

# Identificação e Caracterização de Redes Científicas de Dados Curriculares

Thiago M. R. Dias<sup>1</sup>, Gray F. Moita<sup>1</sup>, Patrícia M. Dias<sup>1</sup>, Tales H. Moreira<sup>1</sup>

<sup>1</sup>Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG)  
Av. Amazonas, 7675 – Nova Gameleira – CEP: 30510-000 – Belo Horizonte – MG –  
Brasil

{thiogomagela,patriciamdias,talesinfo}@gmail.com, gray@dppg.cefetmg.br

**Abstract.** *The analysis of social networks has been the focus of several studies in different areas of knowledge. Mainly scientific collaboration networks, where data on productivity and intensity of collaboration reveal important information for understanding the evolution of research. This article presents the process of data extraction from Lattes platform and molding the extracted data into an immense network of scientific collaboration. The main contributions are the methods of extracting the entire basis of curricula and propose an algorithm for the identification of collaborations with low computing costs, presenting itself as an excellent alternative for the analysis of large networks. With the modeled network, several techniques of social network analysis can be applied in order to obtain results that present a view of productivity and collaboration of all registered users on the platform.*

**Resumo.** *A análise de redes sociais tem sido foco de vários estudos nas diversas áreas do conhecimento. Principalmente as redes de colaboração científica, onde dados sobre produtividade e intensidade de colaboração revelam informações importantes para o entendimento da evolução das pesquisas. Este artigo apresenta o processo de extração de dados da Plataforma Lattes e a modelagem dos dados extraídos em uma imensa rede de colaboração científica. As principais contribuições do trabalho são um método de extração de toda a base de currículos e a proposta de um algoritmo para a identificação das colaborações com baixo custo computacional, se apresentando como uma excelente alternativa para a análise de grandes redes. Com a rede modelada, diversas técnicas de análise de redes sociais podem ser aplicadas com o intuito de obter resultados que apresentam uma visão da produtividade e colaboração de todos os usuários cadastrados na plataforma.*

## 1. Introdução

Esse artigo é uma versão estendida e atualizada de Dias et al. (2013), onde foram apresentados os resultados de um extrator de currículos da Plataforma Lattes, para a modelagem e caracterização de redes de colaboração científica. Nesta versão, é proposto um algoritmo para a identificação de colaboração em grandes bases de dados.

O algoritmo proposto apresenta baixo custo computacional se comparado a métodos que utilizam de técnicas como comparação de títulos de trabalhos publicados para identificar colaborações, apresentando resultados excelentes na base de dados analisada.

Uma rede pode ser caracterizada como um grafo, que consiste de um conjunto de nós (vértices) e ligações (arestas) entre os nós. Estas ligações podem ser, direcionadas ou não direcionadas, e podem, opcionalmente, ter um peso associado. Diversos fenômenos naturais podem ser descritos em termos de uma rede. O cérebro pode ser caracterizado como uma rede de neurônios ligados por sinapses. A Internet também é um exemplo de uma rede importante e amplamente estudada nos dias atuais.

Em redes rodoviárias as cidades e as rodovias que as interligam pode ser considerado como um grafo, onde as cidades se caracterizam por ser os vértices e as rodovias as arestas. Já em redes de comunicação, os computadores interligados por um meio de comunicação pode ser caracterizado como um grafo, onde vértices representam um computador e arestas são representações de ligações físicas entre eles [Liben-Nowell e Kleinberg, 2007].

Entre os vários tipos de redes, existem as redes sociais. Uma rede social é um conjunto de pessoas ou grupos que têm algum tipo de relação entre eles [Newman, 2001].

As redes sociais têm sido objetos de estudos há muitos anos. Em Barabási (2004) são apresentados diversos estudos que motivaram e agregaram valor a teoria dos grafos e que já se utilizava do conceito de redes sociais. Já Freire e Figueiredo (2011), revelam que as relações entre as pessoas podem ser de amizade, parentesco ou colaboração (por exemplo, em um artigo os coautores). Em uma rede social de amigos, a relação entre duas pessoas pode representar uma amizade entre elas. Em uma rede de relações de parentesco entre as pessoas pode indicar que duas pessoas, pertencem à mesma família.

No domínio científico, um exemplo de uma rede social é a de colaboração científica, que pode ser observada como um grafo no qual os vértices correspondem aos autores de publicações científicas e as arestas correspondem à relação de coautoria. Neste tipo de rede, as arestas podem ou não ser ponderadas. A adição de peso representa o número de trabalhos conjuntos que os autores relacionados pela aresta analisada participaram. Dessa forma, a intensidade dos relacionamentos presentes em uma rede de colaboração científica é medida pelo número de colaborações entre um par de autores. A presença do peso é útil representar, por exemplo, a afinidade e interesses comuns entre dois autores da rede [Sonnenwald, 2007].

Newman (2001) apresenta uma avaliação extensa sobre características sociais das redes de coautoria em redes científicas de computação, biologia, física e medicina no período de 1995 a 1999, encontrando diversas características sociais que são comuns a essas redes. Já em Newman (2004), o autor faz análises de coautoria para identificar propriedades estatísticas, procurando por padrões comuns entre os elementos da rede analisada.

É proposto neste trabalho, um método com baixo custo computacional para a construção de grandes redes de colaboração científica, a partir de dados extraídos de um repositório de currículos de pesquisadores e profissionais de diversas áreas de pesquisa. Para caracterizar a relação de colaboração entre pares de pesquisadores, são analisados

trabalhos que dois ou mais pesquisadores realizaram em conjunto. Estes trabalhos podem ser de diversos tipos como artigos publicados em conjunto e orientações de trabalhos ou projetos. Neste trabalhos, são considerados artigos e resumos publicados em anais de congressos e artigos publicados em periódicos para a identificação das colaborações.

## **2. Trabalhos Relacionados**

Com a competição cada vez mais acirrada entre organizações e instituições de pesquisa, torna-se importante para seus integrantes, descobrirem potenciais colaboradores com o intuito de realizarem trabalhos em colaboração objetivando impulsionar a sua produção científica e obter melhores resultados em suas pesquisas. É possível apontar trabalhos recentes onde é mostrado que grupos de pesquisa com uma rede social científica bem conectada geralmente tendem a ser mais produtivos [Lopes et al., 2011; Brandão e Moro, 2012].

As redes de co-autoria de uma comunidade podem revelar fatos interessantes sobre eles, como os grupos que colaboram melhor, a intensidade das relações entre autores ou autores que trabalham com um maior grau de colaboração. O estudo de redes de co-autoria, também pode ser usado para comparar os padrões de colaboração entre diferentes comunidades científicas das diversas áreas do conhecimento [Junior et al., 2011].

No trabalho de Petersen et al. (2012), são analisados fatores relevantes para o sucesso científico em redes de colaboração. Dentre eles são destacados, a quantidade de produção científica que aumenta a atração de novas oportunidades e também tamanho da equipe de colaboradores, bem como, da própria rede. Diante disto, justifica-se o grande potencial de novos estudos na área, objetivando entender como a colaboração científica acontece e como ela vem evoluindo em cada uma das áreas de pesquisa.

No trabalho de Dias e Moita (2014) são geradas redes entre autores de trabalhos científicos onde as ligações entre pares de autores são caracterizadas por palavras-chave que estes autores tenham utilizados em algum de seus trabalhos. Com isto, é possível identificar autores que tem publicado trabalhos cujo conteúdo se assemelham baseado na análise das palavras chave que tenham utilizados. Logo, ligações mais densas, geralmente representam autores com interesses semelhantes.

Outros trabalhos de pesquisa em análise de redes científicas podem ser vistos em (Dias e Moita, 2013; Alvarenga et al., 2012; Freire e Figueiredo, 2011; Oliveira et al., 2011; Mena-Chalco et al., 2009; Reijers et al., 2009). Estas propostas baseiam-se no potencial de extração de conhecimento sobre pesquisadores, instituições e grupos de pesquisa, a partir de suas produções científicas, principalmente artigos científicos produzidos pelos pesquisadores.

## **3. Identificação e Caracterização**

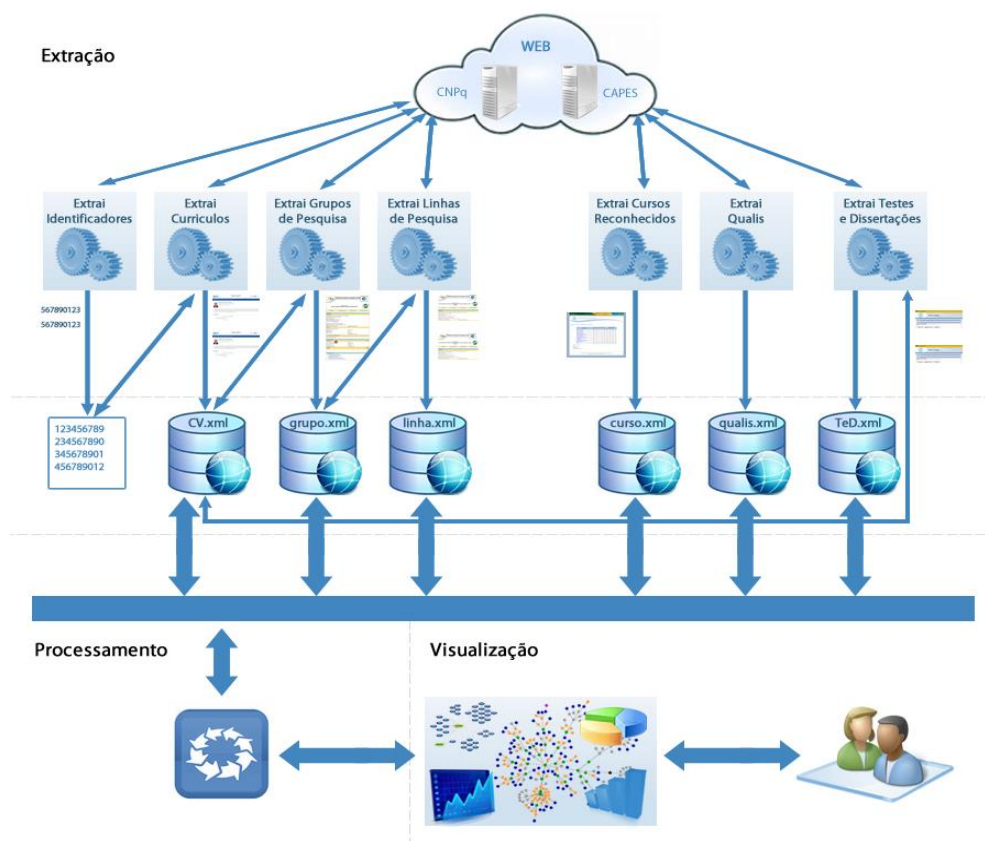
Para a construção das redes de colaboração científica deste trabalho, foi utilizado a Plataforma Lattes como principal fonte de informação. A Plataforma Lattes, criada em 1997 pelo CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico)

com participação do Grupo Stela do PPGE/UFSC, tornou-se grande conhecida da comunidade científica e tecnológica no Brasil, como também, em países latino-americanos e europeus. A Plataforma Lattes foi concebida para integrar os sistemas de informação das agências federais, racionalizando o processo de gestão de Ciência e Tecnologia (C&T), tanto do ponto de vista do usuário quanto das agências de fomento e das instituições de ensino e pesquisa [CNPQ, 2014].

Os currículos que fazem parte da Plataforma Lattes se tornaram um padrão nacional utilizado para avaliação individual das atividades científicas e acadêmicas, já que eles agregam dados de pesquisadores de todas as áreas do conhecimento, tornando a Plataforma Lattes uma fonte extremamente extensa e valiosa para analisar e compreender o comportamento de grupos de pesquisa [Digiampietri et al., 2012].

Diversos trabalhos para análise de dados científicos tem explorado a Plataforma Lattes como principal fonte de dados [Alves et al. 2011; Alves et al. 2011(a); Alves et al. 2011(b); Dias et al. 2013; Fernandes et al. 2011; Farias et al. 2012; Mena-Chalco et al. 2012].

Neste trabalho, são utilizadas tecnologias envolvidas no processo de recuperação de dados da web para realizar a extração de todos os currículos da Plataforma Lattes, bem como dados dos grupos e das linhas de pesquisa do CNPq. Apesar de existir a possibilidade de extração outros dados como o Qualis de periódicos e conceitos de cursos avaliados, o principal elemento de análise neste projeto são os currículos cadastrados na plataforma. Figura 1.



**Figura 1 – Arquitetura da Plataforma de Extração e Integração**

Todo o processo de extração e integração dos dados é dividido em três partes principais denominadas de Extração, Processamento e Visualização. Porém, neste trabalho só foram utilizados os resultados da etapa de extração dos currículos Lattes, tendo em vista, que a identificação das colaborações utiliza-se dos dados das publicações cadastradas em cada um dos artigos, como título e autores.

O processo de extração dos currículos Lattes, principal objeto de estudo deste trabalho, se inicia pela aquisição dos códigos dos currículos Lattes obtidos com uma requisição na interface de consulta da Plataforma Lattes, para que desta forma os identificadores possam ser armazenados localmente. A estratégia para aquisição se inicia com uma solicitação que resulta em uma lista contendo todos os códigos de identificação dos currículos cadastrados.

De posse da listagem contendo todos os identificadores, o extrator possibilita selecionar uma quantidade de currículos a serem extraídos, extrair os currículos a partir de uma determinada data de atualização registrada em cada currículo, ou de todo o repositório.

Todos os currículos extraídos são armazenados em disco em formato *XML* (*eXtensible Markup Language*). Na etapa de extração, várias falhas estruturais que estão nos currículos são corrigidas. Erros estes que dificultam o processo de obtenção dos dados. Exemplos destes problemas são cursos de graduação e pós-graduação que estão informados como concluídos, mas que não possuem ano de conclusão, cursos de pós-graduação em andamento que não possuem ano de início, trabalhos publicados que não possuem ano de publicação, dentre outros. Importante ressaltar que neste caso, tratamento de exceções foram desenvolvidos para contornar estes problemas e permitir que os arquivos sejam salvos normalmente e com o máximo de consistência possível.

Além dos currículos, também é possível extrair dados de outros repositórios como os grupos e as linhas de pesquisa, bem como, dados de outras fontes com o intuito de complementar os dados dos currículos. No entanto, para este trabalho só foram utilizados os dados curriculares. Conforme mencionado, as etapas de processamento e visualização, responsável por realizar a integração das fontes extraídas e oferecer uma interface de visualização dos dados não foram utilizadas. Tendo em vista que para a identificação das colaborações, os dados dos artigos publicados são suficientes para o propósito deste trabalho.

Com isto, diante do conjunto de currículos extraídos, é possível realizar diversos tipos de análises bibliométricas, já que todos os dados dos currículos estão armazenados em arquivos estruturados. Tendo em vista a grande quantidade de dados, um grande desafio para a compreensão de como acontece a colaboração científica nacional passa a ser a identificação de colaborações.

Para a identificação das colaborações tendo em vista a grande quantidade de autores e publicações a serem analisados, um algoritmo que tenha baixo custo computacional e com alta taxa de precisão foi proposto. O intuito é que seja possível realizar a identificações de colaborações científicas de grande quantidade de dados em tempo hábil. Algoritmo 1.

### Identification-Collaboration

1.  $n \leftarrow$  number of articles author
2. **for**  $i \leftarrow 1$  **to**  $n$
3.  $x \leftarrow$  *string*[ $i$ ] //  $x$  is article title [ $i$ ]
4.  $x \leftarrow$  *stopword*[ $x$ ] // removes token without semantic value
5.  $x \leftarrow$  *normalization*[ $x$ ] // remove whitespace and accentuation
6.  $x \leftarrow$  *lowercase*[ $x$ ]
7. **if** *hash*[ $x$ ] **in** *dictionary* // checks whether  $x$  is in the dictionary
8.  $\text{dictionary}[x] \leftarrow \text{id\_author}$
9. **else**  $\text{dictionary} \leftarrow x, \text{id\_author}$

### **Algoritmo 1 – Algoritmo para identificação de colaboração**

Para realizar as identificações das colaborações entre os autores, cada título de um trabalho cadastrado em um determinado currículo passa por uma transformação que tem como objetivo obter o título sem as palavras que não tenham valor semântico, sem acentuação e sem os espaços entre elas. Esta estratégia tem como intuito minimizar a ocorrência de erros gramaticais que podem estar inseridos nos títulos dos artigos. Consequentemente, todo o texto é padronizado em letras minúsculas e a *string* resultante é concatenada com o ano da publicação para posteriormente ser transformada em uma chave que representa o trabalho em análise, passos 2 a 6 do algoritmo.

Após a transformação verifica-se no dicionário utilizado para a caracterização da rede de colaboração se a chave já está presente. Caso a chave já exista no dicionário, o identificador do autor do currículo em análise é vinculado à chave, caso contrário, são inseridos a chave e o identificador no referido dicionário. Tabela 1.

<b>Chave do Dicionário</b>	<b>Autores</b>
modelagemcaracterizacaoredescientificasestudosobreplataformalattes2013	Id01, Id25
studyaboutinfluenceacademiperformancesstudentsuserssocialnetworks2013	Id25, Id175, Id98
analysiscollaborationnetworksscificpublications2013	Id01, Id98, Id67
...	
...	
processoparaidentificacaocolaboradoresredescientificas2013	Id01, Id28, Id174

**Tabela 1 – Exemplo de dicionário construído pelo algoritmo de identificação**

Importante destacar que cada um dos currículos cadastrados na Plataforma Lattes possui identificador único. Este identificador é utilizado tanto para caracterizar o usuário da plataforma como também para permitir o acesso ao currículo individual de cada usuário. Uma dificuldade encontrada pelos autores no cadastro de uma publicação está no momento da citação de seus colaboradores. Devido à falta de padronização e possibilidade de ambiguação no nome de citação dos colaboradores de um trabalho, não é possível garantir que o nome informado na lista de colaboradores pertença a somente

um usuário da plataforma, o que dificulta a caracterização da rede de colaboração pelos nomes de citação informados em cada uma das publicações cadastradas.

Diante disso, a Plataforma Lattes permite que um autor vincule seus coautores pelos seus identificadores. Isso possibilita que, ao contrário de simplesmente informar um nome de citação de um colaborador, seja informado o nome vinculado ao identificador desse colaborador. Logo, quando há a citação deste colaborador, além do nome de citação, o seu código de identificação também é incorporado à publicação cadastrada.

No entanto, nem sempre são realizados estes vínculos, isto ocorre em sua maioria por não ser um processo automático na atualização dos currículos, ou seja, o autor deve vincular seus colaboradores a seus identificadores manualmente. Porém, quando é possível registrar este vínculo faz com que o método de identificação proposto tenha melhores resultados do que outros métodos geralmente utilizados para identificação de colaborações, pois mesmo que apenas um dos autores tenha cadastrado um determinado título, é possível identificar a colaboração, diferentemente dos métodos que trabalham com validações cruzadas.

Para isto, quando o identificador de um co-autor está inserido em uma publicação de um determinado autor que esteja sendo analisada, na inserção desta publicação no dicionário de chaves e identificadores, não apenas o identificador do autor é inserido mas também de todos os co-autores que estiverem informados. Isto faz com que o algoritmo obtenha alta taxa de revocação, já que é possível identificar colaborações mesmo que apenas um dos autores tenha cadastrado a publicação em seu currículo, diferentemente de outros métodos.

Para efeito de comparação, foram utilizados os currículos dos professores integrantes do Programa de Pós Graduação em Modelagem Matemática e Computacional do CEFET-MG (PPGMMC-CEFET MG). Logo, este grupo foi utilizado para avaliar o algoritmo proposto em comparação com uma solução que utiliza a estratégia de comparação entre títulos de artigos. A tabela 2 apresenta os resultados dos testes, bem como o grau relação de relacionamento. O grau real corresponde a quantidade de colaboradores de um determinado professor, calculado manualmente.

A Tabela 2 apresenta o grupo analisado, além dos graus de cada pesquisador que estão representados por identificadores. São apresentados para cada pesquisador o grau real obtido manualmente, o grau obtido pela aplicação do método de comparação de títulos e pelo algoritmo proposto. O grau de um vértice indica a quantidade de ligações de cada pesquisador, ou seja, vínculos que cada vértice do grafo, aqui representado pelos pesquisadores, possui com outros.

Como resultado, o algoritmo obteve melhores resultados já que conseguiu ter uma precisão de 100% nas identificações, ou seja, todas as colaborações indicadas são colaborações reais e um percentual de 97,61% de revocação, superando os 95,23% do método utilizado para fins de comparação. A revocação indica o percentual de colaborações reais que foi possível identificar.

<b>Grupo de Pesquisadores do PPGMMC-CEFET MG</b>			
<b>ID_Pesquisador</b>	<b>Grau Real</b>	<b>Algoritmo Comparação de Títulos</b>	<b>Algoritmo Proposto</b>
1	0	0	0
2	1	1	1
3	0	0	0
4	2	2	1
5	7	7	7
6	5	3	5
7	6	5	5
8	0	0	0
9	0	0	0
10	5	5	5
11	4	4	4
12	6	6	6
13	6	6	6
14	6	6	6
15	3	2	3
16	2	2	2
17	2	2	2
18	3	3	3
19	4	4	4
20	0	0	0
21	5	5	5
22	2	2	2
23	7	7	7
24	0	0	0
25	8	8	8

**Tabela 2 – Análise entre algoritmos de identificação de colaborações**

A grande vantagem da proposta é que como cada título de uma publicação é analisado uma única vez, faz com que o algoritmo tenha custo linear  $O(n)$ , tendo em vista que não são feitas comparações com outros título, a exemplo de outras técnicas onde são comparados todos os títulos com os outros, com o objetivo de encontrar títulos iguais e caracterizar um trabalho em colaboração, tendo seu custo como  $O(n^2)$ .

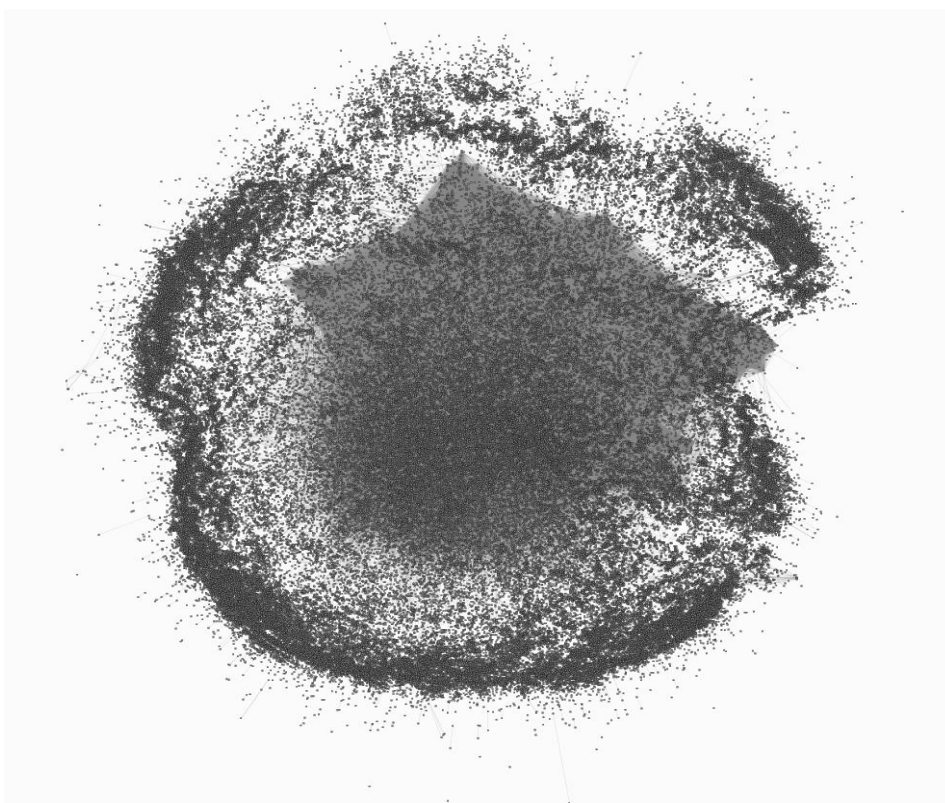
#### **4. Análise dos Resultados**

A Plataforma Lattes é uma fonte extremamente rica de informações. O grande volume de dados presente nos currículos podem ser trabalhados objetivando fornecer



informações valiosas até então desconhecidas. Em Abril de 2014 a Plataforma Lattes alcançou a marca de 3.600.000 currículos cadastrados.

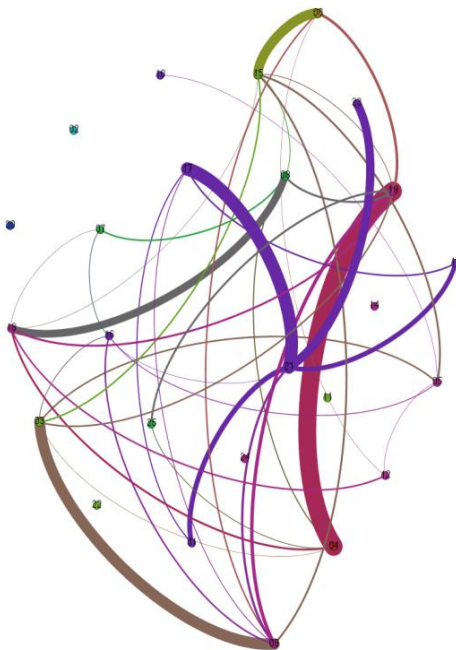
Como resultado do algoritmo de identificação de colaborações científicas descrito anteriormente, é possível gerar diversas redes, podendo especificar o conjunto de currículos a serem analisados. Exemplo da rede que contém todos os currículos da plataforma pode ser observada na Figura 2.



**Figura 2 – Rede com currículos coletados em 2/2013 (Dias et al. 2013)**

Nesta rede, os autores são representados por nós e as arestas são caracterizadas por trabalhos que um par de autores tenham realizados em conjunto. Com a rede gerada, podem ser aplicadas diversas métricas de análise de redes sociais que permitiram obter informações relevantes sobre como estão estruturados os grupos de pesquisa, como os pesquisadores de uma instituição colaboram entre si e entre pesquisadores de outras instituições, além de identificar aqueles elementos que possuem maior grau de conectividade e analisar sua produção.

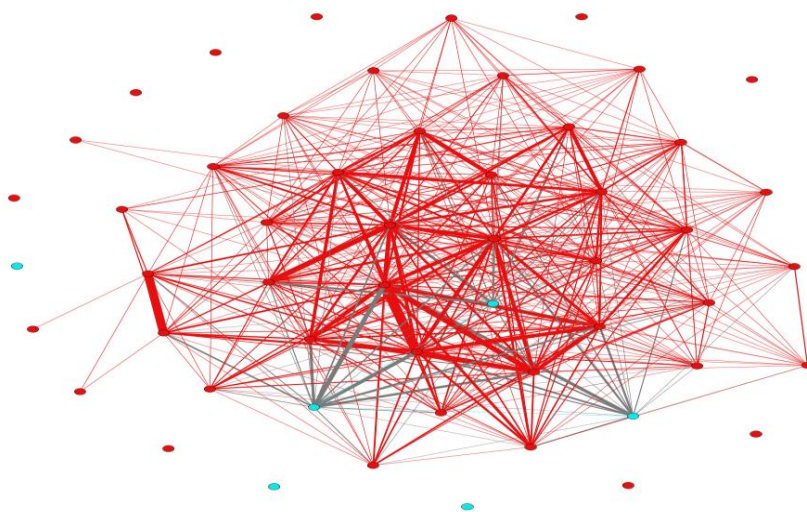
É possível ainda a geração de redes específicas tendo como base informações registradas nos currículos, como área de atuação, formação acadêmica, instituição de ensino ou fluência em um determinado idioma. Para isto é possível realizar uma busca na base de currículos por um ou vários termos inseridos nas seções que compõem os arquivos, com o uso de linguagem de consulta em arquivos XML. Desta forma podem ser geradas redes específicas como por exemplo de uma determinada instituição ou programa de pós graduação. Figura 3.



**Figura 3 – Rede de professores do PPGMMC CEFET MG**

Nesta rede, cada um dos nós representa um professor do PPGMMC CEFET MG e as arestas representam trabalhos realizados em colaboração. A espessura da aresta corresponde a quantidade de trabalhos que professores realizaram em conjunto, neste caso é possível identificar colaboradores mais frequentes e nós isolados, neste caso, correspondem a nós que nunca realizaram trabalhos em colaborações.

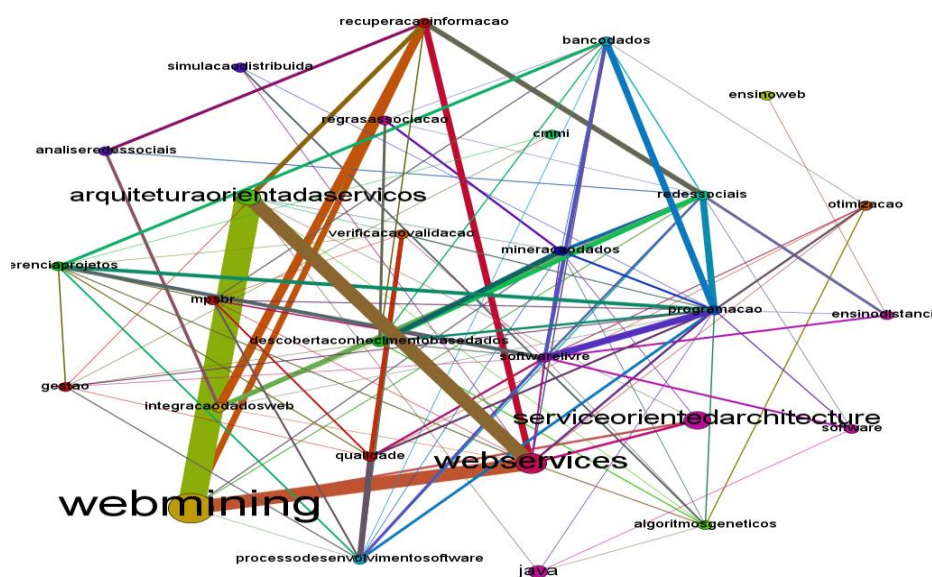
Outro exemplo de rede que pode ser gerada é da Figura 4. Onde, nós representam autores e arestas entre estes autores são caracterizadas por palavras-chave que os autores utilizaram em comum em uma de suas publicações.



**Figura 4 – Exemplo de autores (nós) vinculador por palavras-chave (arestas)**

Portanto, é possível identificar os pesquisadores que trabalharam com palavras-chave em comum, independentemente se já trabalharam em colaboração. A importância desta rede é que ela indica pelas arestas esparsas que um par de autores tem trabalhado com a mesma palavra-chave ou com o mesmo conjunto de palavras-chave, podendo indicar que existe semelhanças em seus trabalhos ou áreas de interesse.

Além das redes de colaboração onde nós representam autores de trabalhos, também é possível gerar redes onde os nós são palavras-chave e as arestas indicam palavras utilizadas na mesma publicação.



**Figura 5 – Exemplo de Redes de Palavras-chave**

Neste caso, a espessura das arestas indicam o número de publicações em que um par de palavras aparecem em conjunto e o tamanho dos nós representam a quantidade de vezes em que uma determinada palavra foi informada no conjunto de artigos analisados. Diante disso, é possível observar as palavras utilizadas em conjunto com maior frequência e as palavras que não são utilizadas com outras em um mesmo trabalho, ou seja, o quanto estão distantes umas das outras, podendo representar temas que estão sendo pesquisados conjuntamente.

A construção deste tipo de rede possibilita verificar o quão uma determinada palavra-chave é importante no conjunto analisado e quais são seus principais vínculos. Podendo desta forma, indicar áreas de pesquisa que estejam trabalhando em conjunto e também áreas que estão distantes umas das outras.

## 5. Conclusões

Este trabalho apresenta uma arquitetura para extração e identificação de colaborações científicas de todos os currículos cadastrados na Plataforma Lattes. A extração acontece com grande eficiência, já que só extrai os currículos atualizados e armazena todos os currículos Lattes em formato *XML* para posterior consultas e análises. Uma técnica eficiente para identificar colaboração entre os currículos é proposta e as redes de

colaboração são geradas com maior precisão. Pode se gerar redes específicas por termos, ou uma única rede de toda a plataforma.

O algoritmo para identificação de colaborações apresenta excelentes resultados com relação a sua precisão por utilizar o identificador que representa de forma única um determinado autor na plataforma. A grande vantagem da adoção deste método é com relação ao seu custo computacional. Como é realizada apenas uma comparação para cada título de artigo, é possível caracterizar a rede de colaboração a um custo linear  $\theta(n)$ , diferentemente dos métodos que trabalham com outras técnicas como, por exemplo, validação cruzada a um custo polinomial  $\theta(n^2)$ . Resultado disto é a viabilidade para se construir redes com um número muito grande de títulos a serem analisados, tornando o algoritmo proposto uma excelente alternativa para a identificação de colaboração em repositórios com grande quantidade de dados.

Além das colaborações é possível ainda identificar redes de palavras-chave de publicações científicas, desta forma, além de detectar temas que podem estar sendo investigados em conjunto, pode-se extrair várias informações relevantes que podem ajudar na compreensão de quais temas de pesquisa estão em evolução.

## Referências

Alvarenga, P. J. L.; Gonçalves, M. A.; Figueiredo, D. R. Ranqueamento Supervisionado de Autores em Redes de Colaboração Científica. SBBB - Simpósio Brasileiro de Banco de Dados, 2012, São Paulo, Brasil.

Alves, A. D.; Yanasse, H. H.; Soma, N. Y. LattesMiner: a multilingual DSL for information extraction from lattes platform. Proceedings of the compilation of the co-located workshops on DSM'11, TMC'11, AGERE!'11, AOOPEs'11, NEAT'11; VMIL'11. Portland, Oregon, USA: ACM: 85-92 p. 2011.

Alves, A. D.; Yanasse, H. H.; Soma, N. Y. Perfil dos bolsistas pq das áreas de engenharia de produção e de transportes do cnpq: enfoque na subárea de pesquisa operacional. XLIII Simpósio Brasileiro de Pesquisa Operacional, 2011a, Ubatuba, SP, Brasil.

\_\_\_\_\_. SUCUPIRA: Um Sistema de Extração de Informações da Plataforma Lattes para Identificação de Redes Sociais Acadêmicas. CISTI'2011 (6ª Conferência Ibérica de Sistemas e Tecnologias de Informação), 2011b, Chaves, Portugal.

Barabasi, A. L.; Oltvai, Z. N. Network biology: understanding the cell's functional organization. Nature Reviews Genetics, v. 5, n. 2, p. 101-113, 2004. ISSN 1471-0056.

Brandão, M. A.; Moro, M. M. Recomendação de Colaboração em Redes Sociais Acadêmicas Baseada na Afiliação dos Pesquisadores. SBBB - Simpósio Brasileiro de Bancos de Dados, 2012, São Paulo, Brasil.

CNPQ. Conselho Nacional de Desenvolvimento Científico e Tecnológico. Brasil, 2014. Disponível em: <<http://lattes.cnpq.br/>>. Acesso em: 19/09/2014.

- Dias, T. M. R. et al. Modeling and Characterization of Scientific Networks: A Study of the Lattes Platform. Brazilian Workshop on Social Network Analysis and Mining (BraSNAM), 2013, Maceió, Brasil.
- Dias, T. M. R.; Moita, G. F. Identifying relevant keywords in Scientific Collaboration Networks. 11th World Congress on Computational Mechanics (WCCM XI), 2014, Barcelona, Espanha.
- Digiampietri, L. A. et al. Minerando e caracterizando dados de currículos lattes. Brazilian Workshop on Social Network Analysis and Mining (BraSNAM), 2012, Curitiba, Brasil.
- Farias, L. R.; Vargas, A. P.; Borges, E. N. Um sistema para análise de redes de pesquisa baseado na Plataforma Lattes. Anais da VIII Escola Regional de Banco de Dados, 2012, Curitiba, Brasil.
- Fernandes, G. O.; Sampaio, J. O.; Souza, J. M. XMLattes - A Tool for Importing and Exporting Curricula Data. WORLDCOMP'11 - The 2011 World Congress in Computer Science, Computer Engineering, and Applied Computing, 2011, Las Vegas, Nevada, USA.
- Freire, V. P.; Figueiredo, D. R. Ranking in collaboration networks using a group based metric. Journal of the Brazilian Computer Society, v. 17, n. 4, p. 255-266, 2011. ISSN 0104-6500.
- Júnior, P. S. P.; Laender, A. H. F.; Moro, M. M. Analysis of Network Co-authoring the Brazilian Symposium on Databases. SBBD - Simpósio Brasileiro de Banco de Dados, 2011, Florianópolis, Brasil.
- Liben-Nowell, D.; Kleinberg, J. The link-prediction problem for social networks. Journal of the American Society for Information Science and Technology, v. 58, n. 7, p. 1019-1031, 2007. ISSN 1532-2890. Disponível em: <<http://dx.doi.org/10.1002/asi.20591>>.
- Lopes, G. R. et al. Ranking Strategy for Graduate Programs Evaluation. ICITA 2011 Journal of Information Technology and Applications, 2011, Sydney, Australia.
- Mena-Chalco, J. P.; Digiampietri, L. A.; Cesar-Jr, R. M. Caracterizando as redes de coautoria de currículos Lattes. Brazilian Workshop on Social Network Analysis and Mining (BraSNAM), 2012, Curitiba, Brasil. p.1-12.
- Mena-Chalco, J. P.; Junior, C.; Marcondes, R. ScriptLattes: an open-source knowledge extraction system from the Lattes platform. Journal of the Brazilian Computer Society, v. 15, n. 4, p. 31-39, 2009. ISSN 0104-6500.
- Newman, M. E. J. The structure of scientific collaboration networks. Proceedings of the National Academy of Sciences, v. 98, n. 2, p. 404-409, 2001. ISSN 0027-8424.
- \_\_\_\_\_. Coauthorship networks and patterns of scientific collaboration. Proceedings of the National Academy of Sciences, v. 101, n. suppl 1, p. 5200-5205, 2004. ISSN 0027-8424.
- Oliveira, J. P.; Lopes, G. R.; Moro, M. M. Academic Social Networks. International Conference on Conceptual Modeling (ER 2011), 2011, Bruxelas, Belgica.

- Petersen, A. M. et al. Persistence and uncertainty in the academic career. Proceedings of the National Academy of Sciences, v. 109, n. 14, p. 5213-5218, 2012. ISSN 0027-8424.
- Reijers, H. A. et al. A collaboration and productiveness analysis of the BPM community. In: (Ed.). Business Process Management: Springer, 2009. p.1-14. ISBN 3642038476.
- Sonnenwald, D. H. Scientific collaboration. Annual Review of Information Science and Technology, v. 41, n. 1, p. 643-681, 2007. ISSN 1550-8382. Disponível em: < <http://dx.doi.org/10.1002/aris.2007.1440410121> >.