

Um Estudo da Colaboração Técnica Brasileira: Análise Baseada em Dados de Patentes

Patrícia M. Dias
UEMG
Divinópolis, Brasil
patriciamdias@gmail.com

João M. Melo
CEFET-MG
Divinópolis, Brasil
joao.galo.2014@gmail.com

Thiago M. R. Dias
CEFET-MG
Divinópolis, Brasil
thiagomagela@cefetmg.br

Abstract — Analyzes of scientific collaboration networks have been extensively explored in research from different areas of knowledge, in view of their ability to identify how groups of researchers have carried out their work collectively. Such studies make it possible to identify how collaboration between individuals occurs through analyzes based on social network metrics. In this context, new studies have been proposed in order to analyze collaboration in the development of technical products, with data on patents being studied in most studies. This type of analysis is relevant because it makes it possible to understand the collaboration process in the proposal of new inventions. In this work, initially a general characterization of the group of individuals analyzed is presented, and afterwards, a global and temporal analysis of the collaboration network is performed in the proposal of patents of Brazilian individuals with curricula registered in the Lattes Platform. For that, all the patents registered in the curricula of these individuals were used for the identification and characterization of the collaboration networks. As a result, it is possible to see how collaboration in the proposed inventions of the analyzed set has been intensified over the years, with an emphasis on the institutions and areas of expertise of each inventor.

Keywords - Patents, Lattes Platform, Data Analysis, Technical Production.

Resumo — As análises de redes de colaboração científica têm sido amplamente exploradas em pesquisas de diferentes áreas do conhecimento, tendo em vista sua capacidade de identificar como grupos de pesquisadores realizam seus trabalhos de forma coletiva. Tais estudos permitem identificar como ocorre a colaboração entre indivíduos por meio de análises baseadas em métricas de redes sociais. Nesse contexto, novos estudos têm sido propostos com o objetivo de analisar a colaboração no desenvolvimento de produtos técnicos, sendo os dados de patentes estudados na maioria dos estudos. Esse tipo de análise é relevante porque permite compreender o processo de colaboração na proposição de novas invenções. Neste trabalho, inicialmente é apresentada uma caracterização geral do grupo

de indivíduos analisados e, posteriormente, é realizada uma análise global e temporal da rede de colaboração na proposta de patentes de indivíduos brasileiros com currículos cadastrados na Plataforma Lattes. Para tanto, todas as patentes registradas nos currículos desses indivíduos foram utilizadas para a identificação e caracterização das redes de colaboração. Com isso, é possível perceber como a colaboração nas invenções propostas do conjunto analisado tem se intensificado ao longo dos anos, com destaque para as instituições e áreas de atuação de cada inventor.

Palavras-chave — Patentes, Plataforma Lattes, Análise de Dados, Produção Técnica.

I. INTRODUÇÃO

A implementação de novas tecnologias se faz um fator fundamental para a propulsão do desenvolvimento e crescimento de um Estado. Segundo [1], o processo de inovação requer vários tipos de tecnologia e conhecimento oriundos de diferentes fontes, incluindo indústria, empresas, laboratórios, institutos de pesquisa e desenvolvimento, academia e consumidores.

Percebe-se que grande parte das invenções e inovações geradas possuem como origem pesquisas iniciadas em instituições de ensino, o que faz delas um grande polo de inovação nacional. Este fator impulsionou a criação de institutos que visam proteger a propriedade intelectual gerada em universidades e também núcleos que visam auxiliar no processo de criação de patentes e na criação de pontes entre os setores acadêmicos e mercadológicos.

Atualmente há diversos repositórios online que possibilitam a pesquisa de produções científicas publicadas, como por exemplo DBLP (Digital Bibliography & Library

Project), ArnetMiner, Google Scholar e Microsoft Academic Search. Importante também destacar a Plataforma Lattes do CNPq, sendo essa um instrumento extremamente rico em dados para estudos sobre a produção científica e técnica brasileira. Os currículos da Plataforma Lattes apresentam a trajetória acadêmica dos pesquisadores e por isso são considerados um padrão nacional para análise de competências e méritos dos mesmos [2]. Por ser constantemente utilizada como parâmetro para avaliação em editais de financiamento de pesquisas, os pesquisadores se empenham em manter os seus currículos atualizados. Contudo, mesmo com a disponibilização livre dos dados e informações no âmbito de produções técnicas, tais como patentes, a base de dados da Plataforma Lattes ainda não foi amplamente analisada.

A partir desse contexto, este trabalho visa ampliar a compreensão sobre as atividades de patenteamento construídas nacionalmente, buscando avaliar os principais atores, as redes de colaboração e o resultado que estas infligem na evolução da ciência. Após a aprendizagem de descrições, identificação de padrões e geração de agrupamentos será possível apresentar resultados quantitativos e temporais sobre as patentes propostas por brasileiros e registradas nos currículos da Plataforma Lattes.

II. TRABALHOS RELACIONADOS

A Plataforma Lattes tem recebido destaque especial no setor acadêmico nacional como um dos principais repositórios para registros de informações de publicações científicas em diversas áreas de conhecimento. Mesmo com um alto repositório de dados, a plataforma possibilita somente a consulta simples de currículos, tornando inviável a construção de análises de um todo da plataforma, como por exemplo, áreas com maior volume de publicações e possíveis integrações entre pesquisadores que abordam tópicos com alta sinergia.

Além de informações referentes a produções científicas, a plataforma também disponibiliza dados de patentes registradas pelos colaboradores, sendo estas de própria autoria ou realizadas em colaboração com empresas ou instituições acadêmicas. Conforme afirmam [1], patentes são títulos de propriedade industrial sobre invenções, outorgados pelo Estado como recompensa ao inventor, garantindo a este segurança nas negociações para venda de tecnologias e processos criados.

Em [3] foi desenvolvido um estudo de caso com análise de redes sociais e entrevistas com executivos de um grupo empresarial para examinar uma rede de inventores e patentes depositadas entre 1978 e 2008 para esse grupo. Trata-se de um estudo de caso, com entrevistas semiestruturadas que foram necessárias para a primeira fase da pesquisa (qualitativa) e uma segunda fase (quantitativa) com um método até então

pouco utilizado em pesquisas de negócios internacionais, a utilização da análise de redes sociais (ARS). Os dados utilizados no trabalho foram obtidos de fontes secundárias e primárias. As principais fontes secundárias foram publicações internas do Grupo e de alguns de seus principais players. Para o estudo de caso, foram utilizadas como fonte primária entrevistas semiestruturadas, feitas com executivo de marketing e com gerente de engenharia de desenvolvimento de produtos. Como resultado desse trabalho, pôde ser avaliada a relação entre teoria de redes sociais e atores envolvidos com a geração de patentes em uma multinacional brasileira. Desse modo foi possível reconhecer a relevância estratégica da subsidiária e como ela aperfeiçoou os mecanismos de geração de patentes e disseminou para a matriz.

Já em [4] foram feitos mapeamentos de redes de inventores de uma empresa de cosméticos, com o objetivo de identificar padrões na organização da rede de inovação, além de identificar formas de promover a capacidade de inovação por meio da interconectividade. A execução da pesquisa foi realizada por meio de métodos de estudo de casos, e, para isso, foram feitos mapeamentos de inventores, com base nos registros de patentes pesquisados previamente, levando-se em consideração o vínculo (interno ou externo) com a empresa de cada inventor e também a quantidade de citações da patente. Para a coleta de dados, utilizou-se a base de dados de patentes da SpaceNet. Foi identificada maior hierarquia nas redes com a presença de colaboradores externos à empresa, bem como um possível maior teor tecnológico, já que a quantidade de citações foi superior às de outras redes.

III. METODOLOGIA

Neste trabalho, realizou-se a análise de dados contidos nos currículos cadastrados na Plataforma Lattes de diversos pesquisadores de todo o Brasil. A escolha da Plataforma Lattes para a extração dos dados 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 C&T, registrando os dados acadêmicos, técnicos e as produções científicas, permitindo ainda que a atualização dos dados individuais seja realizada pelos próprios pesquisadores.

Para a extração dos dados contidos na Plataforma Lattes, foi utilizado o framework *LattesDataExplorer* (Figura 1).

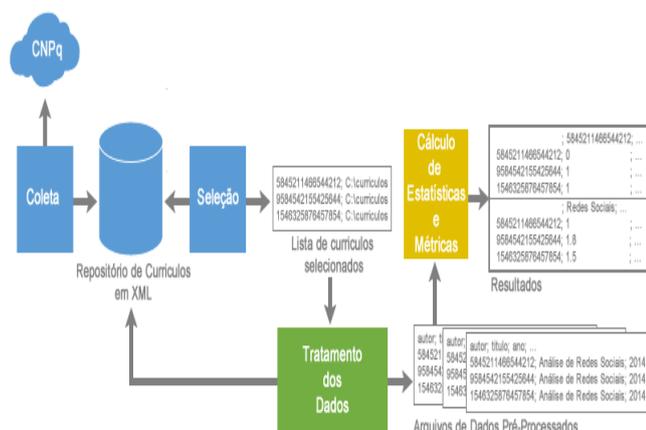


Fig. 1: Arquitetura geral do LattesDataXplorer [5].

O LattesDataXplorer é responsável por englobar um conjunto de técnicas para obtenção de todo o conjunto de currículos cadastrados na Plataforma Lattes. Com auxílio do método de coleta, cada um dos currículos é acessado e posteriormente recuperado para o armazenamento local. Além do método de coleta, ele também engloba mais uma série de métodos que possibilitam a realização de diversas análises bibliométricas tendo os currículos como fonte de dados.

De posse de todos os currículos armazenados localmente em formato .XML, o conjunto de dados passa por algumas fases que possibilitam alcançar os resultados esperados (Figura 2).

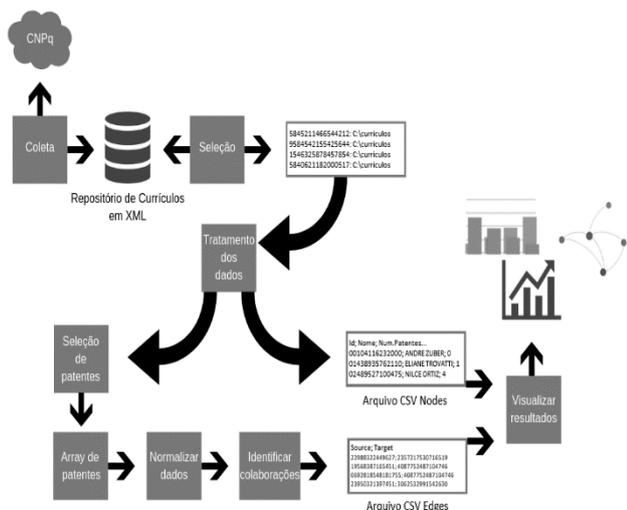


Fig. 2: Etapas do processo de caracterização das redes.

Após a obtenção dos dados, foi realizado um processamento deles a fim de se caracterizar as redes de colaboração em patentes entre pesquisadores. Essa etapa foi feita por um script desenvolvido em linguagem python, linguagem que permite uma escrita mais facilitada de scripts por apresentar diversos módulos com utilidade específica para cada caso e uma sintaxe fácil e maleável. As etapas executadas são basicamente:

1. Selecionar patentes que foram publicadas dentro do período de tempo - informado diretamente no código - que se está analisando. Ao longo desse processo, também são coletados dados de todos os pesquisadores que depositaram patentes.

2. Criar uma lista cujos índice seja o nome da patente, e o valor seja o conjunto dos identificadores de todos seus depositantes. Dessa forma, se torna possível identificar colaborações entre os pesquisadores.

3. Normalizar os nomes das patentes antes de usa-los como índices da lista, evitando assim a redundância de dados. Como exemplo, tem-se o caso de uma mesma patente estar registrada em dois currículos com nomes ligeiramente diferentes. Nesse caso, a normalização faria com que a correspondência entre os dois registros fosse identificada. Esse resultado é atingido através de diversos processos, tais como, radicalização, remoção de “stop-words” e transformação de todas as letras para o maiúsculo.

4. Identificar colaborações após a lista de patentes ter sido gerada. É criado um array contendo, em cada posição, dois identificadores, indicando uma colaboração entre A e B. Para montar esse array, são obtidas todas as permutações de dois elementos, sem repetição, do conjunto de identificadores de depositantes para cada patente.

As informações processadas foram armazenadas em arquivos .CSV. Para se obter uma visualização desses dados, utilizou-se o software Gephi, que auxiliou na construção de grafos, bem como a biblioteca matplotlib do python para gerar gráficos.

IV. RESULTADOS

Tendo em vista os dados analisados neste trabalho, foi possível quantificar a evolução da produção técnica brasileira, especificamente, as patentes registradas nos currículos cadastrados na Plataforma Lattes de forma temporal (Figura 3).

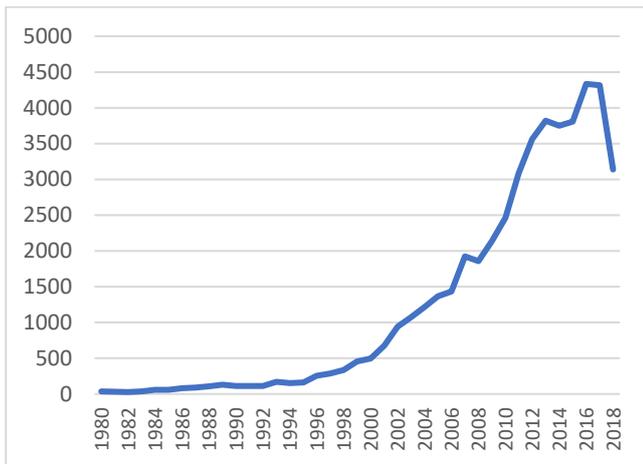


Fig. 3: Evolução temporal da proposta de patentes.

Como pode ser observado, a partir do ano 2000 percebe-se um crescimento constante na quantidade de patentes registradas, com uma pequena redução nos anos de 2007 e 2012. Após, destaca-se considerável queda a partir do ano de 2017, o que pode ser justificado por uma possível falta de atualização dos currículos, tendo em vista que a coleta dos dados foi realizada no ano de 2019. Tal fenômeno também já foi identificado na produção de artigos científicos, conforme apresentado em [5].

No intuito de avaliar a colaboração na proposta das patentes, foi necessário caracterizar as redes de colaboração técnica. Para tanto, a estratégia idealizada por [6] foi aplicada aos títulos das patentes no intuito de identificar as colaborações técnicas. Como resultado, foram caracterizadas as redes considerando todas as patentes recuperadas.

Na Figura 4 é possível observar as instituições com as maiores intensidades de colaborações (grau). Para isso, foram consideradas aquelas instituições que colaboram diretamente com outras na elaboração de patentes, independentemente do período de colaboração. Logo, para as análises aqui realizadas o vínculo institucional atual do indivíduo cadastrado em seu currículo da Plataforma Lattes foi considerado.

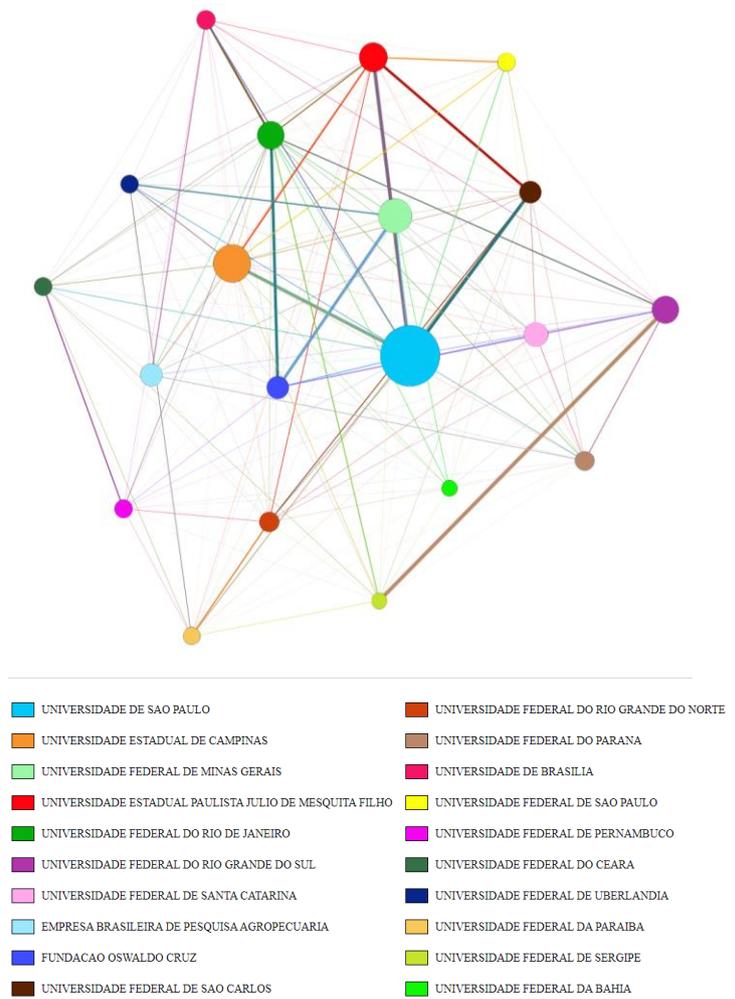


Fig. 4: Rede de colaboração caracterizada.

Como pode ser observado nesta rede, o diâmetro dos vértices caracterizados pelas instituições dos proponentes, leva em consideração a quantidade de patentes que seus pesquisadores possuem registrado em seus currículos. Logo, destaca-se a Universidade de São Paulo, Universidade Estadual de Campinas e Universidade Federal de Minas Gerais com o maior quantitativo de patentes, instituições estas que também figuram como as mais colaborativas.

Tais valores são diretamente influenciados por indivíduos altamente produtivos conforme já apresentado. Diante disso, novos estudos que avaliem em detalhes tais indivíduos podem proporcionar resultados interessantes sobre o processo de registro de patentes no Brasil.



V. CONSIDERAÇÕES

Foi possível observar que a quantidade de patentes vem tendo crescimento constante ao longo dos últimos anos, exceto nos anos de 2017 e 2018 que pode ter influência da falta de atualização de alguns currículos. Além disso, foi possível observar as instituições mais representativas, e ainda, como elas tem colaborado com outras instituições no registro de patentes. Observou-se que a colaboração acontece de forma mais intensa entre instituições que estão geograficamente próximas e que colaboram na proposta de trabalhos científicos ou no processo de formação acadêmica de seus pesquisadores.

REFERÊNCIAS

- [1] Amadei, J. R. P., & Torkomian, A. L. V. (2009). As patentes nas universidades: análise dos depósitos das universidades públicas paulistas (1995-2006). *Ciência da Informação*, 38(2), 9-18.
- [2] Amorin, C. V. (2003). Organização do currículo: Plataforma Lattes. *Pesquisa Odontológica Brasileira, SciELO Brasil*, v. 17, p. 18-22
- [3] João, B. D. N. (2009). Redes em subsidiárias de multinacionais: um estudo de caso com análise de redes sociais de inventores e patentes. *Revista de Administração Pública*, 43(5), 1037-1066.
- [4] Serrano, B. P.; & Junior, J. A. G. (2014). Redes de inovação: mapeamento de inventores de patentes em uma empresa do setor de cosméticos. *Revista GEPROS*, v. 9, n. 1, p. 101.
- [5] Dias, T. M. R. (2016). Um estudo da produção científica brasileira a partir de dados da Plataforma Lattes. 2016. Tese (Doutorado em Modelagem Matemática e Computacional) CEFET-MG, Minas Gerais.
- [6] Dias, T. M. R., Moita, G. F. (2015). A Method for Identification of Collaborations in Large Scientific Databases. *Em questão. Porto Alegre*, v. 21. N. 2, mai./jul.