coautoria segundo os artigos publicados pelos professores na Plataforma Lattes nos anos de 2015 a 2020.**



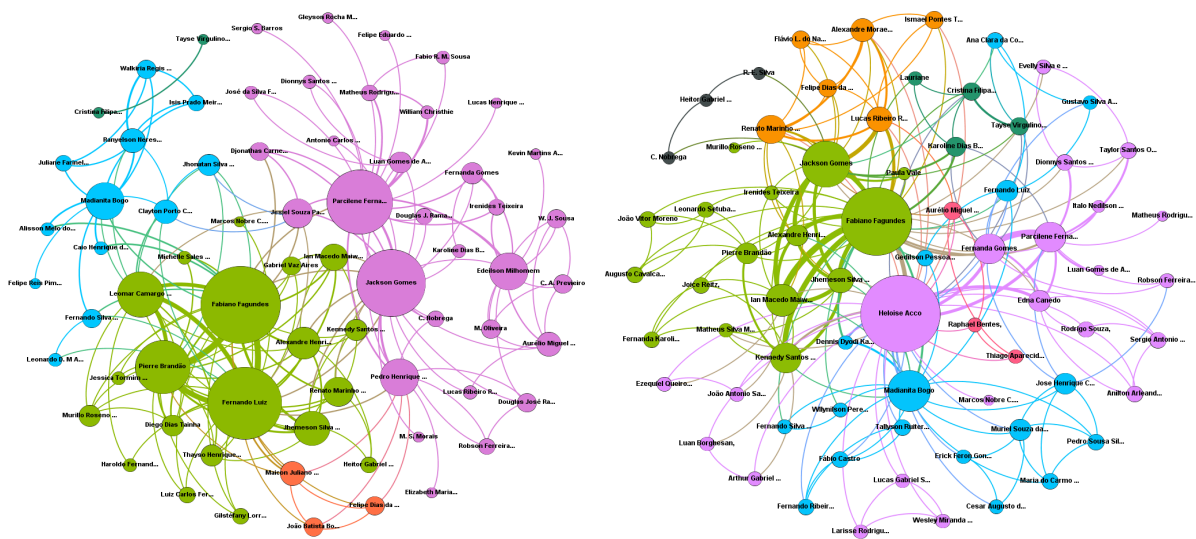
Optou-se também por verificar a evolução da rede de coautoria dentro do *corpus* desta pesquisa. A Figura 26 mostra 2 grafos, A e B, representando, respectivamente, as redes de coautoria dos anos de 2015/16/17 e 2018/19/20.

Diante disso, foi possível obter alguns resultados comparativos entre os dois grafos. No grafo A, tem-se um total de 69 nós e 182 ligações, e os graus de centralização e intermediação são amplamente assegurados pelos professores dos cursos em análise. Além disso, a densidade da rede é de 0.078, evidenciando uma rede com ligações fortes somente entre os nós centrais e pouca ligação distribuída, significando uma baixa variedade de coautorias. Também é possível verificar a existência de 5 comunidades que se comunicam pelos seus nós centrais.

**Figura 26 - Grafo A e B representando respectivamente as redes de coautorias dos anos de 2015/16/17 e 2018/19/20 segundo os artigos publicados pelos professores na Plataforma Lattes.**

**A**

**B**



O grafo B tem um total de 71 nós e 208 ligações. Apesar dos maiores graus de centralização e intermediação da rede ainda serem dos professores dos cursos em análise, a densidade da rede aumentou minimamente, em comparação ao grafo anterior, sendo o valor agora de 0.084, no entanto, esse aumento não foi o suficiente para que se alterasse a conectividade dos nós, pois ainda predominam as ligações mais fortes nos coautores centrais, consequentemente, existindo uma baixa distribuição de colaborações científicas.

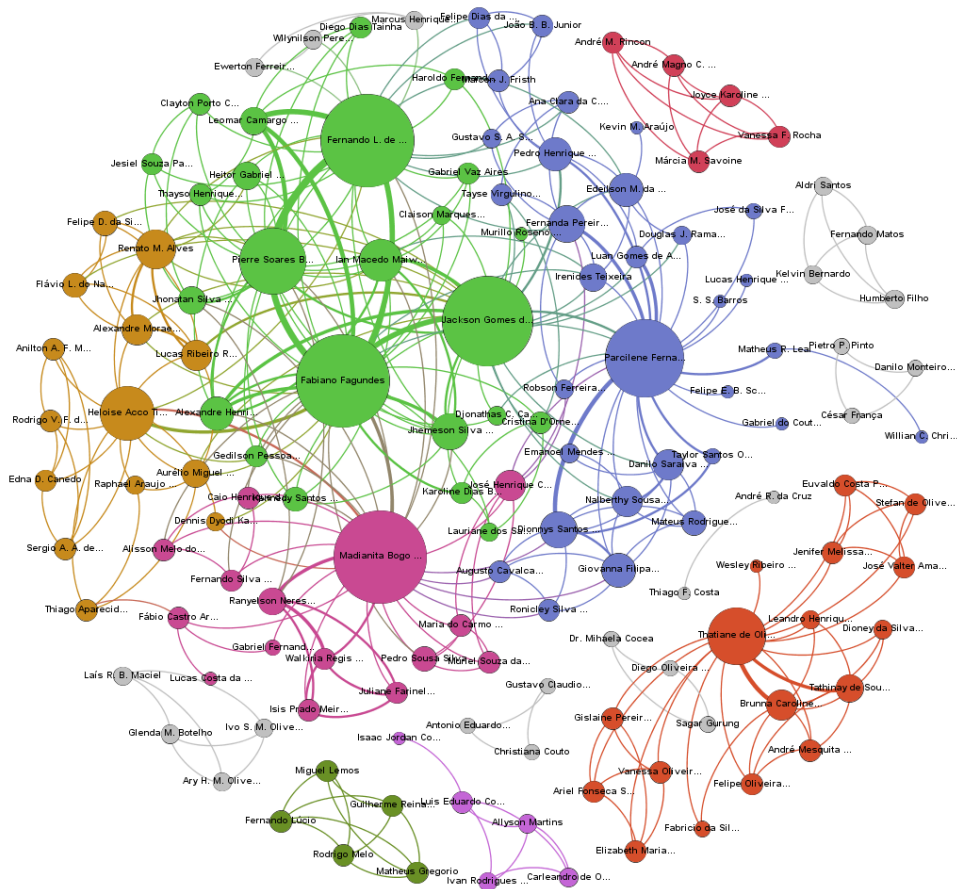
No entanto, é possível verificar um aumento no número de comunidades, sendo agora um total de 7, correlacionando com o aumento de produções científicas e coautorias desse período.

#### 4.4.4 Grafos e parâmetros da rede de coautoria a partir do ENCOINFO

A Figura 27 representa o grafo da rede de coautoria do *corpus* deste projeto. Ele foi analisado segundo os indicadores de densidade, grau de centralidade, grau de intermediação e modularidade.

Dessa forma, foi obtida uma densidade de 0.037, a partir da fórmula inicial  $D = 2 \cdot L/N$ , o que evidencia uma rede semelhante à da Plataforma Lattes. Apesar do valor ser inferior a outra, ambas apresentam pouca conectividade entre os nós e ligações bastante fracas. Essa condição é verificada no grafo através dos diversos grupos afastados e poucas conexões, além disso, com exceção dos coautores mais ao centro da rede que estabelecem relações mais fortes, as ligações são predominantemente tênues.

**Figura 27 - Grafo da rede de coautoria segundo os artigos publicados no ENCOINFO nos anos de 2015 a 2020.**



Os professores dos cursos analisados destacam-se novamente na rede como coautores principais por estarem presentes em quase todos os artigos, sendo assim, eles detêm os maiores graus de centralização. Procedendo disto, é possível verificar que os 3 primeiros professores possuem o mesmo grau, que se mantém acima de 0.20, são eles: Fabiano Fagundes, Fernando Luiz e Madianita Bogo. Além disso, alguns alunos dos cursos, se destacam com graus elevados se comparado a média geral, como o Ian Macedo com 0.09, Renato M. Alves com 0.07 e, também, Fernanda Pereira Gomes com 0.065. É importante ressaltar que no período de análise destes dados a autora Fernanda Pereira Gomes ainda não era docente dos cursos em análise.

Consequentemente, assim como na rede da Plataforma Lattes, os nós intermediários também são os docentes, pois, além de estarem mais ao centro, também possuem maior nível de diversificação do campo de estudo, o que gera novas relações de coautorias e mais ligações na rede. Já os nós mais afastados tendem a formar conectividades menores, pois permanecem dentro de grupos da mesma intenção de pesquisa científica. O grau foi obtido a partir do cálculo utilizado anteriormente, e com isso normalizou-se os valores máximos como 1. Portanto, obtém-se os 5 primeiros coautores com mais importância para a rede, e que em uma eventual remoção, causariam um maior impacto na produção científica, são eles: Madianita Bogo com

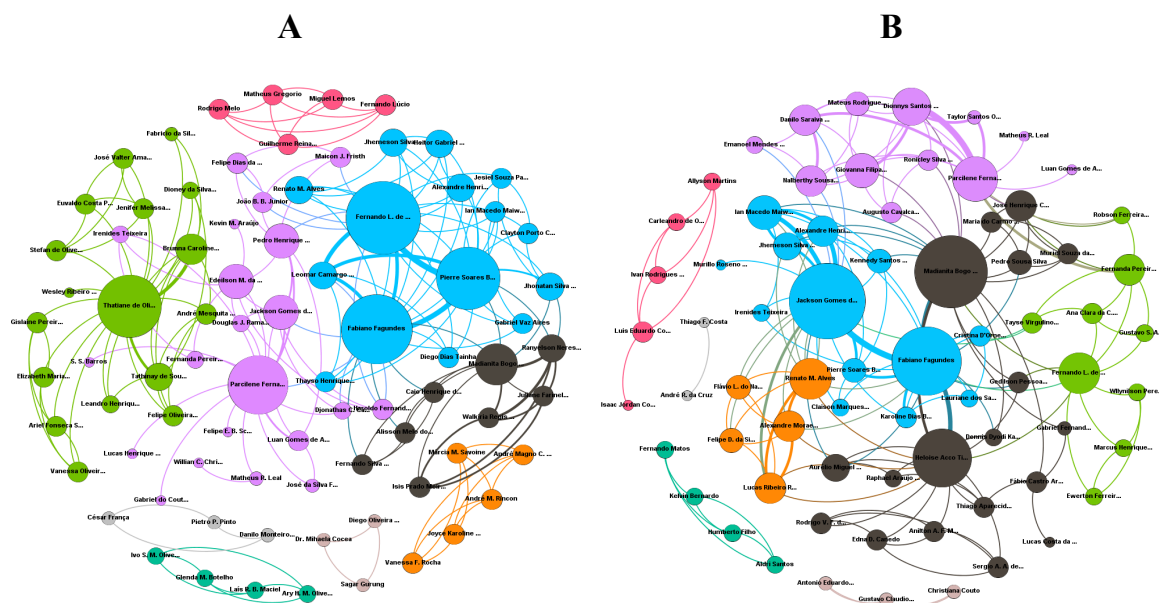
0.1, Fabiano Fagundes com 0.09, Parcilene Fernandes com 0.09, Fernando Luiz com 0.08 e Jackson Gomes com 0.07.

Foi identificado, de acordo com a modularidade, um total de 15 comunidades, definidas pela proximidade de seus vizinhos e pela densidade de ligação entre eles. Quanto mais espessa a aresta no grafo, maior também a ligação entre os nós, ou seja, os coautores possuem mais artigos publicados juntos. Vale destacar que a rede é dominada, com mais de 50% de tamanho, pelas comunidades vermelha, azul e verde, no entanto, existem algumas que possuem poucos nós e são representadas pela cor cinza.

A partir do que já foi apresentado, mostrou-se interessante verificar a evolução da rede de coautoria dentro do *corpus* desta pesquisa. A Figura 28 mostra 2 grafos, A e B, representando, respectivamente, as redes de coautoria dos anos de 2015/16/17 e 2018/19/20.

Diante disso, foi possível obter alguns resultados comparativos entre os dois grafos. No grafo A, tem-se um total de 81 nós e 188 ligações, e os graus de centralização e intermediação são amplamente assegurados pelos professores dos cursos em análise. Além disso, a densidade da rede é de 0.058, o que configura a rede com fracas ligações e predomina a homofilia que significa a ocorrência de relações entre atores que possuem atributos similares, sem diversidade. No entanto, contém uma modularidade formada por um bom número de nós.

**Figura 28 - Grafo A e B representando respectivamente as redes de coautorias dos anos de 2015/16/17 e 2018/19/20 segundo os artigos publicados no ENCOINFO.**



Na sequência, o grafo B tem um total de 71 nós e 177 ligações. Apesar dos maiores graus de centralização e intermediação da rede ainda serem dos professores, a densidade da rede aumentou, em comparação ao grafo anterior, sendo o valor agora de 0.071, portanto, a rede tem

ligações mais fortes e amplamente conectadas, além disso, as comunidades se tornaram mais relacionadas e menos dispersas, fazendo a modularidade da rede aumentar em relação ao grafo A. No entanto, ainda persistem alguns grupos desconexos dentro da rede, como o verde e vermelho.

#### 4.5 GRÁFICOS E TABELAS

Nesta seção os gráficos e as tabelas das redes de coautorias foram separados de acordo com a sua base de dados, sendo a primeira da Plataforma Lattes e a segunda do ENCOINFO.

##### 4.5.1 Gráficos e tabelas da rede de coautoria a partir da Plataforma Lattes

O primeiro dado estatístico é demonstrado na Figura 29 e se refere aos números totais de artigos, coautores e ligações entre eles, provenientes dos artigos publicados pelos professores na Plataforma Lattes entre os anos de 2015 e 2020.

É relevante citar que nessa tabela o número de coautores passou por um filtro, onde foram excluídos os nomes que se repetiam em anos diferentes, como exemplo, o coautor Fabiano Fagundes, que se repetiu em todos os anos analisados pela pesquisa. Dessa forma, para evitar ruídos na contagem absoluta dos dados foi realizada essa exclusão, sendo contabilizado somente uma vez cada coautor.

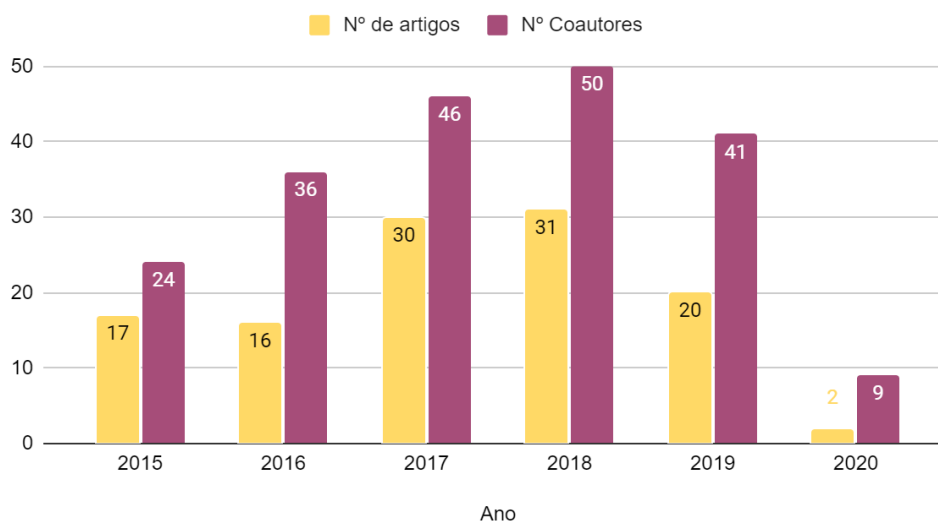
**Figura 29 - Números totais de artigos, coautores e ligações.**

Nº de artigos	Nº Coautores	Nº de ligações
116	113	360

A Figura 30 representa o número de artigos e coautores segundo as publicações dos professores na Plataforma Lattes, separado entre os anos do *corpus* da pesquisa, dessa forma, o número de coautores é evidenciado em roxo e o número de artigos em amarelo.

**Figura 30 - Gráfico do número de artigos e número de coautores por ano.**

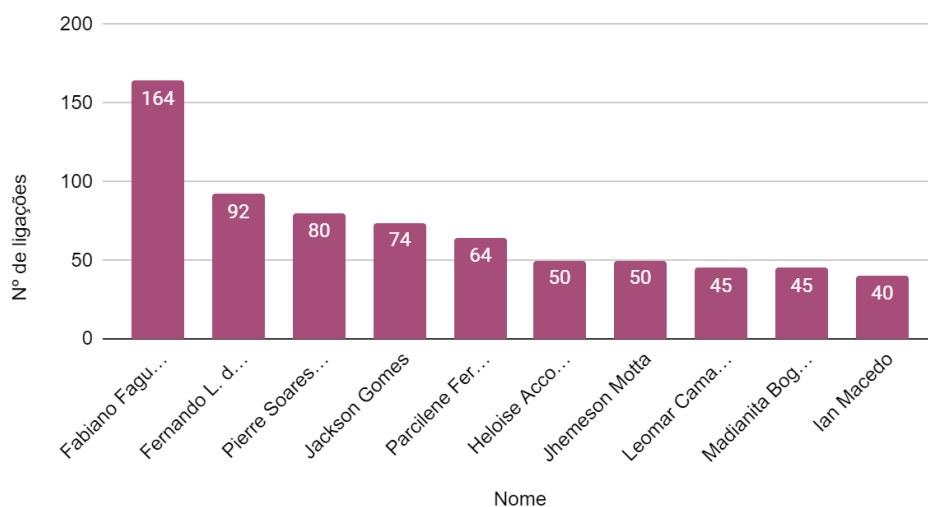
### Nº de artigos e Nº Coautores por ano



Para ampliar a visualização dos dados em relação aos parâmetros de grau de centralidade, intermediação e modularidade, a Figura 31 quantifica o número de ligações dos dez primeiros coautores da rede. Os três primeiros coautores, Fabiano Fagundes, Fernando Luiz e Pierre Brandão representam aproximadamente 40% das ligações totais da rede e, desta forma, obtêm os maiores graus de centralização e intermediação, o que os tornam extremamente imprescindíveis para a rede, oferecendo maior produtividade científica e relevância nas diversificações das áreas de estudo. Vale destacar que estes três primeiros atores pertencem à mesma comunidade, portanto, pode-se inferir que o aumento da produção científica de determinado coautor influenciou no crescimento dos números dos outros coautores do mesmo grupo.

**Figura 31 - Gráfico do número de ligações dos dez primeiros coautores da rede de coautoria.**

Nº de ligações dos 10 primeiros coautores



#### 4.5.2 Gráficos e tabelas da rede de coautoria a partir do ENCOINFO

O primeiro dado estatístico é demonstrado na Figura 32 e se refere aos números totais de artigos, coautores e ligações entre eles, provenientes dos artigos publicados no ENCOINFO entre os anos de 2015 e 2020.

É importante frisar que nessa tabela o número de coautores também passou por um filtro, assim como na da Plataforma Lattes, onde foram excluídos os nomes que se repetiam em anos diferentes, como exemplo, o coautor Fabiano Fagundes, que se repetiu em todos os anos analisados pela pesquisa. Dessa forma, tende-se a evitar os ruídos na contagem absoluta dos dados foi realizada essa exclusão, sendo contabilizado somente uma vez cada coautor.

**Figura 32 - Números totais de artigos, coautores e ligações.**

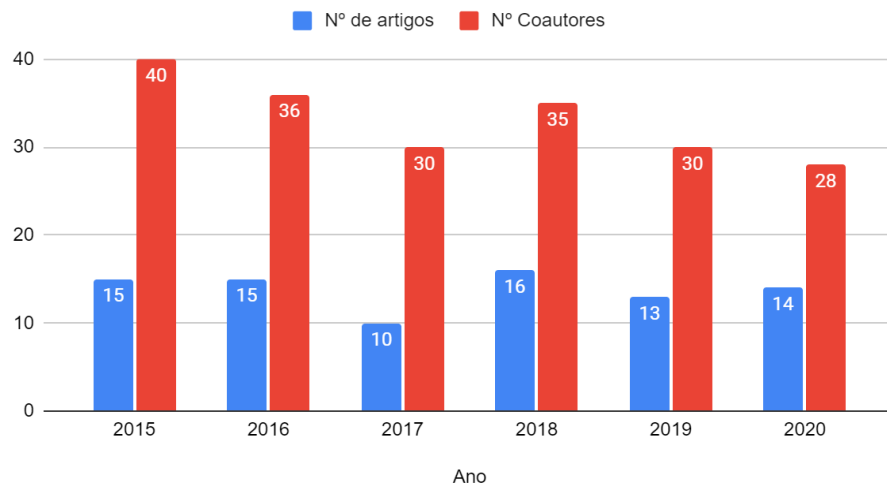
Nº de artigos	Nº Coautores	Nº de ligações
83	138	348

A Figura 33 representa o número de artigos e coautores do ENCOINFO separado entre os anos do *corpus* da pesquisa, dessa forma, o número de coautores é evidenciado em vermelho e o número de artigos em azul.

**Figura 33 - Gráfico do número de artigos e número de coautores por ano.**



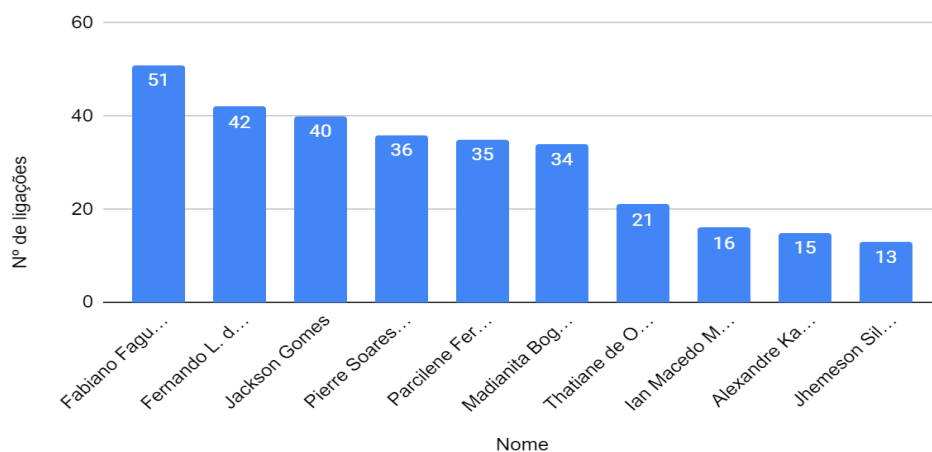
### Nº de artigos e Nº Coautores por ano



Para ampliar a visualização dos dados em relação aos parâmetros de grau de centralidade, intermediação e modularidade, a Figura 34 quantifica o número de ligações dos dez primeiros coautores da rede. Os três primeiros coautores, Fabiano Fagundes, Fernando Luiz e Jackson Gomes representam aproximadamente 38% das ligações totais da rede e, desta forma, obtêm os maiores graus de centralização e intermediação.

**Figura 34 - Gráfico do número de ligações dos dez primeiros coautores da rede de coautoria.**

### Nº de ligações dos 10 primeiros coautores



Além disso, é importante evidenciar que os nós pertencem à mesma comunidade, oferecendo indicação de maior produtividade científica e de relevância nas diversificações das áreas de estudo.

## 5 CONSIDERAÇÕES FINAIS

Este trabalho teve como propósito coletar os dados dos pesquisadores disponibilizados pela Plataforma Lattes e ENCOINFO e criar grafos que apresentem visualmente as redes de coautorias, possibilitando a obtenção de métricas que contribuam para o entendimento e direcionamento das pesquisas nos cursos citados. Para essa finalidade, foram executadas etapas que envolveram diversos objetivos, desde a definição do *corpus* da pesquisa, coleta e tratamento de dados, criação de grafos, a extração de parâmetros, até a apresentação de gráficos e tabelas.

Os objetivos específicos do trabalho foram focados em caracterizar as redes de coautorias dos cursos de Ciências da computação, Sistemas da Informação e Engenharia de Software do CEULP/ULBRA. Assim sendo, foram seguidas as etapas de coleta dos dados dos pesquisadores que estão disponíveis na Plataforma Lattes e anais do ENCOINFO, criação de grafos que organizassem as redes de coautorias, utilização de parâmetros como grau de centralidade, grau de intermediação, densidade e nível de agregação, e por fim a apresentação das informações da rede de coautorias em gráficos e tabelas. Ambos objetivos foram atingidos.

Assim, foi possível coletar os dados dos pesquisadores que estão disponíveis na Plataforma Lattes e anais do ENCOINFO através da utilização da ferramenta *Coletaprod* e de código *Python*. Esta etapa demandou tempo e estudo de ferramentas possíveis a serem utilizadas no processo, além disso, não existia ferramenta que pudesse extrair os dados dos anais do ENCOINFO, e por isso foi necessária a produção de um código em *Python*, diferentemente da Plataforma Lattes que já tinha algumas ferramentas a disposição e só necessitava da compreensão de alguma delas para utilizar.

O objetivo de criar grafos que organizem as redes de coautorias foi alcançado utilizando a plataforma *Gephi*, que é usada para exploração e compreensão de redes. Nesse sentido, no presente trabalho a produção dos grafos passou por algumas fases, indo desde a importação dos dados, até a seleção de algoritmos de visualização e melhorias estéticas para visualização. Também foram identificadas características do campo científico nos cursos citados, utilizando parâmetros como grau de centralidade, grau de intermediação, densidade e nível de agregação, com base nas fórmulas dos parâmetros e métricas disponibilizadas pela ferramenta *Gephi*. Dessa forma, utilizando como base como os grafos criados foi possível extrair essas métricas, aplicadas a sua fórmula, bem como apresentar as informações da rede de coautorias em gráficos e tabelas.

Em vista disso, e através dos resultados apresentados na seção anterior, o objetivo

principal deste trabalho foi atingido através da caracterização das redes de coautorias dos cursos citados. Isso permitirá que os dados contribuam para o entendimento e direcionamento das pesquisas nos cursos citados, além disso, proporcionará a possibilidade de medir a existência de um feito científico real e como ele ocorre.

Para trabalhos futuros, sugere-se a ampliação da rede de coautoria para incluir outros cursos do CEULP/ULBRA e outros eventos de interesse da instituição. Ainda seria interessante caracterizar a rede a partir de mais indicadores, como: componentes conectados, transitividade e distância geodésica, para que os resultados permitissem maior abrangência e assertividade. Pode-se ainda acrescentar a possibilidade de oferecer um *dashboard* com visualização das informações de forma dinâmica.

## REFERÊNCIAS

- ALEJANDRO, Velasquez A. O.; NORMAN, Aguilar G. **Manual introdutório à Análise de Redes Sociais**. 2005. Disponível em: <[http://www2.unicentro.br/lmqqa/files/2016/05/Manualintrodutorio\\_ex\\_ucinet.pdf](http://www2.unicentro.br/lmqqa/files/2016/05/Manualintrodutorio_ex_ucinet.pdf)>.
- ANDRADE, Deyvyd Manoel C.; DAVID, Helena Maria S. L. Análise de redes sociais: uma proposta metodológica para a pesquisa em saúde e na enfermagem. **Revista de Enfermagem UFRJ**, Rio de Janeiro, v. 6, n. 23, p.852-855, 23 dez. 2015.
- ANDRADE, Ricardo L. de. **A Influência das Redes de Coautoria na Performance dos Bolsistas de Produtividade e nos Programas de Pós- Graduação em Engenharia de Produção**. 2016. 145 f. Dissertação (Mestrado) - Curso de Engenharia de Produção, Universidade Federal de Pernambuco, Recife, 2016.
- BALANCIERI, Renato. **Análise de redes de pesquisa em uma plataforma de gestão em ciência e tecnologia**: uma aplicação à plataforma Lattes. 2004. Dissertação (Mestrado em Engenharia de Produção) – Programa de Pós-graduação em Engenharia de Produção – Universidade Federal de Santa Catarina. 2004.
- BARBASTEFANO, R. G.; SOUZA, C.; COSTA, J. S.; TEIXEIRA, P. M. **Impactos dos nomes nas propriedades de redes sociais**: um estudo em rede de coautoria sobre sustentabilidade. *Perspectivas em Ciência da Informação*, v. 18, n. 3, p. 78-95, 2013.
- BASTIAN, Mathieu; HEYMANN, Sebastien; JACOMY, Mathieu. Gephi: An Open Source Software for Exploring and Manipulating Networks. **Proceedings of the International AAAI Conference on Web and Social Media**, [S. l.], v. 3, n. 1, p. 361-362, 2009.
- BEAVER, Donald. B.; ROSEN, R. **Studies in scientific collaboration**: part I: The professional origins of scientific co-authorship. *Scientometrics*, Amsterdam, v.1, n.1, p.65-84, 1978.
- BORGATTI, Stephen P.; FOSTER, Pacey C. The Network Paradigm in Organizational Research: A Review and Typology. **Journal Of Management**. Boston, p. 991-1013. 21 mar. 2003.
- BUFREM, L. S.; GABRIEL JUNIOR, R. F.; SORRIBAS, T. V. Redes sociais na pesquisa científica da área de ciência da informação. **DataGramZero**, v. 12, n. 4, 2011. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/7410>. Acesso em: 18 set. 2021.
- CERVANTES, Evelyn Perez. **Análise de Redes da Colaboração Científica**: Uma abordagem baseada em grafos relacionais com atributos. 2015. 82 f. Dissertação (Mestrado) - Curso de Ciência da Computação, Universidade de São Paulo, São Paulo, 2015.
- CHERVEN, Ken. **Mastering Gephi Network Visualization**: produce advanced network graphs in gephi and gain valuable insights into your network datasets. Birmingham: Packt Publishing, 2015.
- CHERVEN, Ken. **Network graph analysis and visualization with Gephi**. Birmingham: Packt Publishing Ltd, 2013.
- CNPQ - Conselho Nacional de Desenvolvimento Científico e Tecnológico - **Diretório dos grupos de pesquisa**. Disponível em: <http://dgp.cnpq.br/censos/>.
- COTA, Ricardo G.; GONÇALVES, Marcos André; LAENDER, Alberto HF. A Heuristic-based Hierarchical Clustering Method for Author Name Disambiguation in Digital Libraries. In: **SBBB**. 2007. p. 20-34.

- DE CHOUDHURY, M. et al. Inferring relevant social networks from interpersonal communication. In: **INTERNATIONAL CONFERENCE ON WORLD WIDE WEB**, 19., 2010, New York. Proceedings. New York: Association for Computing Machinery, 2010, p. 301–310.
- DEGENNE, Alain; FORSÉ, Michel. **Introducing Social Networks**. London: Sage, 1999.
- DE STEFANO, Domenico; GIORDANO, Giuseppe; VITALE, Maria Prosperina. **Issues in the analysis of co-authorship networks**. *Quality & Quantity*, v. 45, n. 5, p. 1091-1107, 2011.
- DIGIAMPIETRI, L.; MENA-CHALCO, J.; PÉREZ-ALCÁZAR, J.; TUESTA, E.; DELGADO, K.; MUGNAINI, R.; SILVA, G.; LIMA, J. Extração, Caracterização e Análises de Dados de Currículos Lattes. **Revista Eletrônica de Sistemas de Informação**, v. 14, n. 2, p. 1-24, 2015.
- ENCOINFO: **Apresentação**. 2021. Disponível em: <http://ulbra-to.br/encoinfo/apresentacao/>. Acesso em: 26 set. 2021.
- GLÄNZEL, W; SCHUBERT, A. **Analysing scientific networks through co-authorship**. Handbook of quantitative science and technology research: the use of publication and patent statistics in studies of S&T systems. New York: Kluwer Academic Publishers, 2004.
- GRANDJEAN, Martin. **La connaissance est un réseau**. 2014. Disponível em: [https://commons.wikimedia.org/wiki/File:Social\\_Network\\_Analysis\\_Visualization.png](https://commons.wikimedia.org/wiki/File:Social_Network_Analysis_Visualization.png). Acesso em: 30 set. 2021.
- HOUVARDAS, J.; STAMATATOS, E. (2006). N-gram feature selection for authorship identification. In: Euzenat, J., Domingue, J. (eds) **Artificial Intelligence: Methodology, Systems, and Applications**. AIMSA 2006., LNCS 4183 (pp. 77–86). Berlin: Springer-Verlag.
- KATZ, J. S.; MARTIN, B. R. **What is research collaboration?** *Research Policy*. Amsterdam, n. 26, p. 1-18, 1997.
- KANG, I. S. (2009). On co-authorship for author disambiguation. **Information Processing and Management**, 45(1), 84–97.
- KHOKHAR, Devangana. **Gephi cookbook**. Birmingham: Packt Publishing Ltd, 2015.
- LAUMANN, Edward O.; MARSDEN, Peter V.; PRENSKY, David. The boundary specification problem in network analysis. **Research methods in social network analysis**, v. 61, p. 87, 1989.
- LIMA, Murillo R. F.; GOMES, Jackson. Desenvolvimento de um Data Mart e Automatização do Processo ETL no contexto da Produção Acadêmica do CEULP/ULBRA. In: XXI ENCOINFO – Congresso de Computação e Tecnologias da Informação, 21., 2019, Palmas. **Anais [...]**. Palmas: Congresso de Computação e Tecnologias da Informação, 2019. p. 73-79.
- MACEDO, Thais L. da S. **Composição de grupos de pesquisa e sua relação com redes de coautorias: o caso do IFBA**. 2014. 115 f. Dissertação (Mestrado) - Curso de Administração, Universidade Federal da Bahia, Salvador, 2014.
- MARSDEN, Peter. V.: **Recent developments in network measurement**. Models and Methods in Social Network Analysis. Cambridge University Press, New York, 2005.
- MARTINS, Dalton Lopes. **Análise de redes sociais de colaboração científica no ambiente de uma federação de bibliotecas digitais**. 2012. 255 f. Tese (Doutorado) - Curso de Ciência da Informação, Universidade de São Paulo, São Paulo, 2012.
- MEADOWS, A. J. **A comunicação científica**. Brasília: Briquet de Lemos, 1999.

MUELLER, Suzana P. M.; PASSOS, Edilenice. **As questões da comunicação científica e a ciência da informação**. Brasília: Departamento de Ciência da Informação - Universidade de Brasília, 2000.

MURAKAMI, Tiago R. M.; LARA, Jan L. de. **Coletaprod**. Disponível em: < <https://zenodo.org/badge/latestdoi/77038207> >, Acesso em: 08/11/2021

NEWCOMBE, H. B.; KENNEDY, J. M.; AXFORD, S. J.; JAMES, A. P. (1959). Automatic Linkage of Vital Records Computers can be used to extract "follow-up" statistics of families from files of routine record. **Science**, Vol. 130, No. 3381. [S.]

NEWMAN, Mark E. J. **Who is the best connected scientist?** a study of scientific coauthorship networks. Santa Fé: The Santa Fé Institute, 2000.

NEWMAN, M. E. J. **The structure of scientific collaboration networks**. Santa Fe Institute. Santa Fe, p. 404-409. 16 jan. 2001.

OLIVEIRA, Walison A. de. **Colaboração científica nos programas de pós-graduação em educação: uma análise de redes de coautoria**. 2017. Dissertação (Mestrado em Educação) - Universidade Federal de São Carlos, São Carlos, 2017.

PARREIRAS, F. S.; SILVA, A.; MATHEUS, R. F.; BRANDÃO, W. C. Redeci: colaboração e produção científica em ciência da informação no brasil. **Perspectivas em Ciência da Informação**, v. 11, n. 3, 2006. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/35765>. Acesso em: 18 set. 2021.

PHELAN, T. J. (1999). A compendium of issues for citation analysis. **Scientometrics**, 45(1), 117–136.

PLATAFORMA LATTES. Cnpq (org.). **Sobre a plataforma Lattes**. Disponível em: <http://lattes.cnpq.br/web/plataforma-lattes/o-que-e>. Acesso em: 26 set. 2021.

RECUERO, Raquel. **Introdução à análise das redes sociais online**. Salvador: Edufba, 2017.

REICHERT, Taís; NOVAES, Tiago; RECH, Munique; OLEA, Pelayo. Redes de Colaboração Científica: Um Estudo de Coautoria Através da Análise de Redes Sociais. **Mostra de Iniciação Científica, Pós-graduação, Pesquisa e Extensão**. 5. 1-15, 2016.

RODRIGUES, Miguel (2011). **O tratamento e análise de dados**. In Silvestre, Hugo Consciência; Araújo, Joaquim Filipe Cap. IX de Metodologia para a Investigação Social. Lisboa: Escolar Editora. p. 179-230.

SAMPAIO, R. B. et al.. A colaboração científica na pesquisa sobre coautoria: um método baseado na análise de redes. **Revista Perspectivas em Ciência da Informação**, v20, n.4, p.79-92, out./dez. 2015.

SAVIĆ, M.; IVANOVIĆ, M.; DIMIĆ SURLA, B. Analysis of intra-institutional research collaboration: a case of a Serbian faculty of sciences. **Scientometrics**, v. 110, n. 1, p. 195–216, 2017.

SILVA, Alzira K. A. da. **Redes de coautoria em Ciência da Informação no Brasil: dinâmica na produção científica dos atores mediada pela Ancib**. 2012. 252 f. Tese (Doutorado) - Curso de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2012.

SIMÕES-PEREIRA, J. M. S.. **Grafos e Redes: Teoria e Algoritmos básicos**. Rio de Janeiro: Interciência, 2013.

SMALHEISER, N. R.; TORVIK, V. I. Author name disambiguation. **Annual Review of Information Science and Technology**, v. 43, p. 287-313, 2009.

SONNENWALD, D. H. Scientific Collaboration. **Annual Review of Information Science and Technology**, v.42, n.1, p.643-681, 2008.

STROTMANN, Andreas; ZHAO, Dangzhi; BUBELA, Tania. Author name disambiguation for collaboration network analysis and visualization., **Proceedings of the American Society for Information Science and Technology**, Vol. 46, pp. 1-20. 2009.

TANG, Li; WALSH, John P. (2010). Bibliometric fingerprints: name disambiguation based on approximate structure equivalence of cognitive maps. **Scientometrics**. Springer;Akadémiai Kiadó, 84(3), 763–784.

TOMAÉL, Maria Inês. **Redes de conhecimento**: o compartilhamento da informação e do conhecimento em consórcio de exportação do setor moveleiro. Belo Horizonte, 2005. 289f. Tese (Doutorado em Ciência da Informação) - Universidade Federal de Minas Gerais. Belo Horizonte, 2005.

VANZ, Samile A. de S.; STUMPF, Ida R. C. Colaboração científica: revisão teórico-conceitual. **Perspectivas em Ciência da Informação**, Belo Horizonte, v.15, n.2, p.42-55, maio/ago. 2010. Disponível em: < <https://www.scielo.br/j/pci/a/Fz4q6DhPGhjnXmRxLw6Ct/?lang=pt>>.

VANZ, Samile Andréa de Souza. **As redes de colaboração científica no Brasil**. 2009. 204 f. Tese (Doutorado em Comunicação e Informação) – Universidade Federal do Rio Grande do Sul, Faculdade de Biblioteconomia e Comunicação, 2009.

WANG, D. J. et al. Measurement error in network data: a re-classification. **Social Networks**, v. 34, p. 396-409, 2012

WASSERMAN, Stanley; FAUST, Katherine. **Social network analysis: methods and applications**. New York: Cambridge University Press, 1994. p. 857.

## **APÊNDICES**



## APÊNDICE A - Código Python utilizado na coleta dos dados do ENCOINFO.

```
1 import io
2 arq = open('18.html', 'r', encoding='utf8')
3 conteudo = arq.read()
4 arq.close()
5
6 inicioTitulo = conteudo.find('titulo-link')+30
7 lista = []
8
9 if inicioTitulo != '': temAutores = True
10
11 while temAutores:
12
13     fimTitulo = conteudo.find('/h6', inicioTitulo)-27
14     linha = conteudo[inicioTitulo:fimTitulo]+';'
15
16     inicioAutor = conteudo.find('line-content', fimTitulo)
17     inicioAutor = conteudo.find('/span', inicioAutor)+6
18     fimAutor = conteudo.find('span', inicioAutor)-1
19
20     autores = conteudo[inicioAutor:fimAutor]
21     autores=autores.replace(' ', ',')
22     listaDeAutores = autores.split(',')
23
24     for i in listaDeAutores:
25         linha += i+';'
26     lista.append(linha)
27
28     arq = open('autores2.csv', 'w')
29     for i in lista:
30         lista2 = i.split(',')
31         for j in lista2:
32             arq.write(j+';')
33         arq.write('\n')
34
35     temAutores = False
36     inicioTitulo = conteudo.find('titulo-link', fimTitulo)+30
37     if inicioTitulo != '' and inicioTitulo>fimTitulo: temAutores = True
38
39 arq.close()
```

## **ANEXOS**

## ANEXO A - Artigo do XXIII Encoinfo – Congresso de Computação e Tecnologias da Informação.

### **Caracterização da rede de coautoria dos cursos de Ciência da Computação, Sistemas de Informação e Engenharia de Software do CEULP/ULBRA a partir das publicações no ENCOINFO**

**Caio Vinicius Davantel, Fabiano Fagundes**

Curso de Engenharia de Software

Centro Universitário Luterano de Palmas (CEULP/ULBRA) – Palmas, TO – Brasil

{caiodavantel,thilfa}@gmail.com

**Resumo.** *Este trabalho tem como propósito caracterizar a rede de coautoria dos cursos de Ciência da Computação, Sistemas de Informação e Engenharia de Software do CEULP/ULBRA, utilizando os dados disponibilizados pelos anais do ENCOINFO. Foram desenvolvidos e estruturados grafos, tabelas e gráficos com as ferramentas Microsoft Excel e Gephi e, na sequência, aplicaram-se parâmetros de análise de redes de coautoria para obter os resultados. Desta forma, foi possível extrair indicadores que revelam como o campo de pesquisa evoluiu e como ele está sendo gerenciado, fornecendo insights e um agregado de informações para os gestores da instituição e cursos citados, contribuindo na tomada de decisão em relação ao campo científico.*

#### **1. Introdução**

A colaboração científica é uma prática exercida há muito tempo, antes mesmo do século XX. De acordo com Beaver e Rosen (1978), o primeiro artigo publicado em colaboração científica tem a data definida em 1665. Foram encontradas, ainda, mais 21 produções colaborativas entre 1665-1760, com destaque para o tema de astronomia, que mais envolveu os pesquisadores nessa época. As coautorias têm sido estudadas desde o século XVII (BEAVER; ROSEN, 1978), e buscaram entender a produção científica por meio da análise da colaboração entre os pesquisadores (PARREIRAS *et al.*, 2006).

Segundo Vanz e Stump (2010), a coautoria científica é reconhecida a partir do trabalho conjunto de dois ou mais cientistas em projetos de pesquisas, que por consequência compartilham recursos intelectuais, econômicos e físicos. Pode ser definida como uma interação ocorrendo dentro de um contexto social entre dois ou mais cientistas que facilita o compartilhamento de significados e conclusão de tarefas, no que diz respeito a um mutuamente objetivo superordenado e compartilhado (SONNENWALD, 2007).

Devido ao grande número de artigos com coautorias, não é de se espantar que também haja inúmeras motivações para que ela ocorra. Segundo Balancieri (2004), o mais óbvio dos motivos, certamente, é a necessidade de contribuição especializada para alcançar os objetivos da pesquisa, incluindo o ganho de novas habilidades e conhecimento tácito.

Um conjunto de colaborações forma uma rede de coautorias que, segundo Silva (2012), contém estratégias para compartilhar informações e conhecimentos através das comunidades científicas, sendo que, dentre os benefícios da colaboração, pode-se obter uma maior qualidade dos artigos e um incremento do número de publicações. Além disso, o estudo das

redes de coautoria permite identificar e analisar porque os pesquisadores colaboram e quais as consequências dessas colaborações. Pesquisas apontam a melhoria da qualidade dos artigos e o aumento do número de publicações como benefícios da colaboração (ANDRADE, 2016).

Uma Rede de coautoria pode ser estudada através de grafos, que têm sido amplamente utilizados para representar as redes sociais e quantificar a sua estrutura. Além disso, na opinião de Bufrem, Junior e Sorribas (2011), a visualização por meio desse modelo é considerada mais intuitiva do que a visualização na forma de matrizes, embora os próprios autores afirmem que os dados coletados sejam, normalmente, apresentados dessa forma. Como os grafos são baseados na matemática, a partir de sua análise, pode-se obter indicadores que possibilitam a caracterização das redes de coautorias, dessa forma, Silva (2012), argumenta que as pesquisas que envolvem esse tema, em sua maioria analisam os seguintes parâmetros: grau de centralidade, a colaboração, a produtividade dos autores e o elitismo da área, a densidade, a transitividade e o tamanho da rede, a distribuição dos laços e cliques.

Diante dos dados importantes que são fornecidos pelo ENCOINFO – Congresso de Computação e Tecnologias da Informação que é realizado pelo Departamento de Computação e Informática do CEULP/ULBRA, mostra-se interessante caracterizar as redes de coautorias dos cursos de Ciência da Computação, Sistemas de Informação e Engenharia de Software do CEULP/ULBRA. Visto que, identificar como as redes estão se desenvolvendo dentro das Instituições de Ensino Superior auxilia no desenvolvimento pedagógico, na medida em que a formação de pesquisadores parte delas (REICHERT *et al.*, 2016).

Assim, este artigo tem como propósito coletar os dados dos pesquisadores disponibilizados pelo ENCOINFO e criar grafos que organizem as redes de coautorias, possibilitando a obtenção de métricas que contribuam para o entendimento e direcionamento das pesquisas nos cursos citados.

## **2. Fundamentação Teórica**

### **2.1 Rede de coautoria**

Newman (2001) diz que uma definição razoável para as redes de coautoria é quando dois ou mais autores colaboram em pesquisas. Esta rede pode ser representada por um conjunto de vértices que denota as pessoas, unidas em pares, e por arestas que denotam a familiaridade entre elas. Portanto, compreende-se por rede, neste contexto, um grupo de indivíduos que se relacionam uns com os outros, mediante determinados objetivos sendo, assim, a rede de coautoria formada por três elementos: autores ou nós, vínculos ou relações e os fluxos (ALEJANDRO; NORMAN, 2005).

Segundo Mueller e Passos (2000), os estudos sobre as comunidades científicas é um tema que tem crescido constantemente em todas as áreas que se interessam pelo assunto, dentre elas a filosofia, a sociologia e a ciência da informação. Essas pesquisas utilizam-se de metodologias quantitativas como, por exemplo, no estudo de citações, produções e mapeamento de ligações entre os pesquisadores. Além disso, recentemente também estão sendo empregados métodos qualitativos, que provém da sociologia e comunicação, como forma de analisar estas redes de coautoria e conseguir compreender além dos indicadores. Degenne e Forsé (1999) afirmam

que uma rede não pode ser reduzida a uma simples soma de relações e que a sua forma exerce uma influência em cada relação. Desta forma, é necessária a análise das redes de coautorias, pois proporciona um melhor entendimento do seu funcionamento, dos hábitos e das estratégias dos pesquisadores, além de influenciar políticas de pesquisa e incentivo para a articulação e cooperação (OLIVEIRA, 2017). Entretanto, são poucos ou superficiais os estudos que analisam essas redes, o que torna interessante conhecê-las mais, mapeando seus atores e relações (SILVA, 2012).

Para caracterizar uma rede de coautoria são necessários alguns parâmetros matemáticos. Segundo Tomaél (2005), as redes apresentam propriedades que devem fornecer algumas medidas que facilitem a sua análise e entendimento, quais sejam: coesão social, densidade da rede, transitividade, distância geodésica, fluxo máximo e centralidade. Bufrem, Junior e Sorribas (2011) também constatam que as propriedades estatísticas da rede destacam-se, com diferenças nos padrões de colaboração entre os indivíduos estudados, por meio de uma série de medidas de centralidade e conexão.

De acordo com Wasserman e Faust (1994) e Cherven (2015), para estudar e analisar uma rede de coautoria, algumas métricas dos indicadores de rede podem ser usadas, destacando-se:

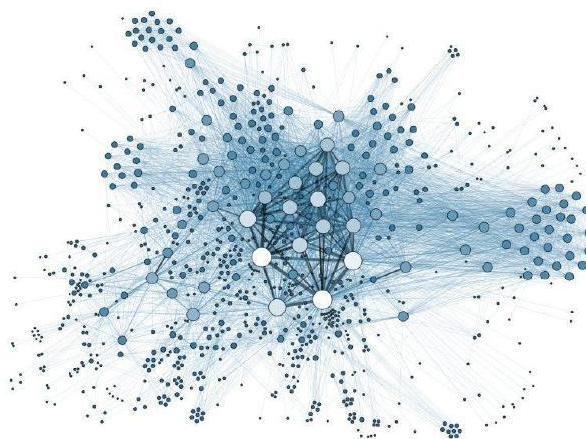
- Nós: podem ser considerados como elementos individuais em uma rede e representar pessoas, lugares ou objetos que definem coletivamente uma rede;
- Ligações: atuam como as conexões que ligam os nós, podem ser representadas por produções científicas, trabalhos, proximidade e qualquer outro objeto que estabeleça uma relação;
- Grau de centralidade: tem relação com o coautor que mais estabeleceu ligações na rede. De acordo com o grafo, os maiores nós são os que têm maior número de ligações e consequentemente estabelecem um maior número de coautorias diferentes;
- Grau de intermediação: é referente ao coautor que mais coordena as ligações dentro da rede, dessa forma, ele é quem detém o maior contato com os outros coautores e também é o que mais se relaciona entre as diversas áreas de estudo;
- Modularidade: define o agrupamento de nós e os separa em comunidades (exemplificadas no grafo através das cores) e que determinam a taxa de representatividade dos coautores e quais os temas de estudo mais publicados dentro da rede de coautoria;
- Densidade: inicialmente a densidade é medida em percentual, sendo que um grafo completo tem percentual igual a 1, sendo assim, quando uma rede apresenta uma densidade muito abaixo desse número, é por não existir alta conectividade entre os nós e ligações fortes entre eles, ou seja, existe uma homofilia, pois muitos grupos estão afastados e pouco se relacionam dentro da rede. O cálculo para obtê-la é:  $D = 2 \cdot L/N \cdot (N-1)$ , de forma que L representa o número de ligações e N representa o número de nós.

Além disso, vale salientar que os indicadores de centralidade permitem analisar a rede tanto no seu conjunto como individualmente, encontrando diversos resultados como: grau de conectividade da rede, indivíduos com o maior e menor número de interações, intermediação de alguns atores nas relações entre indivíduos e a proximidade entre os indivíduos através das suas interações (ALEJANDRO; NORMAN, 2005).

## 2.2 Grafos

Andrade (2016) afirma que as redes de coautoria são um tipo de rede social, logo, sua análise pode ser baseada na premissa de que as relações entre autores da rede podem ser organizadas através de um grafo. Uma definição mais simples de grafo pode ser descrita da seguinte forma: um grafo  $G = (V, E)$  é um sistema formado por um conjunto  $V$  de elementos chamados vértices, pontos ou nodos, e um conjunto  $E$  de pares não ordenados de vértices chamados de arestas. (SIMÕES-PEREIRA, 2013)

Wasserman e Faust (1994) constatam que os grafos têm sido amplamente utilizados na análise de redes de colaboração como uma forma de representar as relações sociais e quantificar as suas propriedades estruturais mais importantes. Também citam que este modelo de visualização beneficia-se da utilização da Teoria dos Grafos que permite a análise das interações sociais, dos padrões e das implicações. A Figura 1 apresenta um grafo utilizado para representar uma rede social.



**Figura 1. Grafo utilizado na representação de uma rede social.**  
**Fonte: Grandjean (2014)**

Um grafo, neste contexto, resulta de uma matriz na qual se introduz previamente a informação de que se dispõe, no entanto, a representação das interações entre os diferentes atores de uma rede faz-se por meio deles, por serem mais compreensíveis e terem uma representação mais agradável para a interpretação (ALEJANDRO; NORMAN, 2005).

## 3. Materiais e métodos

Esta seção aborda os materiais e métodos utilizados na obtenção dos dados e estabelecimento das relações de coautorias dos cursos de Ciência da Computação, Sistemas de Informação e Engenharia de Software do CEULP/ULBRA em suas publicações no ENCOINFO.

O ENCOINFO é o Congresso de Computação e Tecnologias da Informação que é realizado pelo Departamento de Computação e Informática do CEULP/ULBRA e teve sua primeira edição em 1999, então chamado “Encontro de Estudantes de Informática”. Desde então, vinte

edições já foram realizadas, sempre abordando temas-chave da Ciência da Computação com aplicações em diversas áreas (ENCOINFO, 2021). Portanto, sabendo da importância desse evento para a instituição, todos os dados da pesquisa foram extraídos dos anais desta fonte documental eletrônica.

### 3.1 Materiais

A ferramenta Gephi foi utilizada na obtenção dos parâmetros e grafos que foram analisados e mensurados neste trabalho. Este software, objeto de estudo de Cherven (2015), é desenvolvido em código aberto, usado para exploração e compreensão de redes. Oferece inúmeras métricas para análise de redes, como: grau, diâmetro, densidade, coeficiente de cluster, modularidade, componentes conectados, excentricidade e métricas de centralidade (ANDRADE, 2016). Além disso, também foi utilizado o software Microsoft Excel para obter tabelas e dados estatísticos; desta forma, foi possível realizar a análise dos dados quantitativos e mensurar a rede de coautoria.

### 3.2 Métodos

No âmbito dos anais do ENCOINFO, foram definidas como principal *corpus* desta pesquisa, as publicações científicas vinculadas ao evento no intervalo que vai do ano de 2015 ao ano de 2020. Os caminhos metodológicos seguidos no trabalho foram os seguintes:

- a) Coleta manual dos dados dos pesquisadores diretamente dos anais que estão disponíveis de forma documental eletrônica no site do ENCOINFO.
- b) Tratamento dos dados coletados utilizando o software Microsoft Excel. Todo o processo realizado foi de acordo com o modelo de importação da ferramenta Gephi, portanto, os dados foram formatados em matrizes e correlacionados em duas planilhas.

A primeira planilha é relacionada aos nós da rede de coautoria. Desta forma, conforme trecho mostrado na Tabela 1, os nomes, sem repetições, foram colocados na primeira coluna Label, e na segunda coluna foi associado um ID.

Label	ID
Fabiano Fagundes	1
Kennedy Santos Torres	2

**Tabela 1. Formato da Planilha 1.**

**Fonte: Autor**

A segunda planilha é relacionada às arestas da rede, por conseguinte, traz algumas diferenças se comparada a de nós em decorrência das relações em si. A estrutura que é parcialmente apresentada na Tabela 2 ocorre a partir da ligação entre os coautores (*source*) e para onde essa ligação está direcionada (*target*), gerando, assim, as arestas entre autores.

Source	Target
1	2
1	3
1	4

**Tabela 2. Formato da Planilha 2.**

**Fonte: Autor**

- c) Geração dos grafos e medidas com base nos resultados obtidos por meio do software Gephi. Dentre os parâmetros gerados foi utilizado o Grau de centralização, Modularidade, Grau de intermediação e Densidade.
- d) Criação dos gráficos e tabelas estatísticas com a utilização do software Microsoft Excel, tendo como base o grafo de rede coautoria gerado pela ferramenta Gephi. Dentre os dados estatísticos encontrados, foram utilizados para análise os parâmetros: número de artigos, número de coautores, número de ligações e coautores com mais ligações.
- e) Caracterização e interpretação dos resultados obtidos através dos grafos da rede de coautoria, gráficos e tabelas estatísticas .

#### 4. Resultados

Os resultados desta seção foram obtidos através dos métodos descritos anteriormente. O primeiro dado estatístico é demonstrado na Tabela 3 e se refere aos números totais de artigos, coautores e ligações entre eles, provenientes dos artigos publicados no ENCOINFO entre os anos de 2015 e 2020. É importante destacar que nessa tabela o número de coautores passou por um filtro, onde foram excluídos os nomes que se repetiam em anos diferentes, como exemplo, o coautor Fabiano Fagundes, que se repetiu em todos os anos analisados pela pesquisa. Dessa forma, para evitar ruídos na contagem absoluta dos dados foi realizada essa exclusão, sendo contabilizado somente uma vez cada coautor.

Nº de artigos	Nº Coautores	Nº de ligações
83	138	348

**Tabela 3. Números totais de artigos, coautores e ligações.**  
Fonte: Autor

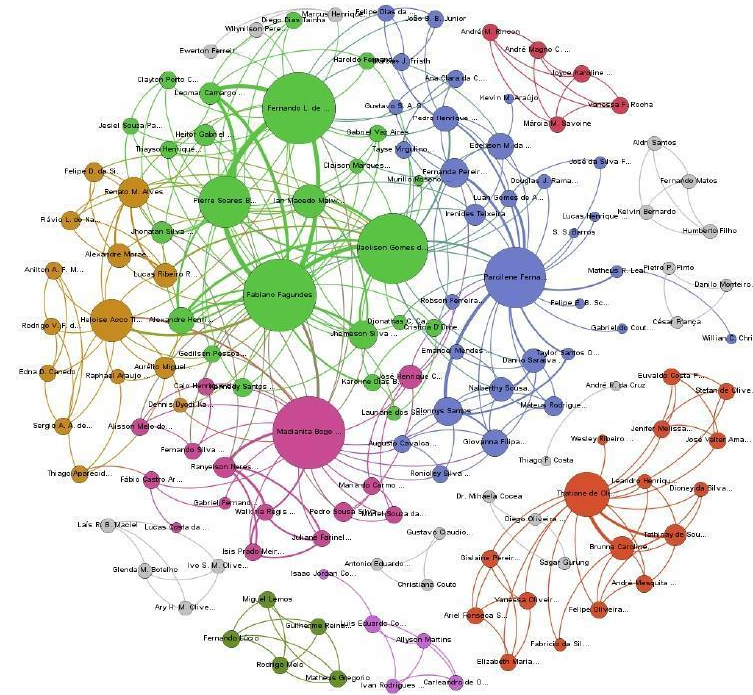
A Figura 2 representa o número de artigos e coautores do ENCOINFO separado entre os anos do *corpus* da pesquisa, dessa forma, o número de coautores é evidenciado em vermelho e o número de artigos em azul.



**Figura 2. Gráfico do número de artigos e número de coautores por ano.**  
Fonte: Autor



A Figura 3 representa o grafo da rede de coautoria do *corpus* desta pesquisa. Ele foi analisado segundo os indicadores de densidade, grau de centralidade, grau de intermediação e modularidade. Dessa forma, foi obtida uma densidade de 0.037, o que demonstra uma rede com pouca conectividade e ligações fracas entre os nós. Essa condição é evidenciada no grafo através dos diversos grupos afastados e poucas conexões, além disso, com exceção dos coautores mais ao centro da rede, as ligações são predominantemente tênues.



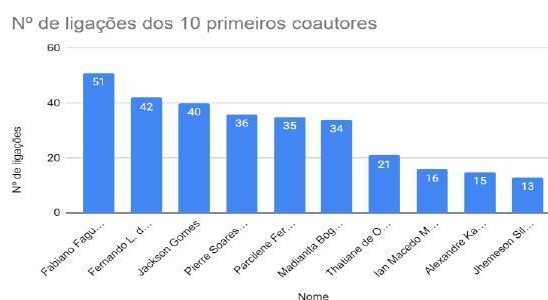
**Figura 3. Grafo da rede de coautoria segundo os artigos publicados no ENCOINFO nos anos de 2015 a 2020.**

Os professores dos cursos analisados destacam-se na rede como coautores assíduos por estarem presentes em quase todos os artigos, sendo assim, eles detêm os maiores graus de centralização. Consequentemente, os nós intermediários também são os docentes, pois além de estarem mais ao centro, também possuem maior nível de diversificação do campo de estudo, o que gera novas relações de coautorias e mais ligações na rede. Já os nós mais afastados tendem a formar conectividades menores, pois permanecem dentro de grupos da mesma intenção de pesquisa científica.

Foi identificado, de acordo com a modularidade, um total de 15 comunidades, definidas pela proximidade de seus vizinhos e pela densidade de ligação entre eles. Desta forma, quanto mais espessa a aresta no grafo, maior também a ligação entre os nós, ou seja, os coautores

possuem mais artigos publicados juntos. Vale destacar que a rede é dominada, com mais de 50% de tamanho, pelas comunidades vermelha, azul e verde, no entanto, existem algumas que possuem poucos nós e são representadas pela cor cinza.

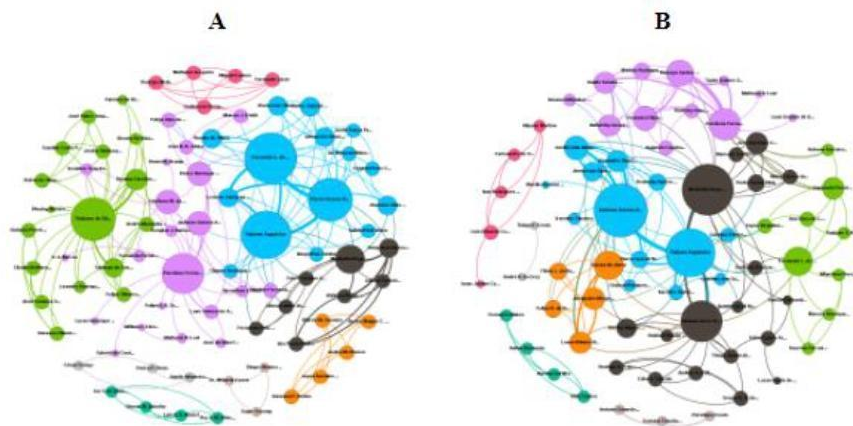
Para ampliar a visualização dos dados em relação aos parâmetros de grau de centralidade, intermediação e modularidade, a Figura 4 quantifica o número de ligações dos dez primeiros coautores da rede. Os três primeiros coautores, Fabiano Fagundes, Fernando Luiz e Jackson Gomes representam aproximadamente 38% das ligações totais da rede e, desta forma, obtêm os maiores graus de centralização e intermediação. Além disso, é importante destacar que os nós pertencem à mesma comunidade, oferecendo indicação de maior produtividade científica e de relevância nas diversificações das áreas de estudo.



**Figura 4. Gráfico do número de ligações dos dez primeiros coautores da rede de coautoria.**

**Fonte: Autor**

A partir do que já foi apresentado, mostrou-se interessante verificar a evolução da rede de coautoria dentro do *corpus* desta pesquisa. A Figura 5 mostra 2 grafos, A e B, representando, respectivamente, as redes de coautoria dos anos de 2015/16/17 e 2018/19/20. Diante disso, foi possível obter alguns resultados comparativos entre os dois grafos. No grafo A, tem-se um total de 81 nós e 188 ligações, e os graus de centralização e intermediação são amplamente assegurados pelos professores dos cursos em análise. Além disso, a densidade da rede é de 0.058, o que configura a rede com fracas ligações, predomina a homofilia, no entanto, contém uma modularidade formada por um bom número de nós.



**Figura 5. Grafo A e B representando respectivamente as redes de coautorias dos anos de 2015/16/17 e 2018/19/20 segundo os artigos publicados no ENCOINFORM.**

Fonte: Autor

Na sequência, o grafo B tem um total de 71 nós e 177 ligações. Apesar dos maiores graus de centralização e intermediação da rede ainda serem dos professores, a densidade da rede aumentou, em comparação ao grafo anterior, sendo o valor agora de 0.071, portanto, a rede tem ligações mais fortes e amplamente conectadas, além disso, as comunidades se tornaram mais relacionadas e menos dispersas, fazendo a modularidade da rede aumentar em relação ao grafo A. No entanto, ainda persistem alguns grupos desconexos dentro da rede, como o verde e vermelho.

### 5. Considerações Finais

Este trabalho teve por objetivo caracterizar a rede de coautoria dos cursos de Ciência da Computação, Sistemas de Informação e Engenharia de Software do CEULP/ULBRA utilizando os artigos publicados entre os anos de 2015 a 2020 no ENCOINFORM. Para obter os resultados para análise foi necessário extrair e tratar os dados do evento, em seguida gerar gráficos, tabelas e grafos para visualização.

Desta forma, foi possível extrair indicadores que revelam como o campo de pesquisa da instituição evolui e como ele está sendo gerenciado. A visão ampla obtida traz facilidade na busca de informações de quais os coautores mais produtivos, quais têm maiores ligações ou ainda quais têm maior repertório de temas científicos. Além disso, possibilita direcionar o planejamento e os investimentos do campo de pesquisa na instituição, indicando onde falta o incentivo e programas pedagógicos.

De forma geral, a caracterização dessa rede de coautoria fornece *insights* e um agregado de informações para os gestores da instituição e cursos citados, o que irá contribuir na tomada de decisão em relação ao campo científico. Como trabalhos futuros sugere-se a ampliação da rede de coautoria para incluir outros cursos do CEULP/ULBRA e outros eventos de interesse da instituição. Ainda seria interessante caracterizar a rede a partir de mais indicadores, como:

componentes conectados, transitividade e distância geodésica, para que os resultados permitissem maior abrangência e assertividade.

### Referências

ALEJANDRO, Velasquez A. O.; NORMAN, Aguilar G. **Manual introdutório à Análise de Redes Sociais**. 2005. Disponível em: <[http://www2.unicentro.br/lmqqa/files/2016/05/Manualintrodutorio\\_ex\\_ucinet.pdf](http://www2.unicentro.br/lmqqa/files/2016/05/Manualintrodutorio_ex_ucinet.pdf)>.

ANDRADE, Ricardo L. de. **A Influência das Redes de Coautoria na Performance dos Bolsistas de Produtividade e nos Programas de Pós- Graduação em Engenharia de Produção**. 2016. 145 f. Dissertação (Mestrado) - Curso de Engenharia de Produção, Universidade Federal de Pernambuco, Recife, 2016.

BALANCIERI, Renato. **Análise de redes de pesquisa em uma plataforma de gestão em ciência e tecnologia**: uma aplicação à plataforma Lattes. 2004. Dissertação (Mestrado em Engenharia de Produção) – Programa de Pós-graduação em Engenharia de Produção – Universidade Federal de Santa Catarina. 2004.

BEAVER, Donald. B.; ROSEN, R. Studies in scientific collaboration: part I: The professional origins of scientific co-authorship. **Scientometrics**, Amsterdam, v.1, n.1, p.65-84, 1978.

BUFREM, L. S.; GABRIEL JUNIOR, R. F.; SORRIBAS, T. V. Redes sociais na pesquisa científica da área de ciência da informação. **DataGramaZero**, v. 12, n. 4, 2011. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/7410>. Acesso em: 18 set. 2021.

CHERVEN, Ken. **Mastering Gephi Network Visualization**: produce advanced network graphs in gephi and gain valuable insights into your network datasets. Birmingham: Packt Publishing, 2015.

DEGENNE, Alain; FORSÉ, Michel. **Introducing Social Networks**. London: Sage, 1999.

ENCOINFO: **Apresentação**. 2021. Disponível em: <http://ulbra-to.br/encoinfo/apresentacao/>. Acesso em: 26 set. 2021.

GRANDJEAN, Martin. La connaissance est un réseau. 2014. Disponível em: [https://commons.wikimedia.org/wiki/File:Social\\_Network\\_Analysis\\_Visualization.png](https://commons.wikimedia.org/wiki/File:Social_Network_Analysis_Visualization.png). Acesso em: 30 set. 2021.

MUELLER, Suzana P. M.; PASSOS, Edilenice. **As questões da comunicação científica e a ciência da informação**. Brasília: Departamento de Ciência da Informação - Universidade de Brasília, 2000. NEWMAN, M. E. J. **The structure of scientific collaboration networks**. Santa Fe Institute. Santa Fe, p. 404-409. 16 jan. 2001. Disponível em: <<http://www.pnas.org/content/98/2/404.full.pdf>>. Acesso em: 20 jul. 2016.

OLIVEIRA, Walison A. de. **Colaboração científica nos programas de pós-graduação em educação**: uma análise de redes de coautoria. 2017. Dissertação (Mestrado em Educação) - Universidade Federal de São Carlos, São Carlos, 2017.

PARREIRAS, F. S.; SILVA, A.; MATHEUS, R. F.; BRANDÃO, W. C. Redeci: colaboração e produção científica em ciência da informação no brasil. **Perspectivas em Ciência da Informação**, v. 11, n. 3, 2006. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/35765>. Acesso em: 18 set. 2021.

REICHERT, Taís; NOVAES, Tiago; RECH, Munique; OLEA, Pelayo. Redes de Colaboração Científica: Um Estudo de Coautoria Através da Análise de Redes Sociais. **Mostra de Iniciação Científica, Pós-graduação, Pesquisa e Extensão**. 5. 1-15, 2016.

SILVA, Alzira K. A. da. **Redes de coautoria em Ciência da Informação no Brasil: dinâmica na produção científica dos atores mediada pela Ancib**. 2012. 252 f. Tese (Doutorado) - Curso de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2012.

SIMÕES-PEREIRA, J. M. S.. **Grafos e Redes: Teoria e Algoritmos básicos**. Rio de Janeiro: Interciência, 2013.

SONNENWALD, D. H. Scientific Collaboration. **Annual Review of Information Science and Technology**, v.42, n.1, p.643-681, 2008.

TOMAÉL, Maria Inês. **Redes de conhecimento: o compartilhamento da informação e do conhecimento em consórcio de exportação do setor moveleiro**. Belo Horizonte, 2005. 289f. Tese (Doutorado em Ciência da Informação) - Universidade Federal de Minas Gerais. Belo Horizonte, 2005.

VANZ, Samile A. de S.; STUMPF, Ida R. C. Colaboração científica: revisão teórico-conceitual. **Perspectivas em Ciência da Informação**, Belo Horizonte, v.15, n.2, p.42-55, maio/ago. 2010. Disponível em: <<https://www.scielo.br/j/pci/a/Fz4q6DhPGhjnhxXmRxLw6Ct/?lang=pt>>.

WASSERMAN, Stanley; FAUST, Katherine. **Social network analysis: methods and applications**. New York: Cambridge University Press, 1994. p. 857.



## CARACTERIZAÇÃO DA REDE DE COAUTORIA DOS CURSOS DE CIÊNCIA DA COMPUTAÇÃO, SISTEMAS DE INFORMAÇÃO E ENGENHARIA DE SOFTWARE DO CEULP/ULBRA A PARTIR DA PLATAFORMA LATTES

Caio Vinicius Davantel<sup>1</sup>, Fabiano Fagundes<sup>2</sup>

<sup>1</sup>Acadêmico do curso de Engenharia de Software no Centro Universitário Luterano de Palmas – CEULP/ULBRA. E-mail: caiodavantel@gmail.com.

<sup>2</sup> Mestre em Ciência da Computação pela UFSC. Professor dos cursos de Ciência da Computação, Sistemas de Informação e Engenharia de Software do CEULP/ULBRA. E-mail: thilfa@gmail.com.

### RESUMO

Este trabalho tem como propósito caracterizar a rede de coautoria dos cursos de Ciência da Computação, Sistemas de Informação e Engenharia de Software do CEULP/ULBRA, utilizando os dados disponibilizados pela Plataforma Lattes. Foram desenvolvidos e estruturados grafos, tabelas e gráficos com as ferramentas Microsoft Excel e Gephi e, na sequência, aplicaram-se parâmetros de análise de redes de coautoria para obter os resultados. Desta forma, foi possível extrair indicadores que revelam como o campo de pesquisa evoluiu e como ele está sendo gerenciado, fornecendo *insights* e um agregado de informações para os gestores da instituição e cursos citados, contribuindo na tomada de decisão em relação ao campo científico.

**PALAVRAS-CHAVE:** Redes de coautoria, coautoria científica, grafos.

### 1 INTRODUÇÃO

Segundo Vanz e Stump (2010), a coautoria científica é reconhecida a partir do trabalho conjunto de dois ou mais cientistas em projetos de pesquisas, que por consequência compartilham recursos intelectuais, econômicos e físicos. Pode ser definida como uma interação ocorrendo dentro de um contexto social entre dois ou mais cientistas que facilita o compartilhamento de significados e conclusão de tarefas, no que diz respeito a um mutuamente objetivo superordenado e compartilhado (SONNENWALD, 2007).

Um conjunto de colaborações forma uma rede de coautorias que, segundo Silva (2012), contém estratégias para compartilhar informações e conhecimentos através das comunidades científicas, obtendo benefícios como uma maior qualidade dos artigos e um incremento do número de publicações. Além disso, o estudo das redes de coautoria permite identificar e analisar porque os pesquisadores colaboram e quais as consequências dessas colaborações. Pesquisas apontam a melhoria da qualidade dos artigos e o aumento do número de publicações como benefícios da colaboração (ANDRADE, 2016).

Uma Rede de coautoria pode ser estudada através de grafos, que têm sido amplamente utilizados para representar as redes sociais e quantificar a sua estrutura. Além disso, na opinião de Bufrem, Gabriel Junior e Sorribas (2011), a visualização por meio desse modelo é considerada mais intuitiva do que a visualização na forma de matrizes, embora os próprios autores afirmem que os dados coletados sejam, normalmente, apresentados dessa forma. Como os grafos são baseados na matemática, a partir de sua análise, pode-se obter indicadores que possibilitam a caracterização das redes de coautorias, dessa forma, Silva (2012), argumenta que as pesquisas que envolvem esse tema, em sua maioria analisam os seguintes parâmetros: grau de

centralidade, a colaboração, a produtividade dos autores e o elitismo da área, a densidade, a transitividade e o tamanho da rede, a distribuição dos laços e cliques.

Diante dos dados importantes que são fornecidos pela Plataforma Lattes - projeto do Conselho Nacional de Desenvolvimento Científico e Tecnológico do Brasil (CNPq), mostra-se interessante caracterizar as redes de coautorias dos cursos de Ciência da Computação, Sistemas de Informação e Engenharia de Software do CEULP/ULBRA. Visto que, identificar como as redes estão se desenvolvendo dentro das Instituições de Ensino Superior auxilia no desenvolvimento pedagógico, na medida em que a formação de pesquisadores parte delas (REICHERT et al., 2016). Assim, este trabalho tem como objetivo coletar os dados dos pesquisadores disponibilizados pela Plataforma Lattes e criar grafos que organizem as redes de coautorias, sendo assim, será possível obter métricas que contribuam para o entendimento e direcionamento das pesquisas nos cursos citados.

## 2 MATERIAL E MÉTODOS

A Plataforma Lattes representa a experiência do CNPq na integração de bases de dados de Currículos, de Grupos de Pesquisa e de Instituições em um único Sistema de Informações. O Currículo Lattes tem grande importância, registrando a vida acadêmica e profissional dos estudantes e pesquisadores do país, sendo adotado pela maioria das instituições de fomento, universidades e institutos de pesquisa do País (PLATAFORMA LATTES, 2021). Sabendo da riqueza de informações e confiabilidade desta plataforma, todos os dados dessa pesquisa foram extraídos a partir dessa fonte documental eletrônica.

Para a extração dos currículos Lattes dos atuais e ex-professores dos cursos analisados foi utilizado Coletaprod. É um software de código aberto, disponibilizado na plataforma GitHub, e foi desenvolvido por Murakami, bibliotecário da Universidade de São Paulo (USP), com o intuito de facilitar a coleta de dados científicos, através de uma interface simples que fornece um catálogo e filtro para pesquisas de registros.

A ferramenta Gephi foi utilizada na obtenção dos parâmetros e grafos que foram analisados e mensurados neste trabalho. Este software, objeto de estudo de Cherven (2015), é usado para exploração e compreensão de redes. Oferece inúmeras métricas para análise de redes, como: grau, diâmetro, densidade, coeficiente de cluster, modularidade, componentes conectados, excentricidade e métricas de centralidade (ANDRADE, 2016). Também foi utilizado o software Microsoft Excel, para obter tabelas e dados estatísticos; desta forma, foi possível realizar a análise dos dados quantitativos e mensurar a rede de coautoria.

A partir dos currículos disponibilizados na Plataforma Lattes, dos atuais e ex-professores dos cursos já citados, foi definido como principal corpus desta pesquisa, as publicações completas dentro do intervalo que vai do ano de 2015 ao ano de 2020, e que estão relacionadas aos eventos científicos que fazem parte do plano de ensino do CEULP/ULBRA. Os caminhos metodológicos seguidos no trabalho podem ser elencados da seguinte forma:

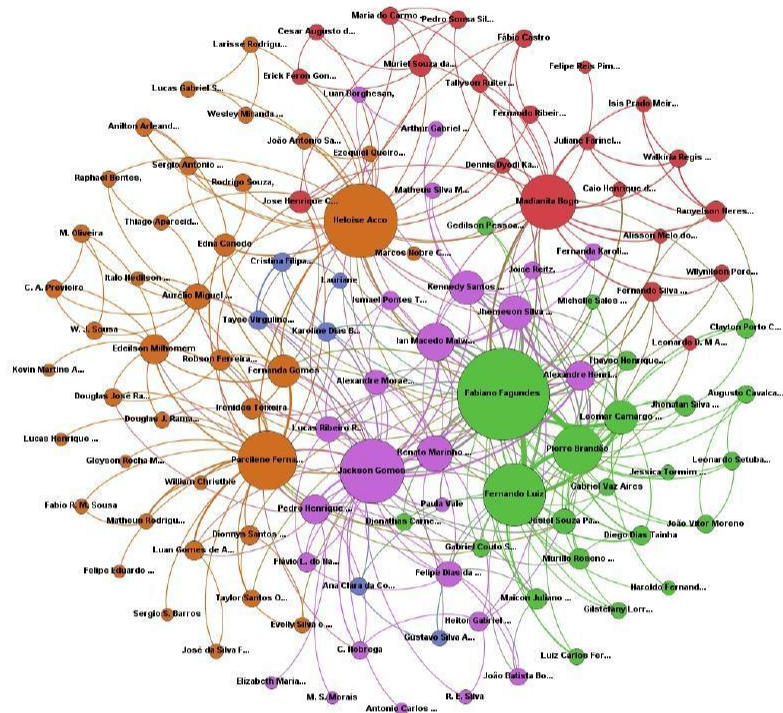
- a) Coleta dos dados realizada através da ferramenta Coletaprod;
- b) Tratamento dos dados coletados utilizando o software Microsoft Excel;
- c) Geração dos grafos com base nos resultados obtidos por meio do software Gephi;
- d) Criação dos gráficos e tabelas estatísticas com a utilização do software Microsoft Excel, tendo como base o grafo de rede coautoria gerado pela ferramenta Gephi;
- e) Caracterização e interpretação dos resultados obtidos por meio dos grafos da rede de coautoria, gráficos e tabelas estatísticas.

## 3 RESULTADOS E DISCUSSÃO

Os resultados desta seção foram obtidos através dos métodos descritos anteriormente. O primeiro dado estatístico se refere aos números totais de artigos, coautores e ligações entre eles, provenientes dos artigos publicados pelos professores na Plataforma Lattes entre os anos de 2015 e 2020. É importante destacar que neste primeiro momento o número de coautores passou por um filtro, onde foram excluídos os nomes que se repetiam em anos diferentes, como exemplo, o coautor Fabiano Fagundes, que se repetiu em todos os anos analisados pela pesquisa. Dessa forma, para evitar ruídos na contagem absoluta dos dados foi realizada essa exclusão, sendo contabilizado somente uma vez cada coautor.

A Figura 1, a seguir, representa o grafo da rede de coautoria do *corpus* desta pesquisa. Ele foi analisado segundo os indicadores de densidade, grau de centralidade, grau de intermediação e modularidade. Desta forma, foi obtida uma densidade de 0.057, o que demonstra uma rede com pouca conectividade e ligações fracas entre os nós. Essa condição é evidenciada no grafo através do grande número de nós pequenos e afastados do centro da rede. Além disso, com exceção dos coautores mais ao centro da rede, as ligações são predominantemente tênues.

**Figura 1. Grafo da rede de coautoria segundo os artigos publicados pelos professores na Plataforma Lattes nos anos de 2015 a 2020.**



**Fonte: Autor**

Os professores dos cursos analisados destacam-se na rede como coautores assíduos por estarem presentes em quase todos os artigos, sendo assim, eles detêm os maiores graus de centralização. Além disso, alguns professores externos aos cursos, como o Pierre Brandão, também recebem bastante destaque na rede. Consequentemente, os nós intermediários também são os docentes, pois além de estarem mais ao centro, também possuem maior nível de diversificação do campo de estudo, o que gera coautorias distintas e mais ligações na rede. Já os nós mais afastados, tendem a formar conectividades menores, mas permanecem dentro de grupos da mesma intenção de pesquisa científica.

Foi identificado, de acordo com a modularidade, um total de cinco comunidades, definidas pela proximidade de seus vizinhos e a densidade de ligação entre eles. Desta forma, quanto mais espessa a aresta no grafo, maior também a ligação entre os nós, ou seja, os coautores possuem mais artigos publicados juntos. A partir disso, pode-se evidenciar o nó centralizador de cada grupo, responsável pelos maiores números de coautorias e diversificação de estudos científicos, além disso, tanto a comunidade que ele se insere, como também a rede, são amplamente influenciadas pelas suas produções, visto que, resulta no aumento de ligações e conectividade.

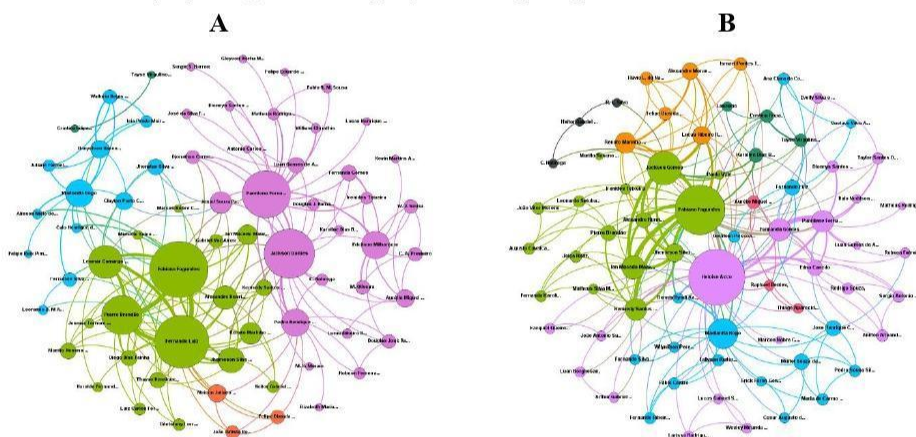


Vale destacar que a rede é dominada, com mais de 75% de seu tamanho, pelas comunidades laranja, roxo e verde, indicando, assim, existir um grande fluxo de produção científica e coautoria entre elas. Por outro lado, a pequena porcentagem restante é composta pela vermelha e azul, sendo essa última extremamente pequena em relação às outras comunidades, contendo apenas seis nós e poucas ligações entre eles.

A Figura 2 apresenta dois grafos, A e B, representando, respectivamente, as redes de coautoria dos anos de 2015/16/17 e 2018/19/20. Diante disso, foi possível obter alguns resultados comparativos entre os dois grafos. No grafo A, tem-se um total de 69 nós e 182 ligações, e os graus de centralização e intermediação são amplamente assegurados pelos professores dos cursos em análise.

Além disso, a densidade da rede é de 0.078, evidenciando uma rede com ligações fortes somente entre os nós centrais e pouca ligação distribuída, significando uma baixa variedade de coautorias. Também é possível verificar a existência de 5 comunidades que se comunicam somente pelos seus nós centrais. O grafo B tem um total de 71 nós e 208 ligações. Apesar dos maiores graus de centralização e intermediação da rede ainda serem dos professores, a densidade da rede aumentou minimamente, em comparação ao grafo anterior, sendo o valor agora de 0.084.

**Figura 2. Grafo A e B representando respectivamente as redes de coautorias dos anos de 2015/16/17 e 2018/19/20 segundo os artigos publicados pelos professores na Plataforma Lattes.**



Fonte: Autor

Pode-se observar, no entanto, que este aumento não foi suficiente para que se alterasse a conectividade dos nós, pois ainda predomina as ligações mais fortes nos coautores centrais, consequentemente, existindo uma baixa distribuição de colaborações científicas. No entanto, é possível verificar um aumento no número de comunidades, sendo agora um total de sete, correlacionando com o aumento de produções científicas e coautorias desse período.

#### 4 CONCLUSÃO

Este trabalho teve por objetivo caracterizar a rede de coautoria dos cursos de Ciência da Computação, Sistemas de Informação e Engenharia de Software do CEULP/ULBRA utilizando os artigos publicados entre os anos de 2015 a 2020 na Plataforma Lattes, pelos atuais e ex-docentes dos cursos em questão. Para obter resultados satisfatórios foi necessário extrair e tratar os dados de eventos relacionados ao plano de ensino da instituição, em seguida gerar gráficos, tabelas e grafos para visualização.

Dessa forma, foi possível extrair indicadores que revelam como o campo de pesquisa da instituição evoluiu. A visão ampla obtida traz facilidade na busca de informações de quais os coautores mais produtivos, os que têm maiores ligações, quais têm maiores repertórios de temas

científicos, além disso, tende a direcionar o planejamento e os investimentos do campo de pesquisa na instituição, indicando onde falta o incentivo e programas pedagógicos.

De forma geral, a caracterização dessa rede de coautoria fornece *insights* e um agregado de informações para os gestores da instituição e cursos citados, o que irá contribuir na tomada de decisão em relação ao campo científico. Como trabalhos futuros sugere-se a ampliação da rede de coautoria para incluir outros cursos do CEULP/ULBRA.

## 5 REFERÊNCIAS

ANDRADE, Ricardo L. de. **A Influência das Redes de Coautoria na Performance dos Bolsistas de Produtividade e nos Programas de Pós- Graduação em Engenharia de Produção**. 2016. 145 f. Dissertação (Mestrado) - Curso de Engenharia de Produção, Universidade Federal de Pernambuco, Recife, 2016.

BUFREM, L. S.; GABRIEL JUNIOR, R. F.; SORRIBAS, T. V. Redes sociais na pesquisa científica da área de ciência da informação. **DataGramZero**, v. 12, n. 4, 2011. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/7410>. Acesso em: 18 set. 2021.

CHERVEN, Ken. **Mastering Gephi Network Visualization**: produce advanced network graphs in gephi and gain valuable insights into your network datasets. Birmingham: Packt Publishing, 2015.

PLATAFORMA LATTES. Cnpq (org.). **Sobre a plataforma Lattes**. Disponível em: <http://lattes.cnpq.br/web/plataforma-lattes/o-que-e>. Acesso em: 26 set. 2021.

REICHERT, Tais; NOVAES, Tiago; RECH, Munique; OLEA, Pelayo. **Redes de Colaboração Científica: Um Estudo de Coautoria Através da Análise de Redes Sociais**. Mostra de Iniciação Científica, Pós-graduação, Pesquisa e Extensão. 5. 1-15, 2016.

SILVA, Alzira K. A. da. **Redes de coautoria em Ciência da Informação no Brasil**: dinâmica na produção científica dos atores mediada pela Ancib. 2012. 252 f. Tese (Doutorado) - Curso de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2012.

SONNENWALD, D. H. Scientific Collaboration. **Annual Review of Information Science and Technology**, v.42, n.1, p.643-681, 2008.

VANZ, Samile A. de S.; STUMPF, Ida R. C. Colaboração científica: revisão teórico-conceitual. **Perspectivas em Ciência da Informação**, Belo Horizonte, v.15, n.2, p.42-55, maio/ago. 2010. Disponível em: <<https://www.scielo.br/j/pci/a/Fz4q6DhPGhjnXmRxLw6Ct/?lang=pt>>.