Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação
E-ISSN: 1518-2924
bibli@ced.ufsc.br
Universidade Federal de Santa Catarina
Brasil

Rodrigues Dias, Thiago Magela; Farias Moita, Gray; Mascarenhas Dias, Patrícia
Adoção da plataforma lattes como fonte de dados para caracterização de redes científicas
Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação, vol. 21, núm. 47, septiembre-diciembre, 2016, pp. 16-26
Universidade Federal de Santa Catarina
Florianopolis, Brasil

**ARTIGO**

**Recebido em:**
**16/03/2016**

**Aceito em:**
**13/06/2016**

*Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação, v. 21, n. 47, p. 16-26, set./dez., 2016. ISSN 1518-2924. DOI: 10.5007/1518-2924.2016v21n47p16*

# Adoção da plataforma lattes como fonte de dados para caracterização de redes científicas

*Adoption of the lattes platform as the data source for the characterization of scientific networks*

Thiago Magela Rodrigues **DIAS**
Professor do Centro Federal de Educação Tecnológico de Minas Gerais – thiagomagela@gmail.com

Gray Farias **MOITA**
Professor do Centro Federal de Educação Tecnológico de Minas Gerais – gray@dpp.cefetmg.br

Patrícia Mascarenhas **DIAS**
Professor da Universidade do Estado de Minas Gerais – patriciamdias@gmail.com

**Resumo**
Os estudos sobre dados de produções científicas têm recebido atenção de pesquisadores, de diversas áreas, que visam obter conhecimento sobre a evolução das pesquisas em geral. Tais estudos possibilitam a análise da produção científica para diversos propósitos e um dos desafios neste tipo de análise está na diversidade de repositórios contendo dados em formatos e estruturas distintas. Os currículos da Plataforma Lattes se caracterizam atualmente como importante ferramenta para que pesquisadores, acadêmicos e estudantes, registrem seus dados, sendo amplamente utilizados, se caracterizando como um dos maiores repositórios de dados sobre produção científica, técnica, artística e profissional, contendo milhões de pesquisadores cadastrados. Neste trabalho é proposta uma plataforma para extração de todo o conjunto de dados dos currículos Lattes compondo um grande repositório de dados científicos, além disso, são implementadas técnicas para análises bibliométricas dos dados e identificação de redes de colaboração científica. Como resultados são apresentados estudos que objetivam obter uma visão geral sobre o repositório de currículos Lattes e como o conteúdo destes currículos pode ser utilizado para a caracterização de redes de colaboração científica. Conclui-se que os currículos Lattes são uma fonte extremamente rica de dados científicos e que sua adoção para estudos bibliométricos e baseados em análise de redes cientificas podem proporcionar resultados importantes para compreensão de como a ciência brasileira tem sido realizada. A grande dificuldade ao se analisar todo o repositório de dados da Plataforma Lattes está relacionada ao grande volume de dados que a compõem e ainda devido ao uso de técnicas, como por exemplo de identificação de colaborações pouco eficientes e computacionalmente complexas. Logo, este trabalho apresenta todo o potencial da Plataforma Lattes para análises bibliométricas de pesquisadores, sendo para isso proposta uma plataforma capaz de coletar e analisar todo o conjunto de dados com baixo custo computacional e com precisão satisfatória.
**Palavras-chave:** Plataforma Lattes. Extração de dados. Recuperação de informações. Colaboração científica.

v. 21, n. 47, 2016
p. 16-26
ISSN 1518-2924

**Abstract**
Studies scientific productions data has received attention from researchers in several areas, which aim to acquire knowledge about the evolution of research general them. Such studies allow for a review of scientific production paragraph several purposes and hum of the challenges this type of analysis is in repositories diversity containing data in different formats and structures. The resumes the Lattes Platform is currently characterized how important tool for what researchers, academics and students, data your register, being widely used, characterizing how largest data repositories about scientific production, technical, artistic and vocational, containing millions of researchers registered. In this work is proposal a platform for extraction entire data set of Curriculum Lattes composing great hum scientific data repository, in addition, are implemented technical for bibliometric analysis of data and identification of scientific collaboration networks. Results are presented as studies aimed that get an overview of the Curriculum Lattes and repository as the content of these resumes may be used one paragraph characterization of scientific collaboration networks. Done is the resumes Lattes that are an extremely rich source of scientific data and your adoption for bibliometric and based on scientific network analysis studies may provide results important for how understanding the Brazilian science has been held. The difficulty when analyzing whole Lattes Platform data repository is related large data repository and still due when using techniques such as collaborations identification shortly efficient and computationally complex. Therefore, this work presents the full potential of the Lattes Platform for bibliometric analyzes of researchers, for this proposal a platform to collect and analyze the entire data set with low computational cost and with satisfactory accuracy.
**Keywords**: Lattes Platform. Data Extract. Information retrieval. Science collaboration.

## 1 INTRODUÇÃO

Atualmente, serviços como bibliotecas digitais, redes de relacionamentos, repositórios bibliográficos e sítios para registro individual de produção científica, são alguns exemplos de como a internet tem contribuído consideravelmente na quantidade de trabalhos publicados permitindo que usuários não apenas acessem conteúdo disponível, mas também possam registrar a sua produção técnica e científica a partir de sua interação com a internet. Dessa forma, trabalhos publicados e disponibilizados pela internet são acessados de forma instantânea por interessados contribuindo de forma significativa para a expansão do conhecimento nas diversas áreas de pesquisa.

Nos últimos anos, além da produção científica, tem havido um constante crescimento no estudo das redes em relação às diversas disciplinas que vão desde a ciência da computação a áreas como a economia e sociologia. 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. A Internet, por exemplo, pode ser considerado um exemplo de uma rede importante e amplamente estudada em diversas áreas atualmente. 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, 2001a).

No domínio científico, um exemplo de uma rede social é a rede de colaboração científica que pode ser observada como um grafo no qual os vértices correspondem aos autores de publicações científicas e as arestas correspondem a relação de co-autoria. Neste tipo de rede, as arestas podem ou não ser ponderadas. A adição de peso pode representar, por exemplo, o número de artigos em conjunto que os autores publicaram. Vários trabalhos tem objetivado analisar redes de colaboração científica para compreensão de como grupos de pesquisa tem realizados seus trabalhado ou como determinada área realiza seus estudos (REVOREDO et al., 2012; BARABASI, OLTVAI, 2004; NEWMAN, 2004; NEWMAN, 2001b; NEWMAN, 2001c).

Com a modelagem e caracterização das redes é possível aplicar diversas técnicas que tem como objetivo, entender como estas redes estão estruturadas e consequentemente fornecer subsídios para diversos estudos como predição de vínculos entre pesquisadores, ranqueamento e classificação. Porém, a identificação destas colaborações em grandes bases de dados não é uma tarefa trivial devido a uma série de fatores inerentes a dificuldade de identificar vínculos, como a ambiguidade nos nomes de autores que são informados nos trabalhos publicados, a falta de citação de um determinado autor por seus pares, erros gramaticais nas citações, dentre outros.

Aliado a isso, a presença de dados de publicações disponíveis em diferentes formatos e em diferentes repositórios dificulta a realização de consultas por parte de usuários que necessitam de uma visão unificada desses dados ou da identificação de grupos de indivíduos que estejam trabalhando com determinado tema em diferentes instituições e regiões.

Diante disso, este trabalho tem como objetivo apresentar como os dados dispostos na Plataforma Lattes caracterizam-se como uma excelente alternativa para análise bibliométricas, possibilitando a adoção de diversas métricas para compreensão da comunidade científica brasileira.

Inicialmente é realizada a extração dos dados curriculares da Plataforma Lattes, e diante disso, é possível caracterizar redes de colaboração científica, bem como realizar análises bibliométricas que visam quantificar os dados extraídos. Essas análises são importantes pois podem revelar como as pesquisas científicas estão sendo conduzidas em todas as áreas do conhecimento.

Para o processo de extração de dados e caracterização das redes científicas deste projeto, foram utilizados dados da Plataforma Lattes do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Grande parte dos editais de financiamento de projetos feitos por instituições de amparo à pesquisa, como o próprio CNPq, utilizam os currículos Lattes dos pesquisadores como uma das formas de avaliação das propostas. Este fato motiva os pesquisadores a manter seus currículos com informações atualizadas, tornando a Plataforma Lattes uma fonte adequada para análise da produção científica brasileira. Figura1.


**Figura 1:** Visão Geral da Plataforma Lattes
**Fonte:** CNPq (2016)

Os currículos que compõem a Plataforma Lattes se tornaram um padrão nacional utilizado na avaliação individual das atividades científicas, acadêmicas e profissionais, agrega dados de pesquisadores de todas as áreas do conhecimento, tornando a Plataforma uma fonte extremamente rica para investigar e compreender o comportamento de diversos grupos de pesquisa (DIGIAMPIETRI et al., 2012).

Em Lane (2010) é destacado que medir e avaliar o desempenho acadêmico passa a ser um fator crucial para a vida científica. Vários fatores para fazer esta avaliação e medição estão vinculados ao cálculos de métricas, porém, os sistemas atuais de medição são insuficientes para determinar respostas confiáveis. No trabalho são descritos os vários problemas na adoção das métricas normalmente utilizadas para, por exemplo classificar grupos ou instituições de pesquisa e os desafios na propostas de métricas eficazes.

No trabalho de Mena-Chalco, Cesar-Junior e Marcondes (2009), são descritos motivos que tornam a Plataforma Lattes um interessante estudo de caso:

Os currículos cadastrados se tornaram um padrão nacional e que vem sendo utilizado na avaliação individual das atividades científicas, acadêmicas e profissionais;

Pesquisadores nacionais de diversas áreas do conhecimento estão cadastrados na plataforma;

Impulsionada pelas políticas de C&T nos últimos anos, a ciência brasileira vem apresentando grande crescimento de produção acadêmica que podem ser acompanhado pela análise dos currículos cadastrados.

A escolha da Plataforma Lattes para a extração está relacionada ao fato de que ela possui uma vasta quantidade de dados, pois trata da integração de dados científicos de currículos e de instituições da área de Ciência e Tecnologia (C&T), registrando os dados acadêmicos, técnicos e as produções científicas, permitindo ainda que a atualização dos dados individuais seja realizada pelos próprios pesquisadores. Atualmente a Plataforma Lattes conta com aproximadamente 4.4 milhões de currículos cadastrados, estes currículos possuem informações sobre formação acadêmica, áreas de pesquisa, atuação profissional e orientações acadêmicas. Diversos trabalhos para análise de dados científicos têm utilizados a Plataforma Lattes como principal fonte de informações (FARIAS et al., 2012; MENA-CHALCO, DIGIAMPIETRI, CESAR-JUNIOR, 2012; ALVES, YANASSE, SOMA, 2011a; ALVES, YANASSE, SOMA, 2011b; ALVES, YANASSE, SOMA, 2011c; FERNANDES, SAMPAIO, SOUZA, 2011).

No entanto, trabalhos encontrados na literatura frequentemente se concentram em grupos específicos para análises, principalmente quando se faz análises baseadas em redes de colaboração, já que as técnicas frequentemente utilizadas possuem alto custo computacional.

## 2 PROCEDIMENTOS METODOLÓGICOS

Apesar dos dados dos currículos da Plataforma Lattes serem disponibilizados livremente, estes são visualizados por interface de consulta disponibilizada pelo CNPq, que apresenta os currículos individualmente, sem possibilidade de análises e comparações com outros currículos. Além disso, todas as possíveis seções contendo informações como áreas de pesquisa, atuação profissional e orientações que compõem os currículos não são obrigatórias, cada currículo apresenta uma estrutura, bem como inconsistências frequentemente encontradas nos dados que compõem cada currículo. Logo, se faz necessário, a adoção de técnicas envolvidas na elaboração de extratores web para a extração dos dados.

Com os currículos extraídos, outras extrações podem ser realizadas com o intuito de enriquecer ainda mais a base de dados, como os pesquisadores vinculados a grupos de pesquisa, os próprios grupos de pesquisa e suas linhas de atuação, que

contém informações relevantes como palavras-chaves de interesse e indicadores de publicações destes grupos, que podem ser utilizados para realizar novos estudos. Todo o processo de extração dos dados foi realizado na plataforma de extração proposta por Dias et al. (2013).

A plataforma proposta pelos autores é responsável por extrair dados de diversos repositórios que contém informações científicas e acadêmicas. Os dados extraídos são integrados e armazenados em formato XML (eXtensible Markup Language) para que posteriormente possam ser analisados. O processo de extração de toda a base de dados curriculares da Plataforma Lattes é dividido em várias etapas que objetivam minimizar o custo computacional já que é preciso manter a base atualizada e nem todos os currículos são atualizados frequentemente, importante ainda considerar que os currículos atualizados podem ter novas informações inseridas, bem como a alteração e exclusão de informações já registradas. Além disto, considerando a necessidade de análises de grupos específicos como de uma instituição ou de um estado em particular, um mecanismo de seleção foi criado, responsável por gerar subgrupos de currículos baseados em critérios de interesse.

Com o auxílio da interface de consulta da Plataforma Lattes é realizada uma solicitação que retorna uma listagem contendo todos os currículos que estão cadastrados na plataforma. Após, são extraídos todos os link´s que possibilita o acesso a cada um dos currículos. Com isto, um extrator é responsável por acessar estes currículos e armazená-los em formato XML. Todos os currículos são armazenados em disco. Na etapa de extração, várias falhas estruturais que estão nos currículos são corrigidas, falhas estas que dificultam o processo de extração. 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.

Além dos currículos, também são extraídos dados de grupos e linhas de pesquisa da plataforma, para isso, de posse dos identificadores de cada um dos currículos Lattes, é possível verificar se um determinado pesquisador está vinculado ou não a algum grupo de pesquisa. Figura 2.


**Figura 2:** Página de um pesquisador vinculado a um grupo de pesquisa.
**Fonte:** CNPq (2016)

Caso ele esteja vinculado a um grupo de pesquisa, é possível extrair o identificar do referido grupo, e desta forma acessá-lo para posterior armazenamento, seguindo a mesma estratégia do extrator de currículos. Figura 3.


**Figura 3:** Página de um grupo de pesquisa.
**Fonte:** CNPq (2016)

Por fim, todas as linhas de pesquisa da Plataforma lattes também são extraídas e armazenadas em formato XML. As linhas de pesquisa são importantes pois elas possuem um conjunto de palavras-chave que podem ser utilizadas para construção de corpus de termos de pesquisa e também, dados sobre áreas do conhecimento a que tais linhas possam estar vinculadas. Figura 4.

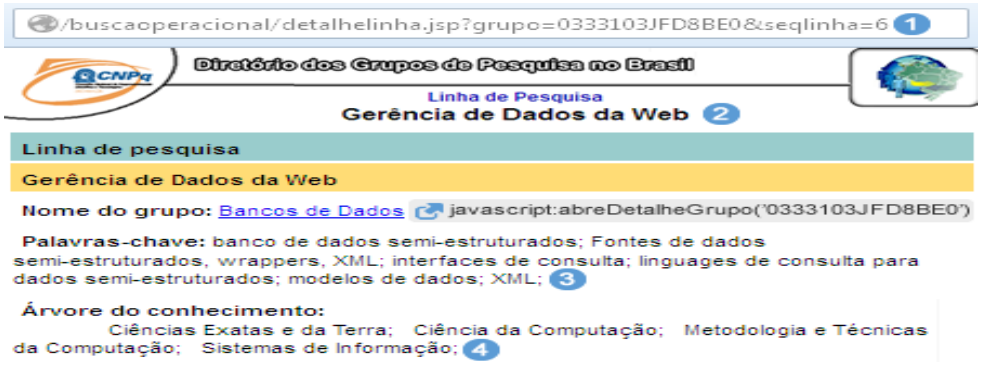


**Figura 4:** Página de uma linha de pesquisa.
**Fonte:** CNPq (2016)

Além destes repositórios que compõem a Plataforma Lattes, também são extraídos junto aos repositórios da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), como por exemplo, dados sobre a relação de cursos de mestrado profissional, mestrado (acadêmico) e doutorado reconhecidos/recomendados pela CAPES, com nota igual ou superior a 3, o quais de publicações, e ainda, possíveis resumos de teses e dissertações de pesquisadores que tenham currículos Lattes cadastrados. Estes dados são importantes pois possibilitam realizar a análise dos cursos de pós-graduação strictu sensu no Brasil, por áreas, instituição ou por regiões. Além disto, também possibilitam complementar as informações dos currículos Lattes, já que nem todos os currículos possuem informações sobre a nota de avaliação dos cursos informados.

## 3 MODELAGEM E CARACTERIZAÇÃO

Aqui é apresentada uma caracterização geral dos dados contidos na Plataforma Lattes e publicados até o dia 05/04/2015, correspondente a 4.156.635 currículos. Como mencionado anteriormente, a Plataforma Lattes conta com milhões de currículos cadastrados e impulsionada pelos órgãos governamentais e por agências de fomento tem tido crescente aumento no número de usuários. É possível identificar que a plataforma vem agregando a cada dia novos usuários que registram informações sobre suas atividades profissionais, acadêmicas e de pesquisa se apresentando como uma fonte extremamente rica e consistente de informações que podem revelar como tem sido feita a pesquisa científica brasileira nas diversas áreas do conhecimento.

Com todos os currículos armazenados localmente em formato XML, a possibilidade de manipulação dos dados com flexibilidade permite explorar todo o potencial que os dados curriculares da Plataforma Lattes oferecem. A Tabela 1 apresenta um resumo de alguns dados que compõem o repositório, os valores apresentados correspondem a um somatória das publicações de cada um dos trabalhos registrados, independentemente se existe algum trabalho em colaboração, que neste caso é contabilizado mais de uma vez.

**Tabela 1:** Resumo dos principais dados da Plataforma Lattes (04/2015)

| Tipo de Trabalho | Geral |
|---|---|
| Artigos em Anais de Congressos | 11.591.142 |
| Artigos em Periódicos | 4.560.921 |
| Capítulos de Livros | 1.055.388 |
| Demais Trabalhos | 667.944 |
| Livros | 455.447 |
| Textos em Jornais e Revistas | 1.353.645 |

**Fonte:** Os autores

Importante ressaltar que cada pesquisador pode possuir um único currículo cadastrado, não existindo a possibilidade de dados duplicados sobre a produção de um mesmo pesquisador. Além disso, em cada currículo é possível identificar a data de sua última atualização. Tendo em vista que a atualização dos currículos é realizada pelos próprios pesquisadores, frequentemente são encontrados currículos que estão há algum tempo sem atualizações. A Figura 5, apresenta uma visão sobre a data da última atualização de todos os currículos extraídos.

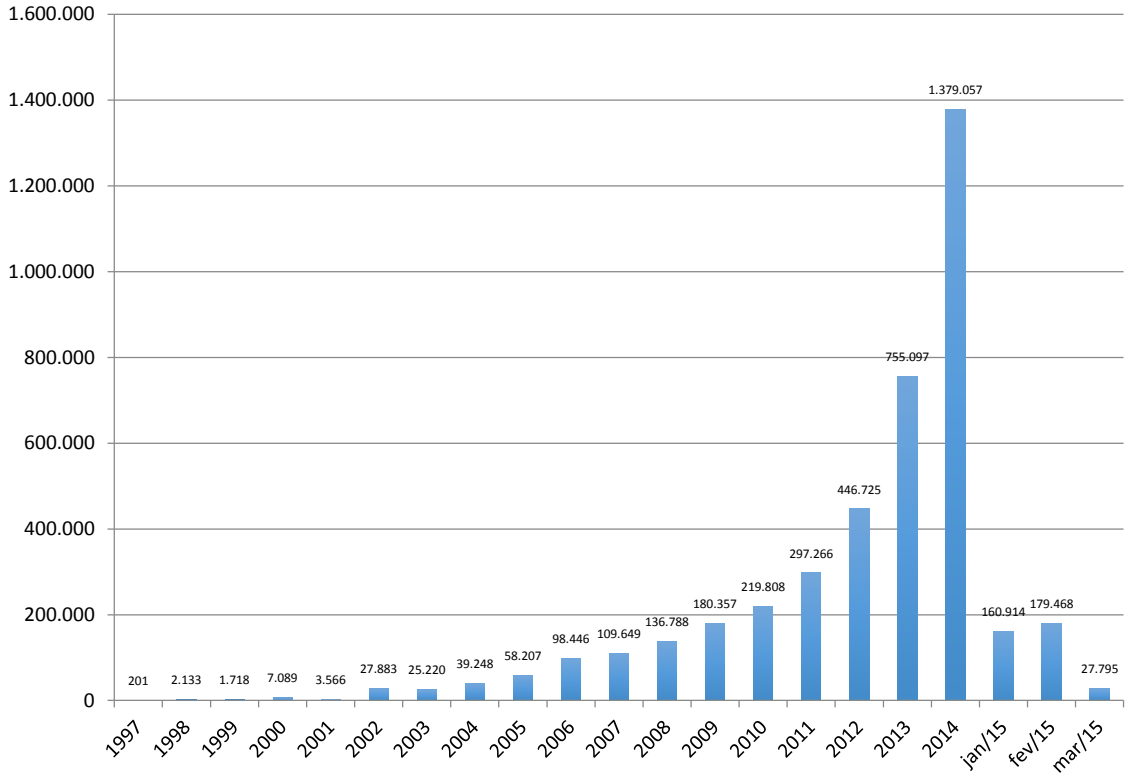


**Figura 5:** Data da última atualização dos currículos
**Fonte:** Os autores

É possível verificar que, apesar de existirem currículos com data da última atualização em 1997, a quantidade é baixa (201 currículos) se comparado com o número daqueles currículos atualizados nos anos posteriores, principalmente a partir de 2004 (39.248). A não atualização desses currículos pode ter motivos diversos e de difícil identificação, já que não é possível verificar as razões da falta de atualização de um determinado currículo. Porém, a grande maioria dos currículos possui data de atualização recente, cerca de 33% dos currículos (1.379.057) possuem data de última atualização em 2014 e 60,20% (2.502.331) foram atualizados nos últimos dois anos.

Com os currículos já extraídos e utilizando o conceito de grafos, onde os identificadores dos currículos são os vértices e as arestas são caracterizadas pela colaboração entre autores de trabalhos realizados em conjunto. O desafio está em identificar as colaborações em um grande volume de dados em tempo hábil, tendo em vista os diversos desafios envolvidos como duplicatas nos nomes de autores e divergência no título da mesma publicação quando cadastrada por co-autores.

O currículo a ser analisado pode possuir várias seções que irão permitir a identificação das colaborações com outros currículos, como por exemplo trabalhos ou orientações realizadas em conjunto. No entanto, por ausência de um processo automático de identificação de colaborações pela própria Plataforma Lattes, a relação de colaboradores de um pesquisador não é identificada automaticamente, como também acontece em outras bases de dados de publicações científicas. Logo, técnicas para a identificação das colaborações de forma automática se fazem necessárias para análise de grande volume de dados. Em Dias e Moita (2015), é proposto um método para a identificação de colaboração científica em grandes bases de dados e que foi adotada no presente trabalho para a construção das redes.

Com a aplicação de métodos para a identificação das colaborações, utilizando elementos como os citados anteriormente, é possível a modelagem das redes de colaboração e diante disso, várias métricas para análise de redes podem ser aplicadas

objetivando extrair conhecimento sobre como estes grupos estão estruturados, como colaboram, possibilitando alavancar a produção científica nacional com este conhecimento adquirido. Exemplo de uma rede de colaboração composta por 23 pesquisadores de um programa de pós graduação do CEFET-MG pode ser observada na Figura 6.

**Figura 6**: Exemplo de uma rede de colaboração científica de um grupo de pesquisadores
**Fonte:** Os autores

Na Figura 6 cada vértice representa um pesquisador e o tamanho do vértice representa a quantidade de publicações que ele possui. Já as arestas entre os vértices são caracterizadas por publicações que pesquisadores realizaram em colaboração e a espessura das arestas indicam a quantidade de trabalhos que dois pesquisadores realizaram em conjunto. As cores dos vértices indicam as áreas de atuação de cada pesquisador. Diante disto, diversas métricas para classificação, agrupamento, ranqueamento, dentre outras podem ser aplicadas.

## 4 CONSIDERAÇÕES FINAIS

A extração e análise de grandes repositórios de dados se mostra como uma importante ferramenta para análise bibliométricas e análises baseadas em colaborações científicas. As análises bibliométricas possibilitam verificar como a produção científica evolui ao longos dos anos, já as colaborações refletem o possível intercâmbio e troca de conhecimento que em sua maioria começa com a relação orientador e orientando ou entre grupos interessados em determinados temas. O grande desafio está no processo de extração e integração dos dados já que repositórios frequentemente utilizados para análises possuem estruturas e formatos distintos.

Neste contexto, a Plataforma Lattes em especial os seus currículos se apresentam como uma excelente fonte de dados científicos, que são atualizados frequentemente e ainda com um crescimento diário no número de novos currículos, agregando novos dados permanentemente, sendo estes passíveis de análise para entendimento da ciência brasileira.

Com a adoção de uma plataforma para a extração e análise dos dados foi possível verificar todo o potencial de análise da produção científica brasileira, tendo como principal fonte de dados os currículos Lattes. Foi possível verificar que grande parte destes currículos possuem atualização recente e que os milhões de artigos

publicados em anais de congressos e em periódicos, possibilitam a adoção de diversas métricas, como por exemplo, as baseadas em redes.

A exemplo do que foi realizado com os dados dos currículos Lattes, com as mesmas técnicas utilizadas na plataforma proposta, outros repositórios de dados científicos também podem ser recuperados, como por exemplo o número de citações de um determinado trabalho de repositórios como a Web of Science e Scopus.

## REFERÊNCIAS


ALVES, A. D.; YANASSE, H. H.; SOMA, N. Y. LattesMiner: a multilingual DSL for information extraction from lattes platform. In: **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, 2011a.

ALVES, A. D.; YANASSE, H. H.; SOMA, N. Y. Perfil dos bolsistas pq das áreas de engenharia de produçao e de transportes do cnpq: enfoque na subárea de pesquisa operacional. In: **XLIII Simpósio Brasileiro de Pesquisa Operacional**, Ubatuba, SP, Brasil, 2011b.

ALVES, A. D.; YANASSE, H. H.; SOMA, N. Y. SUCUPIRA: Um Sistema de Extração de Informações da Plataforma Lattes para Identificação de Redes Sociais Acadêmicas. In: **CISTI'2011 (6ª Conferência Ibérica de Sistemas e Tecnologias de Informação)**, Chaves, Portugal, 2011c.

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.

CNPq. Site institucional. Disponível em: http://www.lattes.cnpq.br. Acesso em: 01 fev. 2016.

DIAS, T. M. R.; MOITA, G. F. A method for thge identification of collaboration in large scientific databases. **Em Questão**, v. 21, n. 2, p. 140-161, 2015. Disponível em: http://seer.ufrgs.br/index.php/EmQuestao/article/download/53259/34340. Acesso em: 01 fev. 2016.

DIAS, T. M. R. et al. Modelagem e Caracterização de Redes Científicas: Um Estudo sobre a Plataforma Lattes. In: **Brazilian Workshop on Social Network Analysis and Mining**. 2013.

DIGIAMPIETRI, L. A. et al. Minerando e caracterizando dados de currıculos lattes. In: **Brazilian Workshop on Social Network Analysis and Mining (BraSNAM)**, Curitiba, Brasil, 2012

FARIAS, L. R. et al. Um sistema para análise de redes de pesquisa baseado na Plataforma Lattes. In: **VIII Escola Regional de Banco De Dados**, Curitiba - PR, 2012.

FERNANDES, G. O.; SAMPAIO, J. O.; SOUZA, J. M. XMLattes - A Tool for Importing and Exporting Curricula Data. In: **WORLDCOMP'11 - THE 2011 World Congress in Computer Science, Computer Engineering, and Applied Computing**, Las Vegas, Nevada, USA, 2011.

LANE, J. Let's make science metrics more scientific. **Nature**, v. 464, n. 7288, p. 488-489, 2010.

MENA-CHALCO, J. P.; DIGIAMPIETRI, L. A.; CESAR-JUNIOR, R. M. Caracterizando as redes de coautoria de currículos Lattes. In: **Brazilian Workshop on Social Network Analysis and Mining (BraSNAM)**, Curitiba, Brasil, 2012.

MENA-CHALCO, J. P; CESAR-JUNIOR, R. M.; 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.

NEWMAN, M. E. J. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. **Physical Review E**, v. 64, n. 1, p. 16132/1-16132/7, 2001.

NEWMAN, M. E. J. Scientific collaboration networks. I. Network construction and fundamental results. **Physical Review E**, v. 64, n. 1, p. 16131/1-16131/8, 2001.

NEWMAN, M. E. J. The structure of scientific collaboration networks. In: **Proceedings of the National Academy of Sciences**, v. 98, n. 2, p. 404-409, 2001.

NEWMAN, M. E. J. Coauthorship networks and patterns of scientific collaboration. In: **Proceedings of the National Academy of Sciences**, v. 101, n. suppl. 1, p. 5200-5205, 2004.

REVOREDO, K. et al. Mining scientific literature for analysis of collaboration in research communities. In: **Brazilian Workshop on Social Network Analysis and Mining (BraSNAM)**, Curitiba, Brasil, 2012.

Editores do artigo: Adilson Luiz Pinto, Rafaela Paula Schmitz e Enrique Muriel-Torrado