em questão
E-ISSN 1808-5245

Colaboração científica intraorganizacional: análise de redes por coocorrência de palavras-chave

Nathalia Mendes Gerotti Franco
Mestre; Universidade Federal de São Carlos, São Carlos, SP, Brasil;
nath-jau@bol.com.br

Leandro Innocentini Lopes de Faria
Doutor; Universidade Federal de São Carlos, São Carlos, SP, Brasil;
leandro@ufscar.br

Resumo: Pesquisas sobre os impactos da colaboração em pesquisas científicas vêm crescendo nos últimos anos, principalmente com um olhar sobre as implicações políticas e a destinação de recursos financeiros. A colaboração científica pode contemplar diversos níveis de análise, dando ênfase nos aspectos internos – intrainstitucional, (intra)nacional – ou em situar o que está sendo analisado em um contexto externo – interinstitucional, internacional e outros. No entanto, a questão da colaboração intraorganizacional e identificação de possíveis colaborações têm pouca abordagem e estudos. Não há conhecimento de estudos que tratem simultaneamente desses dois temas visando à identificação de oportunidades de colaboração interna em uma instituição. Assim, este trabalho apresenta um panorama da colaboração científica entre os docentes de uma universidade federal do estado de São Paulo, e expõe um procedimento para identificar pesquisadores potenciais para futuras colaborações. A identificação dessas colaborações em potencial baseia-se no acoplamento de autores por temáticas através das palavras-chave lançadas pelos próprios docentes em seus Currículos Lattes. Foram identificados 635 docentes e 2.678 palavras. E, como resultado, é apresentada uma rede de afiliação por palavras-chave, onde é possível identificar os pesquisadores com alto potencial de colaboração, o que pode auxiliar os próprios docentes a identificarem seus possíveis parceiros, além de poder ser utilizada por órgãos institucionais estratégicos para fomentar a colaboração entre os professores internos à organização.

Palavras-chave: Análise de Redes. Colaboração científica. Coocorrência de palavras. Acoplamento.

1 Introdução

A medição da ciência sempre foi importante, Horácio dizia que há uma medida para todas as coisas e para Heike K. Onnes, descobridor da supercondutividade, “medir é saber” (VAN RAAN, 2004). A ideia de examinar a literatura remonta ao início do século, Okubo (1997) relata que em 1917, Cole e Eales publicaram

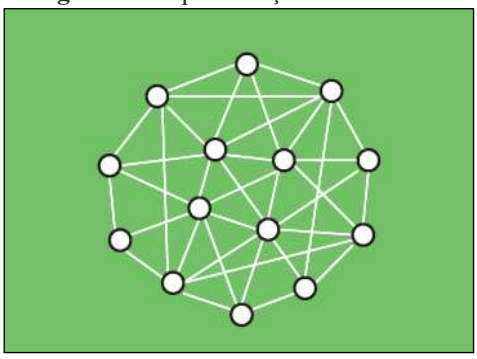
uma análise estatística da história da anatomia comparada, e esta foi um marco na história da análise bibliométrica. Bibliometria é um termo que surgiu em 1969, com Alan Pritchard, e trata-se de um ramo da Ciência que analisa a atividade científica e tecnológica através de dados derivados da literatura científica e de patentes, envolvendo o estudo quantitativo da produção, disseminação e uso da informação registrada, com o principal objetivo de desenvolver indicadores cada vez mais confiáveis (FARIA, 2001; HAYASHI et al., 2007; OKUBO, 1997).

Dentre muitos indicadores bibliométricos destaca-se aqui o indicador de coautoria, que é utilizado para identificar e estudar a colaboração científica acadêmica. Colaboração e coautoria não precisam ser sinônimos, pois a coautoria é apenas uma parte da colaboração científica, uma vez que não mede a colaboração na sua totalidade e complexidade (VANZ, 2009). Para Lima, Velho e Faria (2007), "Quando se fala de cooperação científica na visão da bibliometria, devemos ter em mente que quase sempre estamos nos referindo à análise de trabalhos publicados em coautoria." (LIMA; VELHO; FARIA, 2007, p. 54). Newman (2001) faz uma colocação pertinente quando se trata de colaboração, alegando que, apesar da importância da comunicação escrita na ciência, através de documentos e artigos, pode-se dizer que a maior parte da comunicação científica ainda ocorre, efetivamente, por conversa privada, e atinge a maioria dos membros de uma rede através de conversas particulares e de maneira muito rápida. Assim, ainda que não publiquem conjuntamente, estes atores podem ter colaborações informais entre si que não sejam possíveis de verificar apenas através da análise de rede baseada em publicação de artigos com coautoria, pois estas relações informais não são passíveis de análise quantitativa. Dessa forma, as redes baseadas em coautoria servem de instrumento para analisar as colaborações e parcerias científicas e tecnológicas, permitindo uma visão dos padrões de cooperação entre indivíduos e organizações (SAMPAIO et al., 2015).

Assim, colaboração científica pode ser definida como uma interação entre dois ou mais pesquisadores que compartilham recursos intelectuais, econômicos e/ou físicos, trabalhando juntos em um projeto de pesquisa, dentro de um contexto social, para atingir o mesmo objetivo comum de produzir novos conhecimentos científicos (SAMPAIO et al., 2015; KATZ; MARTIN, 1997; VANZ, 2009).

Este estudo das ligações entre atores pode ser realizado através da Análise de Redes Sociais (ARS), que é um estudo aprofundado dos pontos e conexões das redes, a fim de mapear os relacionamentos através do fluxo de informações (BARROS; GAMA; FERNANDES, 2014), e, além de outras funcionalidades, pode identificar e representar as coautorias e o acoplamento de palavras. Sucintamente, rede é um conjunto de pontos (ou nós) que se ligam a outros pontos por meio de linhas. As redes, para Borgatti, Everett e Johnson (2013), são uma forma de pensar sobre sistemas sociais e focam as relações entre os entes que compõem o sistema, chamados de atores ou nós. Estes nós possuem características, denominadas 'atributos', que distinguem um nó do outro, e podem ser indivíduos ou coletividades. E as relações entre os nós, que são os laços ou links, também com características próprias, entrelaçam-se criando uma teia. As espessuras desiguais dos traços mostram as diferenças do relacionamento entre os atores, ou seja, quanto mais espessa, maior é a qualidade desta relação. A Figura 1 representa uma rede simples, com os nós e laços, sendo as linhas mais importantes que os pontos, uma vez que fazem as conexões. Um conjunto de nós dispersos nada representa se não houver alguma conexão entre eles. Assim "Muitos pontos pouco interligados têm menos qualidade de rede do que poucos pontos profundamente conectados. A medida da rede é o número de conexões, não de pontos." (WWF-Brasil, 2003, p.19).

Figura 1 - Representação de uma rede



Fonte: WWF-Brasil (2003).

Atualmente, com o aumento da capacidade de processamento numérico e o auxílio de softwares especializados e mais sofisticados na criação de matrizes e na geração de imagens e gráficos de redes, o cruzamento destes dados pode ser feito de maneira mais rápida e os sistemas podem ser simulados e avaliados em

maiores detalhes (SAMPAIO et al., 2015). As redes de coautoria remetem à participação ativa dos envolvidos na produção dos trabalhos em colaboração, que é uma tendência do mundo contemporâneo, pois agrega diversas habilidades, interesses e demandas dentro da estrutura do campo científico (BELLO, 2013). No entanto, segundo Faria (2001), há outros critérios que podem ser usados para estabelecer os relacionamentos em redes de autores, chamados de indicadores de ligação ou relacional, que são baseados em coocorrências de publicações, citações e palavras-chaves. Ou seja, os atores podem nunca ter trabalhado juntos, mas publicam trabalhos com as mesmas palavras-chaves e temas de igual interesse. Este indicador auxilia na elaboração de mapas responsáveis pela descrição do conhecimento e relacionamento entre pesquisadores, instituições e países. Segundo Lu e Wolfram (2012), atualmente, existem cinco abordagens utilizadas para medir relações entre autores, que podem ser: citação direta, análise de cocitação, análise de coautoria, análise de acoplamento bibliográfico e análise de co-palavra ou *co-word*. Esta análise baseada em *co-word* permite estabelecer as possíveis forças de ligação entre os termos e assuntos que coocorrem, proporcionando, desta forma, a criação de um mapa do conhecimento. E pode ser obtida através das palavras-chave contidas nos artigos, "[...] palavras do título, palavras do resumo, palavras do texto integral, palavras presentes na classificação dos artigos ou o próprio código de classificação do artigo, neste último caso sendo chamada de análise de co-classificação." (CALLON, 1993[1] apud FARIA, 2001, p. 45).

Esta ideia de relacionar dois autores que compartilham as mesmas temáticas, segundo Kessler (1963[2] apud LU; WOLFRAM, 2012), é um acoplamento bibliográfico, e quanto mais semelhanças dois autores têm em comum, mais intimamente relacionados eles devem ser. Essas possibilidades de análise fornecem dados sobre o impacto científico de grupos de pesquisas ou países na comunidade nacional e internacional, dando cobertura a todas as áreas científicas.

Ainda há poucos estudos conhecidos que abordam a colaboração intrainstitucional. Newman (2004) investigou a aplicação de um método de detecção de comunidades à rede de coautorias do Instituto Santa Fe, e Pepe e

Rodriguez (2010) estudaram padrões de coautoria do *Center for Embedded Networked Sensing*, centro de pesquisa da Universidade da Califórnia. Gazni e Didegah (2011) estudaram diversos aspectos da colaboração científica da Universidade de Harvard, inclusive a colaboração intrainstitucional e seu impacto em termos de citações recebidas quando comparada às publicações em colaboração interinstitucional. Savić, Ivanović e Surla (2017); Bellanca (2009) e Birnholtz e outros (2013) analisaram a colaboração nas Universidades de Novi Sad, York, Salerno e Cornell.

Há ao menos um estudo, desenvolvido por Boyack (2009), que visa à identificação de potenciais colaborações como uma contribuição à elaboração de estratégia e planejamento institucionais. No entanto, a colaboração tratada nessa pesquisa está no nível interinstitucional, uma vez que foca na identificação de oportunidades de colaboração entre o Laboratório Sandia e outras instituições. Também, o estudo não diferencia colaborações já existentes de colaborações ainda inéditas, tratando a todas como potenciais. Finalmente, o método de identificação de oportunidades adotado não utiliza coautorias como critério de medida da ligação entre as instituições e baseia-se na elaboração de redes de publicações com temas comuns, as quais são associadas, posteriormente, às instituições. Assim, o objetivo desta pesquisa é desenvolver um procedimento baseado na análise bibliométrica de publicações científicas e na análise de redes sociais para a identificação de oportunidades de colaboração intrainstitucional. A base de dados escolhida para recuperar as informações dos docentes foi o Currículo Lattes, presente na Plataforma Lattes, com acesso público por meio da ferramenta *SyncLattes*, desenvolvida por Matias (2015).

A Plataforma Lattes é um Sistema de Informações que tem o objetivo de integrar currículos, grupos de pesquisa e Instituições brasileiras numa plataforma única, a Plataforma Lattes (PL). O nome Lattes foi dado em homenagem ao físico César Mansueto Giulio Lattes, mais conhecido como César Lattes, que foi um dos maiores cientistas brasileiro e tornou-se um ícone mundial na produção científica. No final do ano de 2002, o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) passou a licenciar gratuitamente o software e a fornecer consultoria técnica para sua implantação nos países da América Latina,

como Colômbia, Equador, Chile, Peru e Argentina, além de Portugal e Moçambique. Brito, Quoniam e Mena-Chalco (2016) apontam algumas limitações referentes à Plataforma Lattes, que, por ser atualizada pelos próprios usuários, não possui padronização de preenchimento, podendo conter informações incorretas ou incompletas. Além disso, pesquisadores vinculados a programas de Pós-Graduação tem uma obrigatoriedade maior de manter os dados atualizados, enquanto outros ficam anos sem atualizar. Os autores ainda falam sobre o grau de confiabilidade das informações, pois não há supervisão dos dados cadastrados (BRITO; QUONIAM; MENA-CHALCO, 2016). Mas, apesar destes entraves, é uma plataforma inédita que reúne e integra informações acadêmicas, profissionais, institucionais e bibliográficas com 3,5 milhões[3] de currículos cadastrados, sendo, aproximadamente, 598.900 currículos de mestres e doutores disponíveis (CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO, 2016). Ressalta-se ainda que os dados desta pesquisa foram todos retirados da Plataforma Lattes, sem a utilização de outras bases de dados ou repositório institucional.

A Plataforma Lattes permite a atribuição de palavras-chave às publicações cadastradas. As palavras-chave fornecem um breve resumo do conteúdo da publicação, podendo ser entendidas como a menor unidade que sintetiza e identifica o conteúdo de todo o texto, permitindo uma visão simplificada do documento. “Palavras-chave são particularmente úteis porque podem ser interpretadas individualmente e independentemente umas das outras.” (DIAS, 2004, p. 35). Ao mesmo tempo em que ajudam a descrever os assuntos tratados nas publicações, devido ao fato de as palavras chaves serem preenchidas de forma livre no momento em que os pesquisadores cadastram ou atualizam seus currículos, a análise bibliométrica das mesmas, na Plataforma Lattes, representa um desafio. Não há um vocabulário controlado para a escolha das palavras-chave. Não há obrigatoriedade de atribuição de palavras-chave às publicações. Não há verificação se as palavras-chave atribuídas às publicações na Plataforma Lattes são as mesmas que os autores inseriram na publicação original. Não há definição de idioma que deve ser utilizado na escolha das palavras-chave. Autores de uma mesma publicação podem atribuir a ela palavras-chave diferentes, cada um em

seu currículo. E é possível atribuir até 6 palavras-chave por publicação. Mas este problema de análise de palavras-chave não ocorre somente na Plataforma Lattes, bases de dados e indexadores também podem, de acordo com Lu e Wolfram (2012), criar distorções e subjetividades na atribuição dos termos dos artigos indexados.

Em estudos realizados por Gil-Leiva e Alonso-Arroyo (2007), os autores debatem o uso de vocabulários controlados e palavras-chave para recuperação de informações dos artigos científicos, e constataram que, apesar de todos estes percalços enfrentados para a utilização das palavras-chaves, quase 46% das palavras-chave também aparecem em descritores que possuem vocabulário controlado. Isto significa que as palavras-chave fornecidas pelos autores são uma valiosa fonte de informação tanto para indexação humana quanto para sistemas de indexação automática de artigos de periódicos.

2 Dados e métodos

De acordo com Gil (2017, p. 25), a "[...] classificação é uma característica da racionalidade humana.", pois facilita o entendimento e os objetivos da pesquisa a ser desenvolvida. Esta pesquisa, de acordo com a sua finalidade, é classificada como pesquisa aplicada, que é a "[...] aquisição de conhecimentos com vistas à aplicação numa situação específica." (GIL, 2017, p. 27) já que se propõe a estudar as formas de colaboração científica, e aplicá-las numa instituição de ensino superior a fim de verificar o comportamento colaborativo de seus docentes. Segundo seus objetivos, é uma pesquisa descritiva, pois irá descrever e estudar as características de um determinado grupo, e "[...] descobrir a existência de associações entre variáveis, [...]" (GIL, 2017, p. 28). Para que seja possível avaliar a qualidade dos resultados obtidos numa pesquisa, Gil (2017) fala da necessidade de entender como os dados foram obtidos e quais procedimentos foram adotados em sua análise e interpretação. O sistema adotado deve levar em consideração a natureza dos dados, o ambiente de coleta e análise dos dados. Neste caso, é uma pesquisa bibliográfica (elaborada com base em material já publicado), quantitativa (amplas amostras com informações numéricas) e aplicada com

técnicas bibliométricas (análises estatísticas de publicações e atividades científicas).

O objeto de pesquisa é a Universidade Federal de São Carlos (UFSCar), pelo fato de sua relevância científica. A UFSCar tem sua fundação datada de 1968 e, atualmente, possui quatro campi: São Carlos, Araras, Sorocaba e Lagoa do Sino (UNIVERSIDADE FEDERAL DE SÃO CARLOS, 2018). Para esta pesquisa, foram analisadas as produções científicas constantes nos currículos da Plataforma Lattes de todos os docentes da UFSCar, dos quatro campi, desde sua fundação, em 1968, até dezembro de 2016. A última atualização dos dados foi realizada em outubro de 2017, para garantir, assim, que todos os docentes tenham seus documentos relativos ao ano 2016 já lançados na plataforma.

2.1 Coleta dos dados

A extração dos dados constantes na PL foi feita através da ferramenta chamada *SyncLattes*, desenvolvida por Matias (2015), que descreve detalhadamente seu funcionamento em sua dissertação. Esta ferramenta utiliza o IDLattes, que é um identificador único para cada CPF cadastrado na PL. Assim, para verificação dos IdLattes, Bassoli (2017), em sua dissertação, elaborou uma listagem com todos os docentes que são ou já foram vinculados à UFSCar, com seus respectivos departamentos e centros, e o período em que são ou foram docentes da UFSCar. Para Bassoli (2017), os dados extraídos pela ferramenta *SyncLattes* são confiáveis, já que os registros com dados inconsistentes correspondem a apenas 1,56% do total de registros de artigos identificados. Isso resulta em um alto grau de confiabilidade para os dados extraídos dos Currículos Lattes.

2.2 Tratamento dos dados

Para o tratamento e análise dos dados será utilizado o VantagePoint versão 5.0, que é uma ferramenta computacional de mineração e análise de dados para tratar e analisar grandes volumes de dados de forma rápida e eficaz (SEARCH TECHNOLOGY, 2001?). Para realizar as análises de rede, serão utilizados os softwares UCINET, versão 6.631, NetDraw, versão 2.161, e Gephi, versão 0.9.1, que apresenta as redes em 3D (três dimensões). Além do Microsoft Excel®, para elaboração e tratamento das matrizes e gráficos.

Para o desafio de demonstrar quais são as possíveis colaborações existentes que não são exploradas, ou seja, quais são os docentes que trabalham com os mesmos assuntos, mas não tem colaboração ou publicação em coautoria entre si, optou-se por trabalhar com as palavras-chave lançadas pelos docentes na PL, pois as palavras representam, de forma sucinta, os assuntos abordados e de interesse dos docentes.

Entretanto, no conjunto de dados extraído inicialmente, foram identificadas 27.999 palavras, do total de 27.107 artigos, o que representa quase a proporção de uma palavra por artigo publicado. Justificam-se, assim, os cortes feitos para reduzir a amostra de publicação para um conjunto de palavras passíveis de análise. O primeiro critério adotado foi o temporal, fazendo-se um recorte entre os anos de 2013 e 2016, pois este período corresponde à última avaliação realizada pela CAPES, onde se entende que os docentes devam ter atualizado seus currículos. Além disso, recupera os temas que os docentes estão trabalhando atualmente, uma vez que os temas de pesquisa podem se alterar ao longo dos anos. Este primeiro recorte totalizou 8.128 artigos com 8.648 palavras-chave. Posteriormente, utilizando filtros incrementais desenvolvidos, foram selecionados apenas os artigos que possuíam coautoria ao menos com um outro docente vinculado à universidade, chegou-se ao total de 2.009 artigos com 2.678 palavras. E, por fim, foram selecionados apenas os artigos que possuíam ao menos uma palavra-chave, uma vez que as palavras serão as conexões realizadas entre os docentes, totalizando 1.028 artigos com 2.678 palavras. Após realizar o *List Cleanup* das palavras, dentro do Vantage Point, foi feito um refinamento com as

cinco primeiras palavras-chave, buscando por palavras que fossem sinônimas em outras línguas, com ou sem espaçamento, grafia incorreta, totalizando assim, 2.510 palavras e 635 docentes, com os mesmos 2.009 artigos. Na Tabela 1, tem-se destaque os dois conjuntos de dados utilizados, sendo o recorte número 3 utilizado para as análises de coautoria, e o recorte número 7 para os estudos de possíveis colaborações através das palavras-chave.

Tabela 1- Recortes de dados realizados no processo metodológico

	Passos	Nº Publicações	Nº Docentes	Nº de Palavras-chave
1	Original do *SyncLattes*	85.499	1.693	50.419
2	Seleção tipo "Artigo"	28.517	1.598	28.846
3	Recorte data: 1971 a 2016	27.107	1.576	27.999
4	Recorte quadriênio: 2013-2016	8.128	1.240	8.648
5	Artigos com co-autoria	2.009	844	2.678
6	Seleção das palavras com mais de 1 ocorrência	755	478	1.759
7	*List Cleanup* das palavras	755	478	1.717

Fonte: Dados da pesquisa.

A partir deste conjunto de dados, foi elaborada uma rede de dois modos (autor x palavra-chave) que Borgatti, Everett e Johnson (2013) denominam como dados de afiliação. Para melhor entendimento, esta matriz deve ser convertida em uma matriz de dados, de um modo onde tanto as linhas quanto as colunas representem os docentes, e os valores das células indiquem o número de palavras que cada docente cita. Assim, um par de docentes estará vinculado na medida em que citam ao menos uma palavra-chave semelhante, e, na melhor das hipóteses, tem-se aí um potencial de interação ou colaboração, uma vez que quanto mais palavras um par de docentes possui em comum, maiores são as chances de se relacionarem ou colaborarem entre si, pois indica interesses em comum. Borgatti, Everett e Johnson (2013) explicam como construir esta matriz de forma matemática: deve-se multiplicar a matriz de dois modos por sua transposição:

$$C_{ij} = \sum_{k} X_{ik} X_{jk}$$

Ou seja, para cada par de linhas, olha-se para cada coluna e conta-se o número de vezes que ambas são iguais a 1. Quando um par de docentes tem uma

palavra em comum conta-se 1, e se nunca lançaram a mesma palavra na PL, eles não estão conectados.

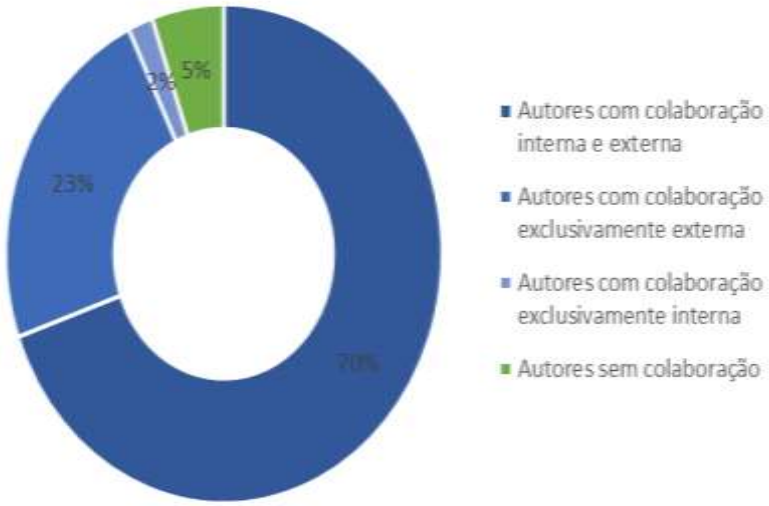
Para chegar às oportunidades de colaboração intrainstitucionais, foi necessário filtrar da rede de autores ligados pela ocorrência de palavras os autores que já colaboravam. Para isso, foi importada a matriz "Autores x Autores (1-mode) " de relação entre docentes por palavras, e a matriz de coautoria do mesmo período para o Microsoft Excel. As duas matrizes foram combinadas para a geração de uma terceira matriz, denominada "Autores x Autores (1-mode) sem colaboração", onde apenas os autores ligados por palavras e sem coautoria entre si deveriam ser apontados.

3 Resultados e discussão

Segundo Bello (2013), observar redes de colaboração permite visualizar as relações constituídas pelos pesquisadores e promover novos direcionamentos no âmbito das publicações científicas. Colaborar tem uma função social, pois é através disso que o pesquisador estabelece laços, acumula créditos científicos e fornece visibilidade aos campos científicos a que pertence. Vale lembrar que o layout das redes é mutável e o espaço e tempo devem ser demarcados, pois as relações continuam acontecendo, fortalecendo e enfraquecendo as ligações.

Para entender como se dá a colaboração científica entre os docentes da UFSCar, o Gráfico 1 apresenta os totais de autores com colaboração, seja interna ou externa, e o total de docentes que não apresentam nenhuma colaboração. Lembrando que se considera aqui, como colaboração, as coautorias preenchidas pelos docentes na PL.

Gráfico 1 - Comparação entre a colaboração científica interna dos docentes da universidade e a colaboração científica com parcerias externas à universidade

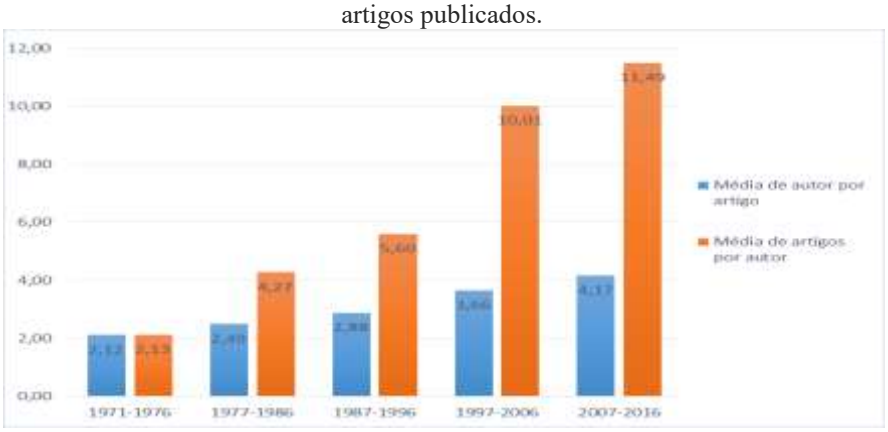


Fonte: Dados da pesquisa.

A partir deste gráfico é possível notar que a grande maioria dos docentes trabalha em colaboração, ou seja, 95% dos docentes trabalham em parcerias, sejam elas internas, externas ou ambas. Supõe-se que estes contatos externos e estas parcerias podem melhorar o fluxo de informação dentro da Universidade, com o compartilhamento de novas tecnologias, métodos e pesquisas. Entretanto, os 5% dos docentes que não possuem qualquer tipo de colaboração, podem não cooperar com a disseminação de conhecimento, pois não tem o compartilhamento das informações com outros pesquisadores.

Outro dado interessante para entender a evolução da colaboração científica intraorganizacional é verificar a relação entre a média de artigos publicados por autor e a média de autores colaboradores por artigo, separados por década, conforme o Gráfico 2.

Gráfico 2 - Média de artigos publicados por autor em comparação com a média de autores por artigos publicados.

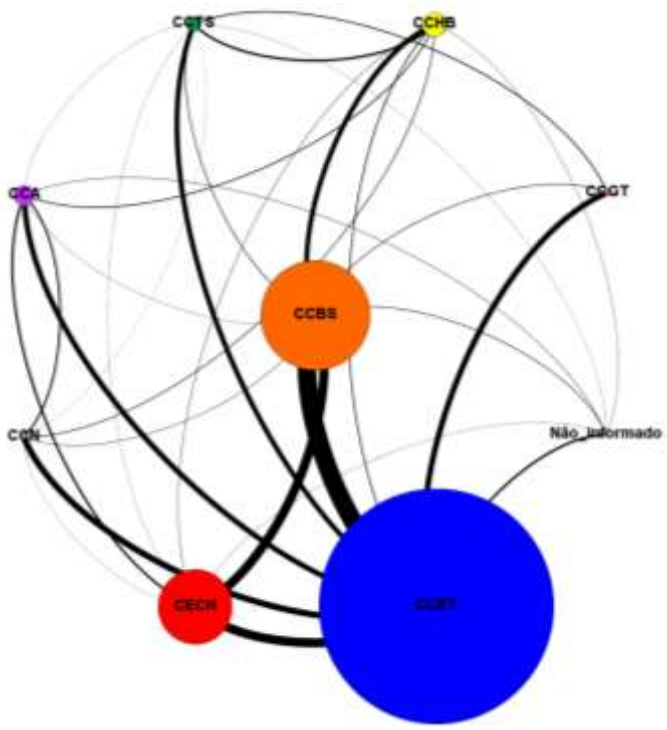


Fonte: Dados da pesquisa.

As barras laranja representam a média de artigos que cada docente publica, passando de dois artigos, no primeiro período, para 11 na última década analisada, ou seja, os docentes estão publicando cinco vezes mais que publicavam no início da fundação da UFSCar. Também houve um acréscimo na média de colaboradores por artigos, que dobrou com o passar dos anos, passando de dois para quatro autores na última década. Fato que corrobora os estudos de Melin e Persson (1996), que afirmam que houve um grande aumento de artigos em coautoria entre cientistas individuais, e, também, entre instituições de pesquisa. Isto nos leva a assumir que a colaboração científica tornou-se um pré-requisito para a ciência moderna nos últimos vinte anos, em função das dinâmicas internas da ciência e iniciativas de políticas científicas. Se 95% dos trabalhos produzidos dentro da UFSCar são em coautoria com pesquisadores internos ou de outras instituições de pesquisa, tem-se, assim, uma rede de cientistas que interagem e comunicam-se entre si. Dessa forma, foi elaborada uma rede, Figura 2, representando a colaboração entre os Centros da UFSCar. As cores representam cada Centro da universidade e o tamanho do nó é proporcional ao número de artigos publicados por cada Centro, a distribuição utilizada foi a Fruchterman Reingold, cujo algoritmo, segundo Sampaio e outros (2015), faz uma distribuição circular dos vértices, colocando os elementos com maior centralidade no centro da esfera. O Centro de Ciências Exatas e de Tecnologia (CCET) é o Centro com o maior número de artigos publicados, e todos os Centros pertencem a mesma rede, não tendo nenhum Centro que não se comunica. O CCET e o Centro de

Ciências Biológicas e da Saúde (CCBS) tem ligação com todos os demais centros, mas a força de ligação entre eles é a mais forte da rede. Em suas pesquisas, Savić, Ivanović e Surla (2017) concluíram que os pesquisadores envolvidos em colaborações entre departamentos tendem a ser mais produtivos, colaborativos e institucionalmente importantes, ao contrário dos pesquisadores que colaboram apenas com colegas de seus próprios departamentos.

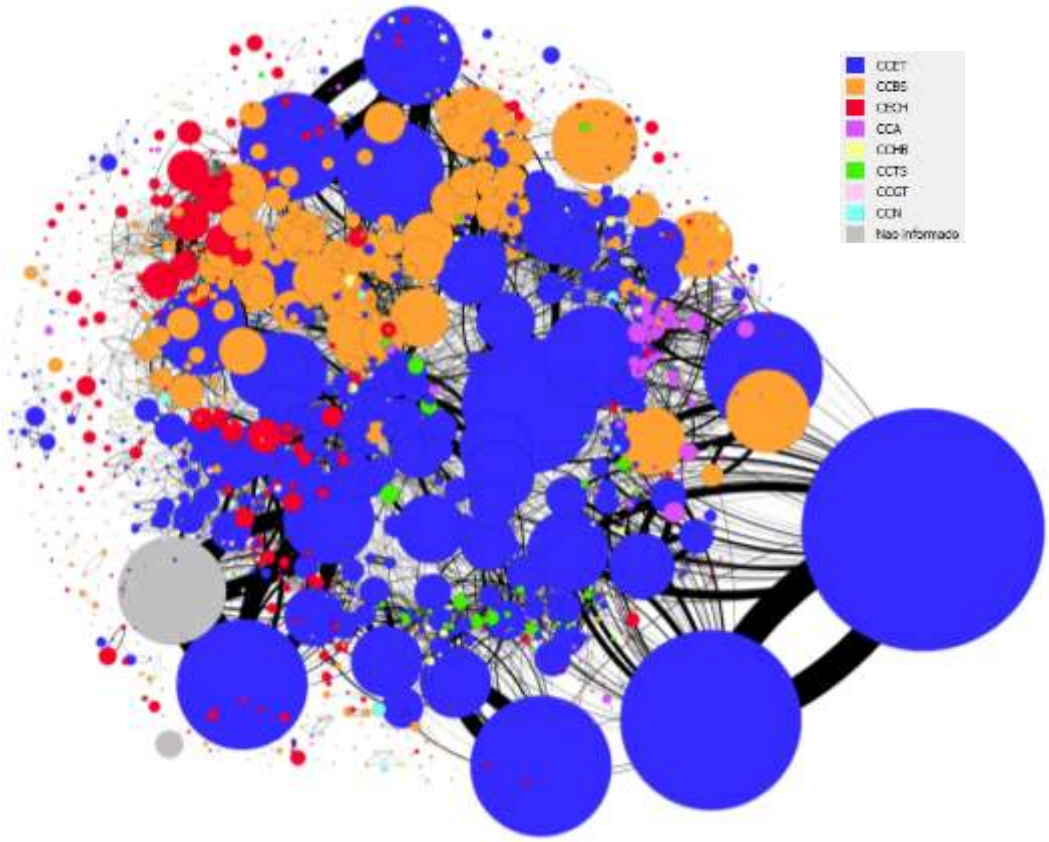
Figura 2 - Rede de colaboração científica entre os Centros da UFSCar destacando a quantidade de artigos publicados por cada Centro



Fonte: Dados da pesquisa a partir do software Gephi.

Uma maneira confiável de verificar a colaboração é através das chamadas redes de autores, uma vez que a coautoria é uma das formas mais concretas e documentadas de colaboração científica. Assim, a Figura 3 apresenta a rede de colaboração intraorganizacional, por coautoria, dos 1.575 docentes estudados.

Figura 3 - Rede de colaboração por coautoria dos docentes da UFSCar, com referência ao centro de vínculo e a quantidade de artigos publicados



Fonte: Dados da pesquisa a partir do software Gephi.

Nesta rede constam apenas os docentes vinculados à universidade, desconsiderando as ligações extraorganizacionais que os docentes possuem com outras instituições ou orientandos. As cores representam os Centros aos quais os docentes estão vinculados, o tamanho do nó é proporcional ao número de artigos publicados e a força de ligação dos nós é representada pelas arestas. Para Newman (2001), a maioria dos pesquisadores que obtêm altas forças de ligação são indivíduos bem conhecidos e bem conectados, e, consequentemente, com alto número de artigos escritos. A coautoria em grande quantidade de *papers* é uma boa maneira de se tornar bem conectado, colocar o pesquisador em contato com seus pares (NEWMAN, 2001). A distribuição utilizada foi a Fruchterman Reingold e posteriormente a distribuição "Não Sobrepor", para que fosse possível identificar os docentes.

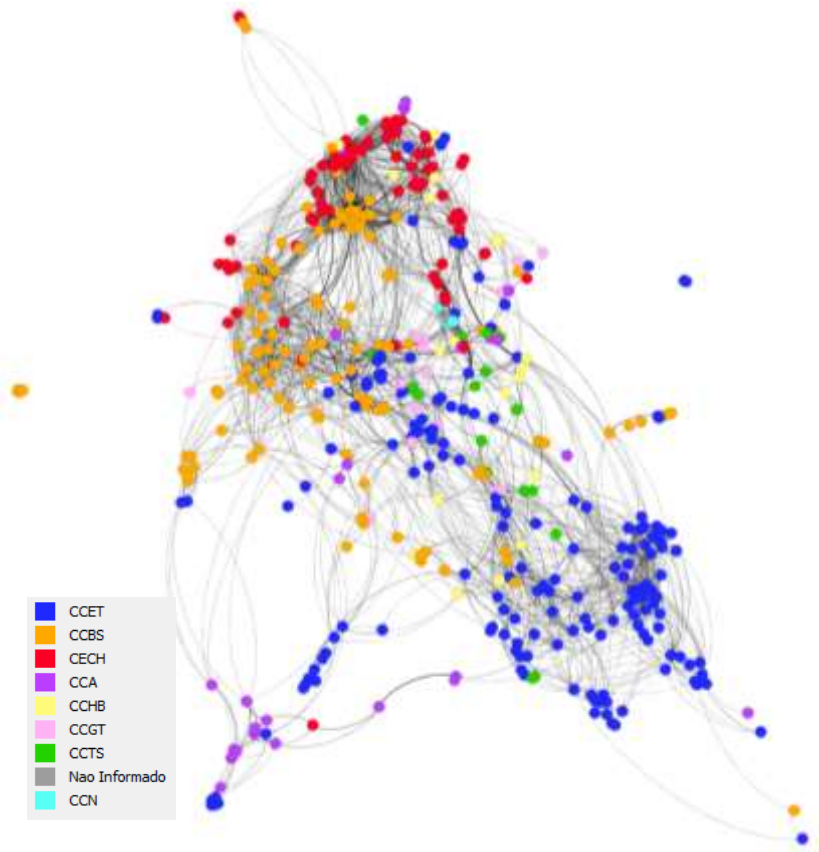
3.1 Análise das palavras-chave

É fato que a colaboração científica tem aumentado ao longo do tempo, assim como as medidas de colaboração e a investigação de seu impacto, utilizando-se principalmente da coautoria. Entretanto, Boyack (2009) coloca a seguinte pergunta: "Com quem eu deveria colaborar?". Sob o ponto de vista estratégico e do planejamento institucional, identificar as melhores oportunidades de colaboração é parte importante. Assim, para identificar futuras colaborações, é preciso apontar os pesquisadores que possam cooperar por trabalhar com os mesmos tópicos. Assim, foi elaborada a rede de temáticas, conforme Figura 4, onde foram excluídas as coautorias já existentes e mantiveram-se as relações por palavras-chave. O tamanho dos nós é igual, as cores estão relacionadas aos Centros de vínculo dos docentes e a distribuição foi a Force Atlas 2, que, de acordo com Sampaio e outros (2015), "[...] foca nos relacionamentos mais fortes e evidentes, demonstrando a coesão dos grupos. " (SAMPAIO et al., 2015, p. 85). Portanto, os nós mais próximos são os que têm maior afinidade de temas.

Nesta rede, é possível identificar os pesquisadores que estão trabalhando nos mesmos assuntos. Boyack (2009) define como "colaboração potencial" as colaborações em coautoria, porque, embora já exista uma colaboração, ela pode continuar no futuro e, também, as potenciais colaborações por aproximação temática, mas aqui demonstra-se apenas as colaborações que ainda não acontecem por coautoria. É possível, ainda, verificar que os Centros acabam ficando mais próximos, por abordarem temas em comum, mas existem diversos nós intercalados, o que propõe uma interdisciplinaridade. Podemos citar como exemplo destas possibilidades, os professores F.A.C., do Departamento de Medicina (DMed - CCBS), e L.P.A., do Departamento de Fisioterapia (DFisio - CCBS). Eles têm o maior peso de ligação desta rede, 0,9233, e trabalham com as temáticas: Alzheimer, envelhecimento, *aging*; entretanto, estes docentes nunca publicaram nenhum trabalho juntos, nem dentro do quadriênio analisado, nem em todo o período de vínculo com a universidade. Outro caso, que tem o segundo maior peso de ligação, igual a 0,9163, são os professores L.I.L.F., do Departamento de Ciência da Informação – Centro de Educação e Ciências Humanas (DCI – CECH), e S.E.A.C., do Departamento de Engenharia de

Produção (DEP – CCET). Ambos utilizam as temáticas: bibliometria, inteligência competitiva e indicadores de Ciência e Tecnologia, mas nunca tiveram nenhum trabalho em coautoria. Neste caso, esta possível colaboração também leva a interdisciplinaridade de departamentos e Centros.

Figura 4 - Rede de possíveis colaborações com acoplamento por palavras-chave entre os docentes da UFSCar, com referência ao Centro de vínculo de cada docente

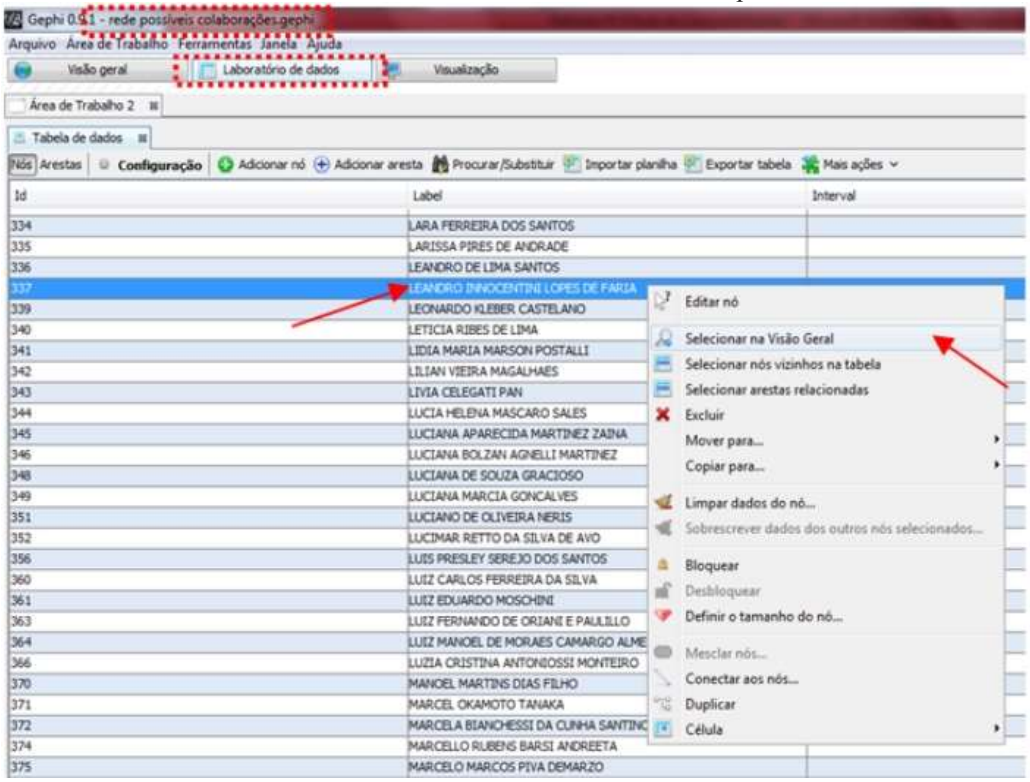


Fonte: Dados da pesquisa a partir do software Gephi.

Esta rede apresenta a situação da UFSCar como um todo, mas é possível selecionar um docente e verificar quais são seus potenciais colaboradores. Como exemplo, será utilizado o professor Leandro Innocentini Lopes de Faria, para encontrar a rede de "vizinhos" deste docente deve-se seguir o seguinte passo a passo: a princípio, dentro do arquivo "rede possíveis colaborações.gephi", no Gephi, deve-se localizar o professor, na aba "Laboratório de dados", clicar com o botão direito do mouse sobre o nome e selecionar o item "Selecionar na Visão Geral". Voltando para a aba "Visão Geral", o software dará um zoom no docente selecionado. Ao clicar sobre o professor, o Gephi mostra a rede de vizinhos dele,

esta rede pode ser colorida de outra cor para destacá-la entre as demais. Na sequência, a Figura 5 e a Figura 6 demonstram o passo a passo deste processo.

Figura 5 - Passo número 1 para selecionar a rede de possíveis colaboradores de um docente na tela de Laboratório de Dados no software Gephi



Fonte: Dados da pesquisa.

Figura 6 - Passo número 2 para selecionar a rede de possíveis colaboradores de um docente na tela Visão Geral do software Gephi

Fonte: Dados da pesquisa

O software permite que várias configurações sejam realizadas, desde a cor e tamanho dos nós, redes de vizinhos e também vizinhos dos vizinhos, depende do que se esta querendo analisar e visualizar. Deste modo, todos os docentes

podem ser identificados e ter sua rede de potenciais colaboradores selecionada. Este tipo de pesquisa pode ser tanto do interesse dos docentes, para identificar aqueles que trabalham com a mesma temática e buscar novas parcerias, já que existem campi que estão fisicamente longe, e mesmo dentro de um mesmo campus existem muitos docentes que não se conhecem e não sabem com quais assuntos cada um trabalha; como também pode ser do interesse de departamentos como a Secretaria Geral de Planejamento e Desenvolvimento Institucionais (SPDI), Pró Reitoria de Pesquisa (ProPq) ou Biblioteca Comunitária (Bco), que podem utilizar estes dados para incentivar e fomentar parceiras com alto grau de potencialidade.

Com esta rede, é possível analisar todos os docentes que podem ser possíveis colaboradores e fazer com que toda a universidade ganhe com isso, em termos de disseminação do conhecimento e compartilhamento de ferramentas, laboratórios e experiências. Ainda que nem todas estas potenciais colaborações sejam passíveis de acontecer na prática, por diversos fatores, devem existir sim, colaborações potenciais que não ocorrem atualmente, e podem ser despertadas através desta análise da rede.

Vale ressaltar, ainda, que essa pesquisa não tem a finalidade de investigar a vida acadêmica dos pesquisadores em si, mas sim, como eles podem contribuir para o fortalecimento do conhecimento dentro da Universidade e a formação de redes de colaboração com suas características próprias e determinantes.

4 Considerações finais

Atualmente, com o avanço tecnológico e as inúmeras maneiras de se obter informações, a colaboração surge como possibilidade de construir novos conhecimentos de forma coletiva. Além do mais, são exigidos elevados níveis de produtividade e reconhecimento dos pesquisadores, que se utilizam das relações de cooperação, compartilhamento e parcerias para o crescimento da ciência como um todo. Então, com essa demanda de produção e avaliação na área acadêmica, os cientistas precisam construir redes que propiciem o acúmulo de capital científico, a interdisciplinaridade, a redução dos custos de pesquisas e as trocas

de experiências entre os pesquisadores da mesma instituição e de instituições diferentes (BELLO, 2013).

O presente trabalho apresentou um procedimento para identificar futuras colaborações intraorganizacionais, tema pouco abordado nos trabalhos científicos. As colaborações potenciais são aquelas em que os autores trabalham com as mesmas palavras-chave lançadas na Plataforma Lattes e extraídas pela ferramenta *SyncLattes*, mas este processo pode ser expandido e aplicado para outros conjuntos de dados e instituições, como também pode ser analisado em níveis menores, como por departamento, centro ou temática.

Esta pesquisa mostrou que é possível elaborar e analisar redes sociais a partir dos dados extraídos da Plataforma Lattes, e que são dados consistentes e confiáveis. Demonstrou, ainda, que na instituição foco do estudo existe um grande potencial de colaboração que já é realizado, mas também, existem parcerias não exploradas. Para tal, foi elaborada uma maneira eficiente para identificar oportunidades de colaboração entre professores com assuntos específicos. E, assim, tentar trazer contribuições em termos de gestão, ou seja, os gestores devem estimular e aproximar estes docentes para que desenvolvam trabalhos em conjunto, pois a técnica de Análise de Redes Sociais permite criar redes que proporcionam informações sobre o relacionamento dos professores dentro da Universidade, sendo úteis aos gestores para a administração estratégica.

Além disso, é importante destacar que, do total de docentes, apenas 635 tinham ao menos uma palavra registrada neste campo na Plataforma Lattes. Corroborando os estudos de Gil-Leiva e Alonso-Arroyo, nos quais cerca de 45% dos docentes puderam ser recuperados. Isso mostra que, para trabalhos futuros, é possível adotar este procedimento em conjunto com outras técnicas, ou ainda adotar um sistema com vocabulário controlado, para que os docentes possam registrar as palavras-chave que mais representam sua pesquisa. Validando, assim, o método adotado, já que, com os professores que realizam o preenchimento, foi possível realizar a análise proposta.

Referências

BARROS, C.; GAMA, R.; FERNANDES, R. Conhecimento, redes e universidades: as redes de colaboração científica das Universidades de Lisboa, Porto e Coimbra. In: ASSOCIAÇÃO PORTUGUESA PARA O DESENVOLVIMENTO REGIONAL CONGRESS, 20., 2014, Évora. **Anais...** Évora: University of Évora, 2014. p. 1-20.

BASSOLI, M. **Avaliação do currículo lattes como fonte de informação para construção de indicadores**: o caso da UFSCar. 2017. Dissertação (Mestrado em Ciência, Tecnologia e Sociedade) - Universidade Federal de São Carlos, São Carlos, 2017.

BELLANCA, L. Measuring interdisciplinary research: analysis of co-authorship for research staff at the University of York. **Bioscience Horizons**, Oxford, v. 2, n. 2, p. 99-112, Apr. 2009.

BELLO, S. F. **Análise de redes de colaboração científica entre a Educação Especial e a Fonoaudiologia**. 2013. Tese (Doutorado em Educação Especial) - Universidade Federal de São Carlos, São Carlos, 2013.

BIRNHOLTZ, J. et al. Cross-campus collaboration: a scientometric and network case study of publication activity across two campuses of a single institution. **Journal of the American Society for Information Science and Technology,** New York, v. 64, n. 1, p. 162-172, Jan. 2013.

BORGATTI, S. P.; EVERETT, M. G.; JOHNSON, J. C. **Analyzing social networks**. Los Angeles: Sage, 2013.

BOYACK, K. W. Using detailed maps of science to identify potential collaborations. **Scientometrics**, Dordrecht, v. 79, n. 1, p. 27-44, Apr. 2009.

BRITO, A. G. C. de; QUONIAM, L.; MENA-CHALCO, J. P. Exploração da Plataforma Lattes por assunto: proposta de metodologia. **TransInformação,** Campinas, v. 28, n. 1, p. 77-86, jan./abr. 2016.

CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO. **Painel Lattes.** [S.l.]: CNPq, 2016.

DIAS, M. A. L. **Extração automática de palavras-chave na língua portuguesa aplicada a dissertações e teses da área das engenharias.** 2004. Dissertação (Mestrado em Engenharia Elétrica) - Faculdade de Engenharia Elétrica e de Computação, Universidade de Campinas, Campinas, 2004.

FARIA, L. I. L. de. **Prospecção tecnológica em materiais:** aumento da eficiência do tratamento bibliométrico. Aplicação na análise de tratamentos de superfície resistentes ao desgaste. 2001. Tese (Doutorado em Ciência e Engenharia de Materiais) - Universidade Federal de São Carlos, São Carlos, 2001.

GAZNI, A.; DIDEGAH, F. Investigating different types of research collaboration and citation impact: a case study of Harvard University's publications. **Scientometrics**, Dordrecht, v. 87, n. 2, p. 251-265, Jan. 2011.

GIL, A. C. **Como elaborar projetos de pesquisa.** 5. ed. São Paulo: Atlas, 2017.

GIL-LEIVA, I.; ALONSO-ARROYO, A. Keywords given by authors of scientific articles in database descriptors. **Journal of the American Society for Information Science and Technology**, New York, v. 58, n. 8, p. 1175-1187, 2007.

HAYASHI, M. C. P. I. et al. Um estudo bibliométrico da produção científica sobre a educação jesuítica no Brasil colonial. **Biblios**, Lima, v. 8, n. 27, p. 1-18, jan./mar. 2007.

KATZ, J. S.; MARTIN, B. R. What is research collaboration? **Research Policy,** Oxford, v. 26, n. 1, p. 1-18, 1997.

LIMA, R. A. de; VELHO, L. M. L. S.; FARIA, L. I. L. de. Indicadores bibliométricos de cooperação científica internacional em bioprospecção. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 12, n. 1, p. 50-64, jan./abr. 2007.

LU, K.; WOLFRAM, D. Measuring author research relatedness: a comparison of word-based, topic-based, and author cocitation approaches. **Journal of the American Society for Information Science and Technology**, New York, v. 63, n. 10, p. 1973-1986, Sept. 2012.

MATIAS, M. S. de O. **Base referencial para o povoamento de repositórios institucionais:** coleta automatizada de metadados da Plataforma Lattes. 2015. Dissertação (Mestrado em Gestão de Organizações e Sistemas Públicos) - Universidade Federal de São Carlos, São Carlos, 2015.

MELIN, G.; PERSSON, O. Studying research collaboration using co-authorships. **Scientometrics**, Dordrecht, v. 36, n. 3, p. 363-377, 1996.

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, Jan. 2001.

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

OKUBO, Y. **Bibliometric indicators and analysis of research systems:** methods and examples. Paris: OECD, 1997.

PEPE, A.; RODRIGUEZ, M. A. Collaboration in sensor network research: an

in-depth longitudinal analysis of assortative mixing patterns. **Scientometrics**, Dordrecht, v. 84, n. 3, p. 687-701, Sept. 2010.

SAMPAIO, R. B. et al. A colaboração científica na pesquisa sobre coautoria: um método baseado na análise de redes. **Perspectivas em Ciência da Informação**, Belo Horionte, v. 20, n. 4, p. 79-92, out./dez. 2015.

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

SEARCH TECHNOLOGY. **Vantage Point**. [S.l.]: Search Technology, [2001?].

UNIVERSIDADE FEDERAL DE SÃO CARLOS. **Apresentação**. [S.l.]: UFSCar, 2018. Portal da universidade.

VAN RAAN, A. F. J. Science meansuring. In: MOED, H. F.; GLÄNZEL, W.; SCHMOCH, U. (Ed.). **Handbook of quantitative science and technology research**: the use of publication and patent statistics in studies of S&T systems. New York: Kluwer Academic Publishers, 2004. p. 19-50.

VANZ, S. A. de S. **As redes de colaboração científica no Brasil.** 2009. Tese (Doutorado em Comunicação e Informação) - Faculdade de Biblioteconomia e Comunicação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2009.

WWF-Brasil. **Redes**: uma introdução às dinâmicas da conectividade e da auto-organização. Brasília: WWF-Brasil, 2003.

Intraorganizational scientific collaboration: analysis of networks by co-occurrence of keywords.

Abstract: Research on the impacts of collaboration on scientific research has been growing in recent years, mainly focused on the policy implications and allocation of financial resources. Scientific collaboration can encompass various levels of analysis, with emphasis on the internal - intrainstitutional, (intra) national - or in situating what is being analyzed in an external context - interinstitutional, international and others. However, the issue of intra-organizational collaboration and identification of possible collaborations have little approach and studies. There is no knowledge of studies that deal with these two themes in order to identify opportunities for internal collaboration in an institution. Thus, this paper presents an overview of the scientific collaboration between the professors of a federal university in the state of São Paulo, and presents a method to identify potential researchers for future collaborations. The identification of these potential collaborations is based on the coupling of authors

by subject matter through the keywords that the professors themselves fill in their Lattes Curricula. 635 professors and 2,678 words were identified. As a result, a keyword affiliation network where it is possible to identify researchers with high potential for collaboration is presented, which can help teachers identify their potential partners, as well as be used by strategic institutional bodies to foster collaboration between internal professors to the organization.

Keywords: Network Analysis. Scientific collaboration. Co-occurrence of words. Coupling.

Recebido: 24/01/2018
Aceito: 19/06/2018

CC BY

[1] CALLON, M.; COURTIAL, J.P.; PENAN, H. **La scientométrie**. Paris: Presses Universitaires de France, 1993. (Coleção Que sais-je?).

[2] Kessler, M. M. Bibliographic coupling between scientific papers. **American Documentation**, *v.* 14, n. 1, p. 10-25, 1963.

[3] Dados extraídos da base de Currículos Lattes em 29 de maio de 2018, com a última atualização datada de 30 de novembro de 2016, constando um total de 3.520.867 currículos cadastrados.