



Um estudo sobre a rede de colaboração científica dos pesquisadores brasileiros com currículos cadastrados na Plataforma Lattes

Thiago Magela Rodrigues Dias

Doutor; Centro Federal de Educação Tecnológica de Minas Gerais, Belo Horizonte, MG, Brasil;
thiogomagela@gmail.com

Gray Farias Moita

Doutor; Centro Federal de Educação Tecnológica de Minas Gerais, Belo Horizonte, MG, Brasil;
gray@dppg.cefetmg.br

Patricia Mascarenhas Dias

Doutoranda; Centro Federal de Educação Tecnológica de Minas Gerais, Belo Horizonte, MG, Brasil;
patriciamdias@gmail.com

Resumo: As análises sobre as redes de colaboração científica vêm sendo amplamente exploradas em pesquisas de diversas áreas do conhecimento há algum tempo, tendo em vista a sua capacidade de identificar como grupos de pesquisadores têm realizado seus trabalhos coletivamente. Tais estudos possibilitam identificar como ocorre a colaboração entre os indivíduos através de análises baseadas em métricas de redes sociais. Neste trabalho, inicialmente é apresentada uma caracterização geral do conjunto analisado e, após, é realizada uma análise temporal da rede de colaboração científica dos doutores brasileiros com currículos cadastrados na Plataforma Lattes. Para isso, todas as publicações científicas dos doutores foram utilizadas para a identificação e caracterização das redes de colaboração. Como resultado, é possível visualizar como a colaboração científica do conjunto analisado foi sendo intensificada ao longo dos anos, com destaque para as grandes áreas de atuação de cada doutor.

Palavras-chave: Redes de colaboração. Análises de redes. Produção científica. Plataforma Lattes.

1 Introdução

“Nos últimos anos, além da análise bibliométrica de publicações científicas, diversos outros estudos têm procurado compreender como a ciência tem evoluído e como a colaboração científica ocorre.” (DIAS, MOREIRA; DIAS, 2018, p. 65). Para isso, diversas técnicas baseadas em análises de redes surgem

como uma alternativa para verificar esse fenômeno. De modo geral, uma rede pode ser caracterizada como um grafo, que consiste em um conjunto de nós (vértices) e ligações entre eles (arestas). (SZWARCFITER, 1986). Essas ligações podem ser direcionadas ou não e, opcionalmente, podem ainda ter um peso associado. Diversos fenômenos naturais podem ser descritos em termos de uma rede: em redes rodoviárias, as cidades e as rodovias que as interligam podem ser consideradas um grafo, no qual as cidades se caracterizam como nós, e as rodovias, como arestas. Já nas redes de comunicação, computadores interligados por um meio de comunicação constituem um grafo no qual os computadores representam os nós, e as ligações entre eles, as arestas (LIBENNOWELL; KLEINBERG, 2007).

Entre os vários tipos de rede, existem as redes sociais. Uma rede social é um conjunto de pessoas ou grupos que têm algum tipo de relação entre si (NEWMAN, 2001a). As redes sociais têm sido objeto de estudos há muitos anos. Barabási e Albert (1999) apresentam diversos estudos que motivaram e agregaram valor à teoria dos grafos, já que faziam uso do conceito de redes sociais. Em uma rede social de amigos, a relação entre duas pessoas pode representar uma amizade entre elas. Em uma rede de relações de parentesco, a relação 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 uma rede na qual os nós correspondem aos autores de publicações científicas, e as arestas correspondem à relação de coautoria. Nesse tipo de rede, as arestas podem ou não ser ponderadas. A adição de um peso representa o número de trabalhos em que os autores relacionados pela aresta considerada participaram conjuntamente. 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 para representar, por exemplo, a afinidade e os interesses comuns entre dois autores da rede (SONNENWALD, 2007¹). (DIAS; MOREIRA; DIAS, 2018, p. 65).

Newman (2001a) apresenta uma avaliação extensa sobre características sociais das redes de colaboração científica nas áreas de Computação, Biologia, Física e Medicina no período de 1995 a 1999, encontrando diversas

características sociais que são comuns a essas redes. Já em outro trabalho (NEWMAN, 2003), o autor analisa redes de coautoria nas áreas de Biologia, Física e Matemática, visando a identificar propriedades estatísticas, como número médio de publicações e quantidade média de autores, de modo a determinar padrões comuns entre os elementos da rede analisada.

Com a modelagem e caracterização das redes de colaboração científica, é possível aplicar diversas técnicas que permitem entender como essas redes são estruturadas, fornecendo assim subsídios para diversos estudos, como predição de vínculos entre pesquisadores, recomendação de especialistas e identificação de grupos de pesquisa. (DIAS; MOREIRA; DIAS, 2018, p. 65).

Porém, a identificação dessas colaborações em grandes repositórios de dados não é uma tarefa trivial devido a uma série de fatores, como ambiguidade entre nomes de autores, relação de autores incompleta, erros ortográficos nas citações ou nos títulos das publicações, entre outros.

Diante disso, este trabalho apresenta um estudo sobre as redes de colaboração científica dos doutores brasileiros, baseado em análises de redes que possibilitam compreender como ocorre a colaboração científica do conjunto analisado e como os pesquisadores estão colaborando nas diversas áreas do conhecimento. Para isso, são analisados dados dos currículos cadastrados na Plataforma Lattes do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), responsável por armazenar informações curriculares da comunidade científica nacional. A Plataforma Lattes é considerada uma importante fonte de dados sobre a produção científica brasileira (LANE, 2010).

2 Trabalhos relacionados

Nesta seção, são apresentados os trabalhos relacionados que utilizam dados de publicações científicas para analisar e compreender como ocorre a colaboração científica dos principais pesquisadores em diversas áreas de atuação.

Meadows e Lemos (1999) destacam um aumento no número de trabalhos científicos compartilhados, impulsionados principalmente pela facilidade na

forma de distribuir os trabalhos, possibilitando maior velocidade no intercâmbio de informações. Aliado a isso, os autores ressaltam que os trabalhos em colaboração permitem a economia de tempo e de recursos financeiros e materiais, resultando em um processo cada vez mais estimulado por instituições e agências que financiam pesquisas. Para Sonnenwald (2007), a interação entre dois ou mais cientistas que, em determinado contexto, objetiva compartilhar atividades para obter resultados e atingir objetivos comuns é definida como colaboração científica e caracteriza-se por ser um fenômeno complexo.

Diante do crescente aumento dos trabalhos realizados em colaboração e da facilidade de acesso a esses trabalhos, a análise das redes de coautoria apresenta-se como uma importante ferramenta para compreender, por exemplo, como tem acontecido a colaboração científica, identificando a existência de grupos com colaborações mais intensas e autores com maior grau de colaboração. Redes de colaboração científica vêm sendo amplamente estudadas, como mostram os trabalhos de Newman (2001a; 2001b; 2001c; 2004) e Kumar (2015). A estrutura das redes de colaboração possibilita revelar características interessantes sobre as comunidades científicas (NEWMAN, 2004).

Segundo Ströele, Zimbrão e Souza (2012²), a análise de redes de colaboração científica possibilita identificar como os grupos de pesquisadores e centros de estudos estão desenvolvendo seus trabalhos, verificar e entender qual o grau de envolvimento entre os pesquisadores de determinados grupos, áreas do conhecimento ou instituições de ensino e pesquisa, e também permitir a indicação de padrões de colaboração que poderiam proporcionar um grande avanço na área, visando melhorias na comunicação e colaboração de toda a comunidade científica. (DIAS; MOREIRA; DIAS, 2018, p. 66).

Petersen et al. (2012) analisam fatores relevantes para o sucesso científico em redes de colaboração. Entre eles, são destacados a quantidade de produção científica, que aumenta a atração de novas oportunidades, e também o tamanho da equipe de colaboradores e da própria rede. Diante disso, justifica-se o grande potencial de novos estudos que visam à análise das redes de colaboração, objetivando entender como a colaboração científica acontece e como ela vem evoluindo em cada uma das áreas de pesquisa.

Barabási et al. (2002³) apresentam um estudo das redes de colaboração científica nas áreas de Neurociência e Matemática entre os anos de 1991 e 1998. Os autores concluíram que a rede evolui a partir do momento em que novos nós e vínculos entre os nós existentes são incluídos, e que es[s]a evolução segue um processo denominado preferential attachment, em que novos nós tendem a criar seu primeiro vínculo com nós que possuem grande número de vínculos. Como resultado, autores experientes tendem a aumentar seu número de colaboradores com mais frequência do que autores novatos. (DIAS; MOREIRA; DIAS, 2018, p. 66).

Todos esses trabalhos foram importantes para demonstrar o potencial que a análise das redes de colaboração científica pode proporcionar para a melhor compreensão sobre como um determinado conjunto de pesquisadores colabora ao longo do tempo e ainda sobre como ocorre a colaboração nas diversas áreas do conhecimento.

Com o intuito de compreender o funcionamento da colaboração científica brasileira, diversos trabalhos exploraram dados de publicações registrados nos currículos cadastrados na Plataforma Lattes para a caracterização e análise das redes de colaboração científica.

Perez-Cervantes et al. (2013) introduzem novas medidas para estimar a influência da colaboração em redes científicas. A abordagem é baseada na técnica de predição de links e avalia como a presença ou ausência de um pesquisador afeta o processo de predição na rede em análise. Para isso, os cientistas são representados por nós em uma rede de colaboração e, após a remoção de nós, o processo de predição de links é realizado de forma iterativa para todos os outros nós. O modelo de aprendizagem supervisionado Support Vector Machines (SVM) foi utilizado como preditor e avaliou redes reais de colaboração em vários períodos. Nesse trabalho, os dados utilizados foram os currículos da Plataforma Lattes. Para os testes, foram selecionados 889 pesquisadores afiliados a 45 departamentos de Ciência da Computação de instituições de ensino brasileiras, e as publicações consideradas para a identificação das colaborações ocorreram entre 2000 e 2012.

Digiampietri et al. (2014) realizaram um estudo da rede acadêmica dos doutores que atuam no Brasil a fim de identificar características da rede. No trabalho, os autores caracterizaram as redes sociais acadêmicas a partir dos relacionamentos presentes nos currículos cadastrados na Plataforma Lattes dos

doutores brasileiros que atuam no Brasil. Após a caracterização das redes, foram aplicadas diversas métricas de análise de redes sociais. Além disso, também foram aplicadas técnicas de mineração de textos para identificar quais são as expressões mais utilizadas pelos doutores. Como resultado, destaca-se que as redes são heterogêneas, pouco densas e com caminhos mínimos_médios relativamente baixos, e ainda que, ao analisar os títulos das publicações, expressões regionais destacam-se devido a sua frequência.

Mena-Chalco et al. (2014⁴) utilizam dados dos currículos da Plataforma Lattes para identificar e caracterizar a rede de colaboração de pesquisadores brasileiros. O trabalho objetivou extrair os dados de currículos cadastrados na Plataforma Lattes, identificar automaticamente a colaboração baseada em informações bibliométricas – produzindo uma rede de colaboração –, e aplicar métricas baseadas em [uma] análise topológica para compreender como ocorre a interação entre os pesquisadores. (DIAS; MOREIRA; DIAS, 2018, p. 66).

Boaventura et al. (2014) caracterizam a evolução das redes de colaboração científica, representadas pelas redes de coautoria das seguintes universidades brasileiras: Universidade Federal do Amazonas (UFAM), Universidade Federal de Minas Gerais (UFMG), Universidade Federal de Pernambuco (UFPE), Universidade Federal do Rio Grande do Sul (UFRGS), Universidade Federal do Rio de Janeiro (UFRJ), Universidade de Brasília (UnB), Universidade Estadual de Campinas (Unicamp) e Universidade de São Paulo (USP). A análise temporal considera os anos de 2000 a 2013 e o seu foco são pesquisadores (professores) dessas universidades. Métricas tradicionais de redes sociais, tais como densidade e diâmetro da rede, grau de centralidade, grau de proximidade e grau de intermediação, foram calculadas. Os autores apresentam também a análise da endogamia dessas instituições e a sua correlação com a avaliação trienal dos programas de pós-graduação realizada pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). Por fim, analisam a existência de grupos de pesquisadores que frequentemente publicam em conjunto.

As principais conclusões da análise apresentada por Boaventura et al. (2014) são as seguintes: (1) a rede de colaboração das universidades analisadas

possui a mesma característica de densificação das colaborações entre os seus pesquisadores – diâmetro relativamente baixo (comparado à cardinalidade da rede) indica que as redes possuem características de redes de mundo pequeno; (2) universidades que possuem a maior porcentagem de programas de pós-graduação com conceitos acima de cinco possuem valores menores de endogamia.

O conteúdo deste trabalho está inserido no contexto de caracterização e análise de redes de colaboração científica. Entretanto, ao contrário dos trabalhos anteriormente citados, o presente trabalho não se restringe a redes de colaboração científica específicas de uma determinada região, pequenos grupos de pesquisadores ou conjuntos de publicações. A novidade está na análise de todo o conjunto de currículos dos doutores brasileiros cadastrados na Plataforma Lattes.

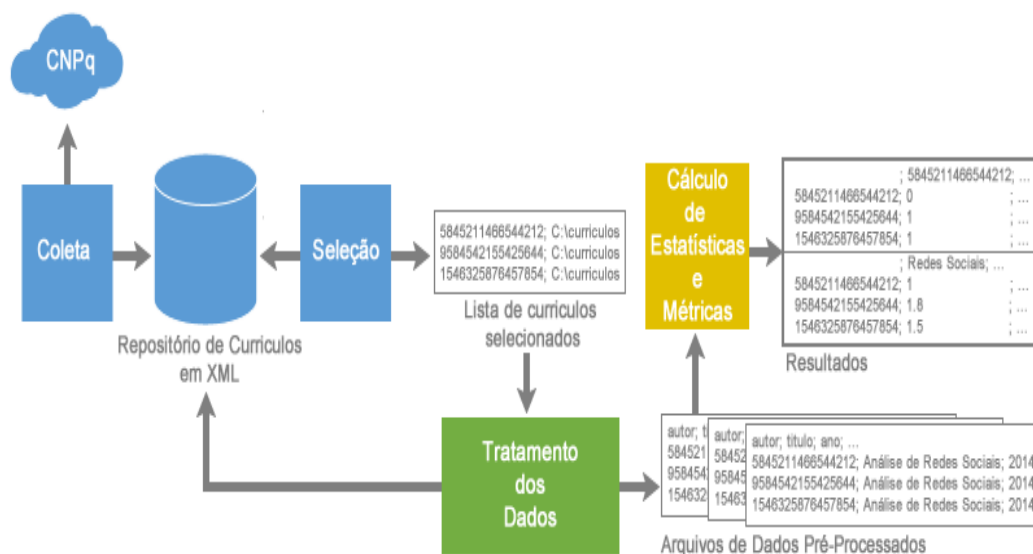
3 Metodologia

Uma grande parte dos editais de financiamento de projetos de pesquisa realizados por agências de amparo à pesquisa, instituições de ensino, bem como pelo próprio CNPq utilizam dados cadastrados nos currículos Lattes dos proponentes como uma das formas de avaliação das propostas. Isso passou a ser um grande incentivo para que os pesquisadores mantivessem seus currículos com informações atualizadas, tornando a Plataforma Lattes uma fonte extremamente rica para análise da produção científica brasileira. (DIAS; MOREIRA; DIAS, 2018, p. 67).

Durante o período compreendido entre maio de 2014 e abril de 2016, houve um aumento de aproximadamente 26% na quantidade de currículos armazenados na plataforma, mantendo um significativo crescimento nesse intervalo.

O *LattesDataXplorer* (DIAS, 2016), um *framework* para extração e tratamento dos dados, foi utilizado (Figura 1), considerando que os trabalhos correlatos analisaram, em sua maioria, apenas grupos específicos de indivíduos e tendo em vista que a manipulação de grandes quantidades de currículos da Plataforma Lattes não é uma tarefa trivial, já que existem problemas que envolvem recuperação de informação e algoritmos eficientes para manipulação de grande volume de dados.

Figura 1 – LattesDataXplorer, arcabouço utilizado para extração e tratamento de dados curriculares da Plataforma Lattes



Fonte: Dias (2016).

O processo de coleta de todos os dados curriculares da Plataforma Lattes é dividido em três componentes que objetivam minimizar o custo computacional: 1) extração de URLs, que é responsável por extrair as referências únicas para todos os currículos cadastrados e desta forma possibilitar o acesso individual a cada currículo, 2) extração de Identificadores (Ids) e Data, que visa acessar cada currículo e extrair o seu identificador individual, bem como a data de última atualização, 3) extração de currículos, que é responsável por extrair e armazenar os currículos (em formato XML) cuja data de atualização na Plataforma Lattes seja divergente da data de atualização do currículo armazenado localmente.

Todas essas etapas [fazem-se] necessárias, já que o ideal é manter os dados curriculares atualizados com a maior frequência possível, possibilitando a realização de análises [a partir de] dados atualizados, tendo em vista que, com a estratégia adotada, não é necessário extrair todo o repositório de dados a cada nova extração. É importante ainda considerar que os currículos atualizados podem ter novas informações inseridas, bem como a alteração e exclusão de informações já registradas, o que torna o processo de atualização de campos específicos uma tarefa complexa. Na abordagem deste trabalho, todo currículo atualizado é substituído pelo mais recente. Além disso, considerando a necessidade de análises de grupos específicos, como, por exemplo, de docentes de uma instituição ou de uma área de pesquisa em particular, um componente denominado Seleção caracteriza-se como [um] importante mecanismo no contexto deste trabalho. (DIAS; MOREIRA; DIAS, 2018, p. 68).

O extrator é responsável por verificar, em um arquivo de identificação, se existem currículos cuja data de atualização no CNPq é divergente da data de

atualização local; em caso afirmativo, o currículo é extraído e substituído o que estava na base local, alterando também a data de atualização local no arquivo de identificação. Caso contrário, significa que o currículo local já é a última versão atualizada na Plataforma Lattes, tendo em vista que as duas datas são iguais. Nesse caso, não é realizada a extração do currículo, e o próximo identificador que representa um currículo é analisado. Finalmente, são extraídos os novos currículos cadastrados, pois foram inseridos, ao final do arquivo de identificação, seus códigos, seus identificadores e datas de atualização no CNPq, já que não possuem data de atualização local, uma vez que não foram extraídos. Diante disso, esses currículos são extraídos pela primeira vez e sua data de atualização local é registrada no arquivo de identificação.

Todo esse processo possibilita manter uma base atualizada com baixo custo computacional, já que um percentual considerável de currículos não é atualizado frequentemente. Com exceção da primeira extração, o processo apresentado neste trabalho possibilita extrair somente os currículos novos ou recentemente atualizados, resultando em uma base atualizada. Com todos os currículos armazenados localmente em formato XML (*eXtensible Markup Language*), a possibilidade de manipulação dos dados com flexibilidade permite explorar todo o potencial que os dados curriculares da Plataforma Lattes oferecem. Com o intuito de analisar grupos específicos de currículos, como os indivíduos que possuem uma determinada titulação concluída, é utilizado neste trabalho o componente de Seleção para a composição do conjunto de indivíduos que possuem, registrada em seus currículos, capacitação a nível de doutorado concluída.

Para isso:

O componente de Seleção utiliza a linguagem de consulta XPath (XML Path Language) para pesquisa e posterior geração do conjunto. A linguagem XPath permite construir expressões que vão processar e percorrer um documento XML de forma similar ao uso de expressões regulares. Logo, é possível agrupar um conjunto de currículos com base em parâmetros desejados. (DIAS; MOREIRA; DIAS, 2018, p. 68).

Diante disso, foi possível identificar um total de 223.477 indivíduos que possuem o doutorado concluído e registrado em seus currículos. Logo, os dados contidos nos currículos desse conjunto de indivíduos foram analisados, sendo apresentada uma visão geral deles; posteriormente, são caracterizadas as redes de colaboração científica para análises cujos resultados são apresentados na próxima seção. Apesar dos dados terem sido coletados em abril de 2015, período em que se iniciou este estudo, as análises aqui apresentadas irão englobar publicações até o ano de 2014.

4 Resultados

Para as análises da colaboração científica dos principais pesquisadores brasileiros, considerou-se o conjunto de doutores com currículos cadastrados na Plataforma Lattes. O motivo da seleção desse conjunto de indivíduos para análises específicas tem como motivação o fato de ele, apesar de corresponder a somente 5,38% do total de indivíduos, ser responsável por 74,51% dos artigos publicados em periódico e 64,67% dos artigos publicados em anais de congresso (DIAS, 2016). Esse grupo de indivíduos que, em sua maioria, têm atuado em pesquisas, seja em instituições de ensino ou em institutos de ciência e tecnologia, ainda é responsável pela formação dos alunos nos principais programas de pós-graduação no Brasil, e vários deles são reconhecidos por sua elevada produção científica. Com isso, ressalta-se que o conjunto de indivíduos analisado neste trabalho compreende grande parte dos pesquisadores em atuação no Brasil.

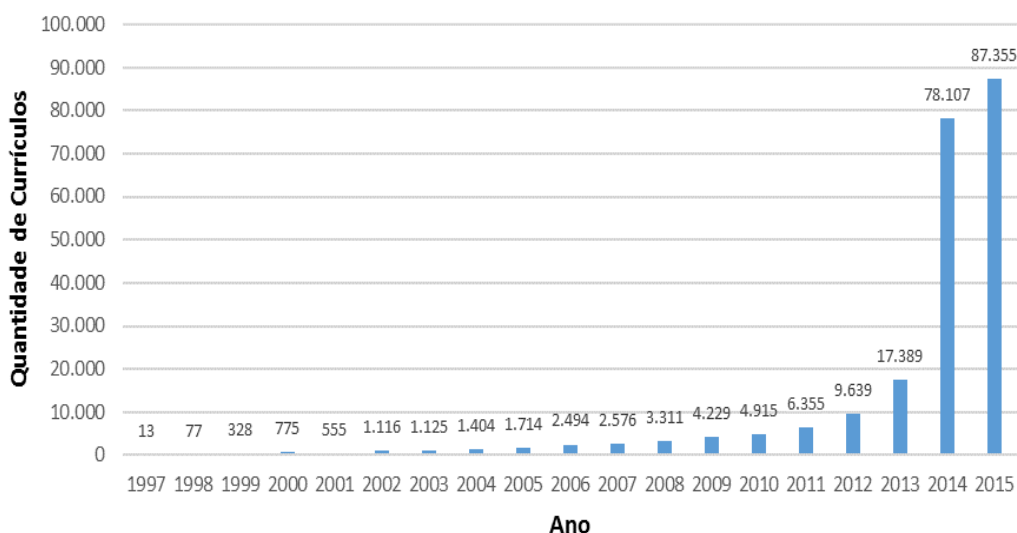
A análise dos doutores que possuem currículos cadastrados na Plataforma Lattes tem como objetivo avaliar o conjunto de pesquisadores que possuem o nível mais alto de qualificação acadêmica. Em geral, ter doutorado concluído é necessário para atuar em programas de pós-graduação, como também para participar de diversos editais de fomento à pesquisa no Brasil. Tendo em vista traçar a tendência de publicação desse conjunto de indivíduos, verificando os principais meios de publicação do conjunto, percebe-se que a

maioria possui grande quantidade de artigos publicados em anais de congresso, seguido por artigos em periódico.

4.1 Caracterização geral

Foi observado que, do total de currículos cadastrados na Plataforma Lattes (4.156.635 em abril de 2015), 223.477 correspondem a indivíduos que possuem como titulação máxima concluída o doutorado. Como vários desses doutores estão vinculados a programas de pós-graduação, e como grande parte dos doutores participa de grupos de pesquisa, grupos estes registrados no Diretório de Grupos de Pesquisa no Brasil do CNPq, devido ao Censo 2014, realizado no final do ano de 2014, todos os pesquisadores que estavam integrando algum desses grupos foram recomendados a atualizar seus currículos. Com isso, é possível observar que esses currículos que fazem parte das análises apresentadas neste estudo possuem, em geral, dados recentemente atualizados (Figura 2).

Figura 2 - Data da última atualização registrada nos currículos cadastrados na Plataforma Lattes dos doutores



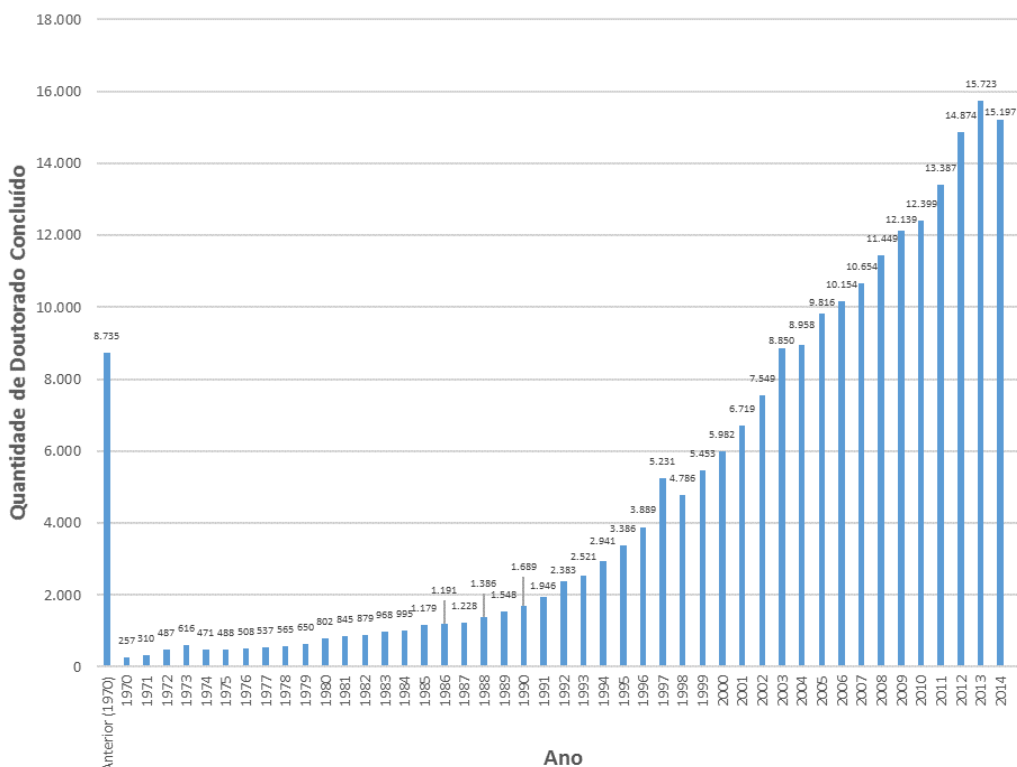
Fonte: Elaborado pelos autores.

Aproximadamente 70% dos currículos dos indivíduos com doutorado concluído foram atualizados há menos de um ano; desses, 39% já nos primeiros meses de 2015. Os currículos não atualizados recentemente, em geral, são de

indivíduos que informaram apenas informações básicas e possuem poucos registros sobre publicações ou dados sobre atuações profissionais. A grande vantagem de analisar um conjunto de currículos recentemente atualizado, é que trabalhos recém-publicados provavelmente já estarão registrados nesses currículos, proporcionando uma visão atual da produção científica brasileira.

Outra informação importante para compreensão do perfil científico nacional está relacionada à quantidade de doutores que se formam a cada ano. A partir da Figura 3, é possível verificar temporalmente que a quantidade de doutores tem aumentado consideravelmente.

Figura 3 - Quantidade de doutores formados por ano, tendo como base a data de conclusão dos cursos registrados nos currículos da Plataforma Lattes



Fonte: Elaborado pelos autores.

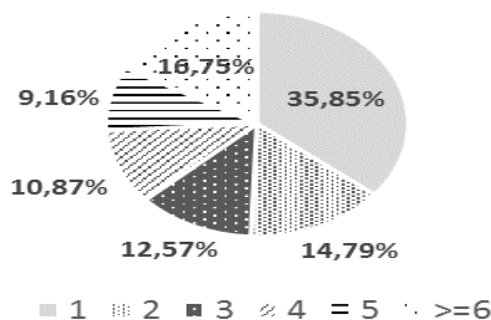
O doutorado concluído mais antigo identificado no conjunto de indivíduos analisado é de 1932, seguido por outro concluído em 1939 e, posteriormente, um terceiro em 1944. Em seguida, o número de conclusões aumenta de forma muito modesta até meados de 1970, com algumas dezenas de

doutores formados por ano. De 1975 até 1996, ocorreu um crescimento constante no número de doutores formados e, em 1997, houve um aumento significativo nesse número – que não se manteve no ano seguinte, mas que voltou a se manifestar de forma constante a partir de 1998. Entre 2011 e 2013, a taxa de crescimento aumenta novamente de forma significativa, sofrendo, no entanto, uma queda em 2014. Esse período coincide exatamente com a expansão dos cursos de doutorado ocorrido nos últimos anos no país, conforme destacado nas avaliações trienais da CAPES de 2007 e 2010.

4.2 Análise das redes de colaboração científica

Um tipo de análise que comumente vem sendo realizada é a análise sobre como a comunidade científica brasileira tem colaborado (MENA-CHALCO; DIGIAMPIETRI; CESAR-JUNIOR, 2012; BOAVENTURA et al., 2014; MENA-CHALCO et al. 2014; DIGIAMPIETRI, 2015). Neste trabalho, são analisadas as redes de colaboração científica dos doutores brasileiros, considerando suas grandes áreas de atuação no período de 1960 a 2014, que compreende o intervalo entre ano de publicação da produção mais antiga cadastrada na Plataforma Lattes e o último ano do período de análise considerado. É possível observar que uma grande quantidade (35,85%) dos artigos publicados em anais de congresso e periódicos é de autoria única e, conseqüentemente, não caracteriza colaborações (ver Figura 4).

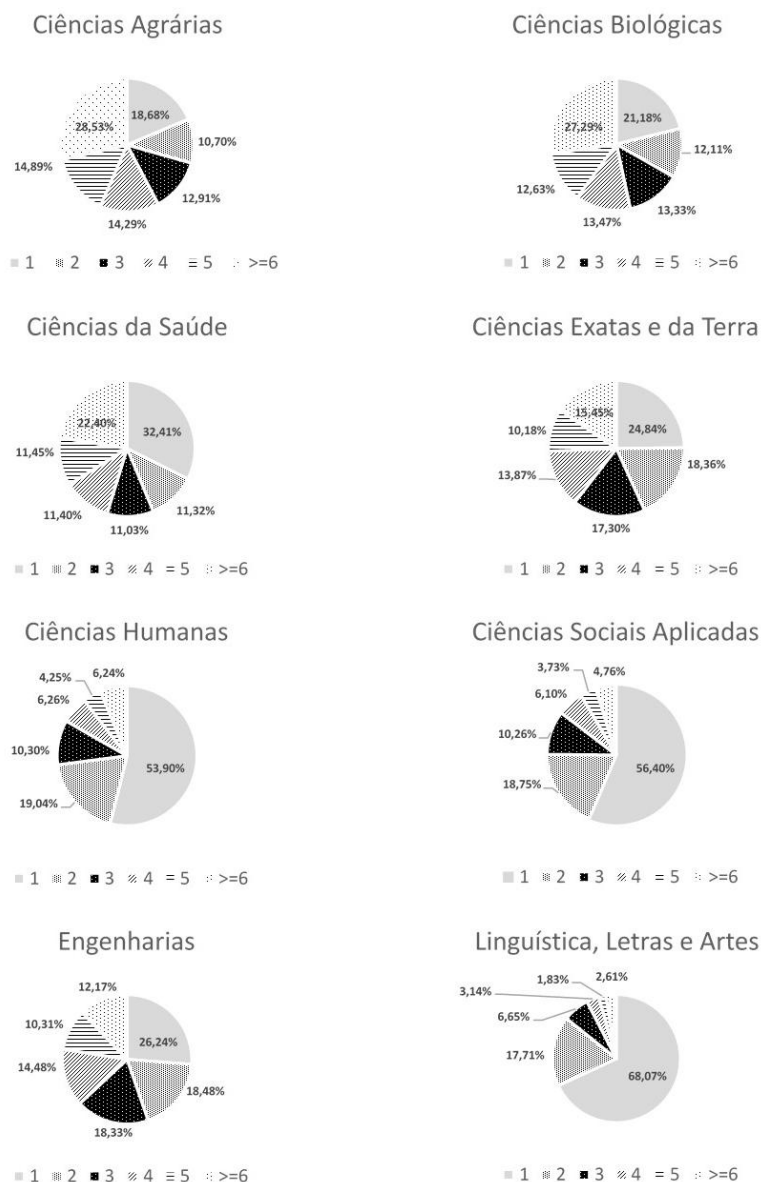
Figura 4 - Quantidade de coautores em artigos publicados em anais de congresso e periódicos do conjunto de doutores



Fonte: Elaborado pelos autores.

Observa-se ainda que 14,79% dos artigos possuem dois coautores e 12,57% têm três coautores, mostrando que a maioria dos trabalhos é realizada com a participação de poucos autores. Além disso, foi identificada a existência de artigos com centenas de coautores, fato esse influenciado por algumas áreas do conhecimento que tendem a realizar seus trabalhos de forma mais colaborativa. Para melhor visualização dessa distinção no perfil de colaboração dos pesquisadores brasileiros, é apresentada a quantidade de coautores em cada grande área do conhecimento (ver Figura 5).

Figura 5 - Quantidade de coautores em artigos publicados de cada grande área



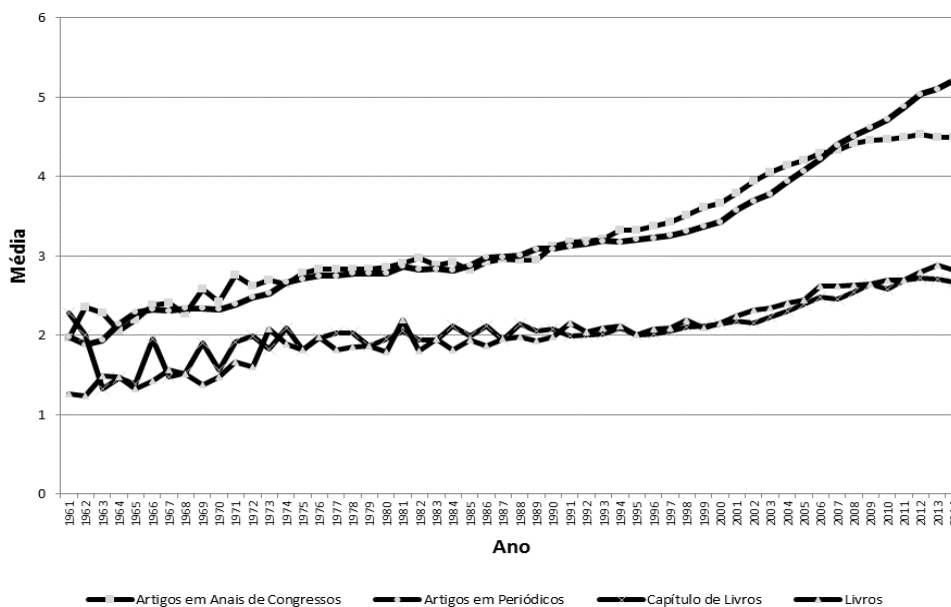
Fonte: Elaborado pelos autores.

Tendo em vista que podem ser informadas diversas grandes áreas de atuação nos currículos de cada indivíduo, a primeira grande área informada foi considerada como a principal.

Como pode ser observado, os perfis de publicação dos pesquisadores em cada uma das grandes áreas são distintos. Pesquisadores das grandes áreas de Ciências Agrárias e Ciências Biológicas tendem a realizar seus trabalhos de forma mais colaborativa, uma vez que uma quantidade considerável de seus trabalhos foi realizada com seis ou mais coautores. Outras grandes áreas como Ciências Humanas, Ciências Sociais Aplicadas e Linguística, Letras e Artes possuem pesquisadores que, em sua maioria, publicam de forma isolada, ou seja, sem colaboradores. Em especial, destaca-se a grande área de Linguística, Letras e Artes, em que 68,07% dos seus artigos foram realizados por um único autor e 17,71%, com dois coautores.

No entanto, de maneira geral, a colaboração científica entre os doutores brasileiros tem aumentado nos últimos anos, conforme mostrado na Figura 6.

Figura 6 - Média de autores por publicação com pelo menos um doutor



Fonte: Elaborado pelos autores.

A publicação de livros e capítulos de livro tem comportamento similar quando se analisa o número de autores envolvidos, tendo havido um tímido crescimento na quantidade média de autores nos últimos anos. Já os artigos

publicados em anais de congresso e em periódicos tinham médias de autores similares até 2007, quando os artigos em periódico passaram a ser produzidos de forma mais colaborativa. Atualmente, a média de autores dos artigos em periódico é de 5,22, já a dos artigos em anais de congresso é de 4,49. É importante destacar que, como as publicações estão sendo realizadas cada vez mais de forma colaborativa, esperava-se que essa colaboração influenciasse no número de trabalhos publicados de forma mais significativa, já que grupos mais colaborativos tendem a ser bem mais produtivos, o que não foi efetivamente observado.

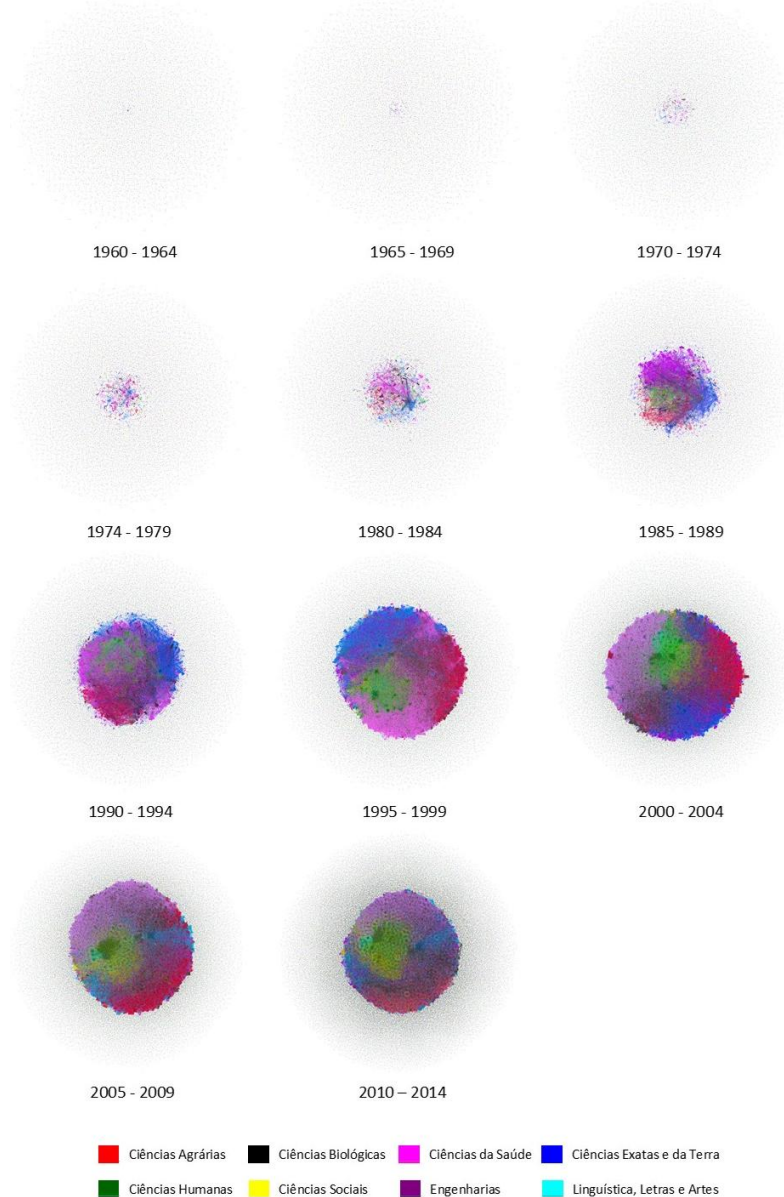
A análise da colaboração científica, entretanto, pode ser melhor evidenciada com a caracterização das respectivas redes de colaboração. Tais redes possibilitam analisar, com a adoção de métricas específicas, como a colaboração tem evoluído ao longo dos anos. Neste trabalho, foram caracterizadas redes temporais e, para suas análises, algumas métricas de análise de redes sociais foram aplicadas.

As métricas adotadas (grau médio dos nós, total de nós no componente gigante, densidade da rede, diâmetro da rede e caminho mínimo médio) são métricas clássicas, usualmente adotadas por diversos trabalhos que analisam redes de colaboração (SZWARCFITER, 1986; NEWMAN, 2003; LEMIEUX; OUIOMET, 2008; SCOTT, 2009; WASSERMAN; FAUST, 2009), de forma que se possa observar como a colaboração científica evolui ao longo dos anos. Tendo em vista a realização de uma análise temporal dessas redes, foram definidas janelas temporais. Tais janelas cobrem um período de cinco anos e iniciam-se em 1960, ano que engloba as primeiras publicações identificadas na Plataforma Lattes, e terminam em 2014, último ano considerado em todas as análises apresentadas neste trabalho. Para cada janela de tempo, foram caracterizadas as redes contendo todos os indivíduos que possuíam doutorado concluído no período em questão, independentemente de seu tempo de titulação como doutor. Somente a produção científica do conjunto de doutores publicada no período foi considerada para a caracterização das redes de colaboração. As cores dos nós que compõem as redes estão relacionadas às suas grandes áreas de atuação.

Para a identificação das colaborações, um método denominado Identification of Scientific Collaboration (ISColl), proposto por Dias e Moita (2015), foi utilizado. Já, na Figura 7:

Para caracterização das redes que contêm todos os elementos, foram adotados métodos que visam concentrar os nós mais conectados no centro da rede e, conseqüentemente, aqueles nós com menor quantidade de ligações ou isolados são deslocados para as extremidades das redes [...] (DIAS; MOREIRA; DIAS, 2018, p. 70).

Figura 7 - Evolução das redes de colaboração científica dos doutores ao longo dos anos

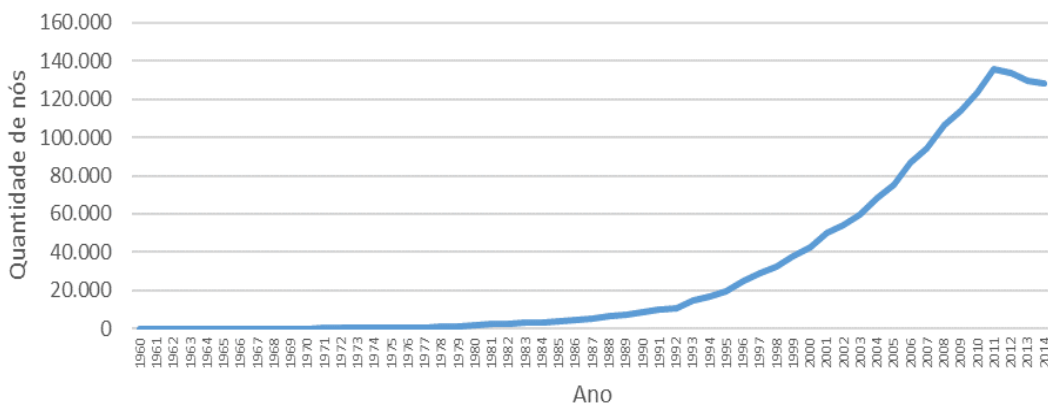


Fonte: Elaborado pelos autores.

Como pode ser observado, as colaborações entre os doutores iniciam-se no início da década de 1960 e são intensificadas a partir da década de 1990. É possível observar como as colaborações, em algumas grandes áreas, estabilizaram-se ao longo do tempo, por exemplo, as grandes áreas de Ciências da Saúde e Ciências Biológicas, bem como outras que passaram a ter um maior grau de colaboração nos últimos anos, sendo elas as grandes áreas de Ciências Exatas e da Terra e Ciências Agrárias.

A Figura 8 apresenta a evolução temporal do tamanho do componente gigante da rede de doutores. É possível observar um crescimento mais intenso desse componente a partir do início da década de 1990, que se estende até o ano de 2011, período coincidente com o crescimento da produção científica brasileira. Esse componente cresce ao longo dos anos, já que outros componentes vão sendo conectados a ele. Verifica-se ainda que, em 2011, o componente gigante representa cerca de 61% da rede. Após, o componente gigante tem redução em todos os outros anos, período também coincidente com a queda da produção científica brasileira (DIAS, 2016).

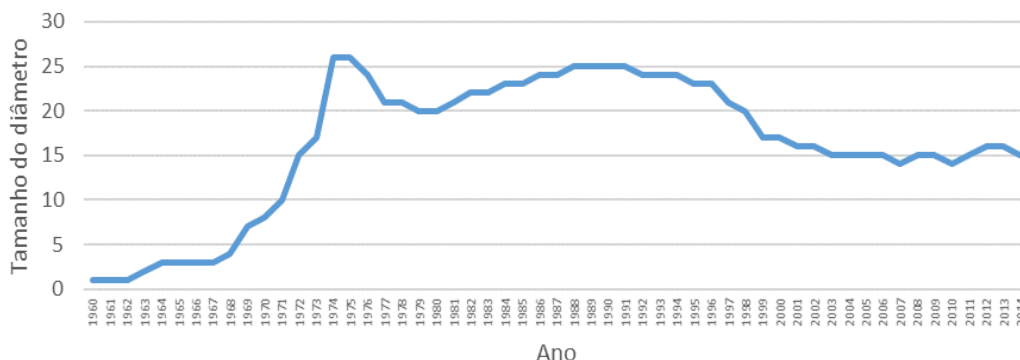
Figura 8 - Evolução temporal do componente gigante



Fonte: Elaborado pelos autores.

Na Figura 9, é possível observar a evolução temporal do diâmetro da rede, o qual é influenciado pela evolução do componente gigante. Uma vez que o componente gigante se torna mais conectado, o diâmetro tende a diminuir.

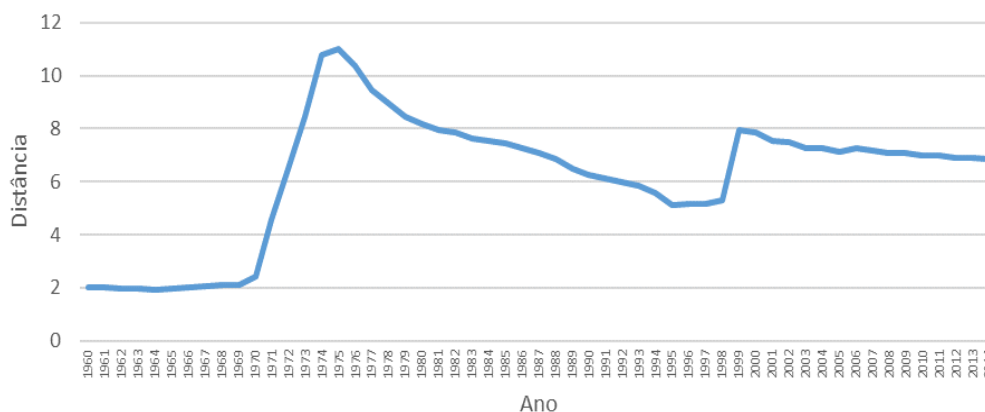
Figura 9 - Evolução temporal do diâmetro da rede



Fonte: Elaborado pelos autores.

A Figura 10 apresenta a evolução temporal do caminho mínimo médio da rede. Observa-se que, inicialmente, essa medida segue um padrão semelhante ao do diâmetro da rede. No entanto, após atingir seu maior valor (11,02) em 1975, tem redução constante até 1995. Posteriormente, volta a ter aumento no final da década de 1990 e, a partir do ano 2000, apresenta uma redução estável. Tal redução é reflexo do crescimento da conexão do componente gigante. É importante destacar o aumento do caminho mínimo médio no quinquênio compreendido entre 1995 a 1999, período que também coincide com um significativo aumento na quantidade de arestas e, ainda, com o lançamento e padronização do Currículo Lattes (agosto de 1999) como sendo o formulário de currículo a ser utilizado no âmbito do Ministério da Ciência e Tecnologia e CNPq.

Figura 10 - Evolução temporal do caminho mínimo médio



Fonte: Elaborado pelos autores.

A Tabela 1 apresenta uma sumarização das redes caracterizadas e das métricas adotadas em todos os períodos analisados. É possível observar a evolução das redes, sendo que a rede inicial possuía apenas 8.186 nós e um componente gigante composto por poucos nós. Ressalta-se ainda a quantidade de arestas que, a exemplo da produção científica, tem crescimento em todos os períodos, com exceção do quinquênio compreendido entre 2010 a 2014, em que a produção científica coincidentemente apresenta redução.

Tabela 1 - Resultado das métricas adotadas

Período	Métrica									
	Total de Nós	Total de Arestas	Grau Médio dos Nós	Total de Nós no Componente Gigante	% de Nós no Componente Gigante	Total de Arestas no Componente Gigante	Densidade da Rede	Diâmetro da Rede	Caminho Mínimo Médio	Total de Componentes Isolados
1960 - 1964	8.186	33	0,008	6	0,07%	9	0,00000098	3	1,939	8.140
1965 - 1969	8.735	129	0,028	12	0,13%	14	0,00000338	7	2,101	8.576
1970 - 1974	10.876	939	0,167	300	2,76%	455	0,00001587	26	10,804	10.036
1975 - 1979	13.624	3.176	0,455	1.333	9,78%	2.618	0,00003422	20	8,449	11.553
1980 - 1984	18.113	9.200	0,997	3.200	17,67%	8.283	0,00005608	23	7,551	13.711
1985 - 1989	24.645	26.781	2,145	7.310	29,66%	25.817	0,00008818	25	6,484	15.948
1990 - 1994	36.125	104.560	5,740	16.402	45,40%	103.691	0,00016024	24	5,601	18.409
1995 - 1999	58.870	688.180	5,970	37.970	64,50%	687.757	0,00039714	17	7,965	20.167
2000 - 2004	96.928	1.011.010	6,769	68.677	70,85%	1.009.655	0,00021522	15	7,256	27.443
2005 - 2009	151.140	1.225.113	7,380	114.292	75,62%	1.218.245	0,00010726	15	7,089	36.170
2010 - 2014	222.720	997.751	6,725	128.059	57,50%	992.195	0,00004022	15	6,875	67.240

Fonte: Elaborado pelos autores.

“Pode-se afirmar que a análise temporal das redes de colaboração científica do conjunto de doutores com currículos cadastrados na Plataforma Lattes possibilita obter uma visão de como ocorreu o processo de coautoria [...]” (DIAS; MOREIRA; DIAS, 2018, p. 73) ao longo dos anos. Tendo como base as publicações científicas desses doutores, a primeira colaboração remete ao início dos anos 1960 e, como pôde ser observado, a cada quinquênio, o número de

colaborações tem aumentado, tendo como base a quantidade de arestas e o grau médio das redes que tiveram crescimento significativo, com exceção do último período analisado, em que a quantidade de componentes isolados cresceu de forma substancial. Destaca-se o período compreendido entre 2005 a 2009, que possui o maior valor do grau médio e a maior quantidade de arestas, podendo ser considerado o período de maior colaboração da comunidade científica brasileira. Foi possível ainda identificar que, apesar das redes possuírem baixa densidade, ao longo dos anos, a quantidade de componentes isolados, que correspondem a autores que publicaram sem nenhuma colaboração ou que não publicaram no período, teve crescimento bem inferior ao crescimento dos nós que compõem as redes, exceto no último quinquênio (2010-2014).

“Também foi possível verificar como alguns pesquisadores têm trabalhado em colaboração de forma muito intensa. Analisando as arestas mais densas das redes [...],” desde o primeiro período analisado, percebe-se que alguns pesquisadores trabalham em colaboração com uma frequência muito alta (DIAS; MOREIRA; DIAS, 2018, p. 73). Em geral, tais colaborações “[...] são entre pesquisadores que atuam na mesma área ou em áreas correlatadas, que possuem a mesma formação acadêmica e, em determinadas situações, orientados pelos mesmos orientadores.” (DIAS; MOREIRA; DIAS, 2018, p. 74).

As análises apresentadas possibilitam compreender como ocorreu a evolução da colaboração científica brasileira de forma temporal, com quinquênios que compreendem todas as publicações encontradas na Plataforma Lattes, considerando o conjunto de doutores analisados.

5 Considerações finais

Como resultado deste estudo, foi possível verificar como a colaboração científica do conjunto de doutores brasileiros evoluiu ao longo dos anos. Para isso, foram considerados quinquênios desde a primeira publicação identificada até os dias atuais. Foi possível observar como os pesquisadores de determinadas grandes áreas do conhecimento tendem a realizar seus trabalhos de forma mais colaborativa e ainda como essas grandes áreas intensificaram seus trabalhos em

colaboração em cada período analisado. Além disso, verificando as arestas mais densas das redes caracterizadas, foi identificado como pequenos grupos têm trabalhado em colaboração de forma intensa, aumentando de forma impactante o número de publicações de seus pesquisadores, tendo em vista a participação conjunta dos mesmos coautores, geralmente das mesmas áreas de atuação e das mesmas instituições.

Referências

- BARABÁSI, A. L.; ALBERT, R. Emergence of scaling in random networks. **Science**, Washington, v. 286, n. 5439, p. 509-512, 1999.
- BOAVENTURA, M. et al. Caracterização temporal das redes de colaboração científica nas universidades brasileiras: anos 2000-2013. In: BRAZILIAN WORKSHOP ON SOCIAL NETWORK ANALYSIS AND MINING, 3., 2014, Brasília. **Anais...** Brasília: Sociedade Brasileira de Computação, 2014.
- DIAS, T. M. R.; MOITA, G. F. Um método para identificação de colaborações em grandes bases de dados científicos. **Em Questão**, Porto Alegre, v. 21, n. 2, p. 140-161. 2015.
- DIAS, T. M. R.; MOREIRA, T. H. J.; DIAS, P. M. Caracterização e análise das redes de colaboração científica dos bolsistas de produtividade em pesquisa do CNPq. In: Brazilian Workshop on Social Network Analysis and Mining, 7., 2018, Natal. **Anais...** Natal: BraSNAM, 2018. p. 64-75.
- DIAS, T. M. R. **Um estudo sobre a produção científica brasileira a partir de dados da plataforma Lattes**. 2016. (Doutorado em Modelagem Matemática e Computacional) - Centro Federal de Educação Tecnológica de Minas Gerais, Belo Horizonte, 2016.
- DIGIAMPIETRI, L. A. et al. Análise da rede de relacionamentos dos doutores brasileiros. In: BRAZILIAN E-SCIENCE WORKSHOP, 8., 2014, Brasília. **Anais...** Brasília: Sociedade Brasileira de Computação, 2014.
- DIGIAMPIETRI, L. **Análise da rede social acadêmica brasileira**. Monografia (Livre Docência) - Escola de Artes Ciências e Humanidades da Universidade de São Paulo. São Paulo, 2015.
- KUMAR, S. Co-authorship networks: a review of the literature. **Aslib Journal of Information Management**, Bingley, v. 67, n. 1, p. 55-73, 2015.

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

LEMIEUX, V.; OUMET, M. **Análise estrutural das redes sociais**. Lisboa: Instituto Piaget, 2008.

LIBEN-NOWELL, D.; KLEINBERG, J. The link-prediction problem for social networks. **Journal of the American Society for Information Science and Technology**, New York, v. 58, n. 7, p. 1019-1031, 2007.

MEADOWS, A. J.; LEMOS, A. A. B. **A comunicação científica**. Brasília: Briquet de Lemos, 1999.

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, 1., 2012, Curitiba **Anais...** Curitiba: Brazilian Computing Society, 2012.

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

NEWMAN, M. E. J. Scientific collaboration networks: I network construction and fundamental results. **Physical Review E**, Maryland, v. 64, n. 1, p. 0161311-0161318, 2001b.

NEWMAN, M. E. J. Scientific collaboration networks: II shortest paths, weighted networks, and centrality. **Physical Review E**, Maryland, v. 64, n. 1, p. 0161321-0161327, 2001c.

NEWMAN, M. E. J. The structure and function of complex networks. **SIAM Review**, Philadelphia, v. 45, n. 2, p. 167-256, 2003.

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

PETERSEN, A. M. et al. Persistence and uncertainty in the academic career. **Proceedings of the National Academy of Sciences**, Washington, DC, v. 109, n. 14, p. 5213-5218, 2012.

PEREZ-CERVANTES, E. et al. Using link prediction to estimate the collaborative influence of researchers. In: IEEE INTERNATIONAL CONFERENCE ON ESCIENCE, 9., 2013, Beijin. **Anais...** Beijin: IEEE, p. 293-300, 2013.

SZWARCFITER, J. L. **Grafos e algoritmos computacionais**. 2. ed. Rio de Janeiro: Campus, 1986.

SCOTT, J. **Social network analysis: a handbook**. 2. ed. London: SAGE, 2009.

WASSERMAN, S.; FAUST, K. **Social network analysis: methods and applications**. 19. ed. Cambridge: Cambridge University Press, 2009.

A study on the network of the scientific collaboration of Brazilian researchers with curricula registered in the Lattes Platform

Abstract: Analyses of scientific collaboration networks have been widely explored in researches within different knowledge fields for quite some time, taking into account its ability to identify how groups of researchers are performing their work collectively. These studies make it possible to identify how collaboration occurs between individuals through investigations based on social network metrics. In this paper, at first, a general characterization of the analyzed data is presented, and afterward, a temporal analysis of the Brazilian PhDs scientific collaboration network with registered curricula in the Lattes Platform is conducted. For this purpose, all their scientific publications were used for the identification and characterization of the collaboration networks. As a result, it is possible to visualize how the scientific collaboration of the analyzed set was intensified over the years, highlighting the areas of activity of each Brazilian PhD.

Keywords: Collaboration Networks. Network Analysis. Scientific production. Lattes Platform.

Recebido: 22/03/2018

Aceito: 11/07/2018

¹ SONNENWALD, D. H. Scientific collaboration. *Annual Review of Information Science and Technology*, White Plains, v. 41, n. 1, p. 643-681, 2007. Apud Dias, Moreira, Dias (2018).

² STRÖELE, V.; ZIMBRÃO, G.; SOUZA, J. M. Análise de redes sociais científicas: modelagem multi-relacional. In: BRAZILIAN WORKSHOP ON SOCIAL NETWORK ANALYSIS AND MINING, 1., 2012, Curitiba: Sociedade Brasileira de Computação. *Anais...* Curitiba, 2012. Apud Dias, Moreira, Dias (2018).

³ BARABÁSI, A. L. et al. Evolution of the social network of scientific collaborations. *Physica A: statistical mechanics and its applications*, Amsterdam, v. 311, n. 3, p. 590-614, 2002. Apud Dias, Moreira, Dias (2018).

⁴ MENA-CHALCO, J. P. et al. Brazilian bibliometric coauthorship networks. *Journal of the Association for Information Science and Technology*, New York, v. 65, n. 7, p. 1424-1445, 2014. Apud Dias, Moreira, Dias (2018).