

ANÁLISE DE REDES POR COOCORRÊNCIA DE PALAVRAS-CHAVE: identificação de possíveis colaborações científicas

Nathalia Mendes Gerotti Franco

Universidade Federal de São Carlos, São Carlos, SP, Brasil
nath-jau@bol.com.br

Denilson de Oliveira Sarvo

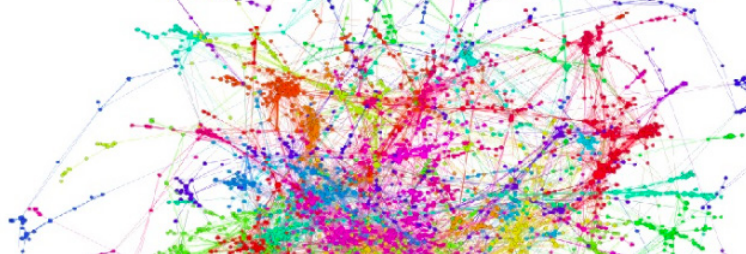
Universidade Federal de São Carlos, São Carlos, SP, Brasil
denilsonsarvo@gmail.com

Leandro Innocentini Lopes de Faria

Universidade Federal de São Carlos, São Carlos, SP, Brasil
leandro@ufscar.br

1 INTRODUÇÃO

Com as recentes mudanças na sociedade, o arranjo organizacional deixou de privilegiar a hierarquia e passou a se organizar em redes, que é um conjunto de pontos (ou nós) com atributos próprios que se ligam a outros pontos por meio de linhas ou *links*. Nas redes as espessuras desiguais dos traços mostram as diferenças do relacionamento entre os atores, ou seja, quanto mais espessa, maior é a qualidade desta relação, 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, 2003, p. 19). Quando se fala em Análise de Rede Social (ARS) focando a colaboração científica, analisa-se dentre diversos indicadores a rede de coautoria, que segundo Vanz (2009), forma-se quando dois ou mais pesquisadores publicam trabalhos juntos. Nesta rede podem-se identificar os nós como sendo os pesquisadores e a ligação entre eles representa ao menos, um artigo publicado em coautoria. O uso da análise de redes de coautoria tem se



mostrado como uma tendência no mundo contemporâneo para a análise do campo científico, uma vez que possibilita a identificação e compreensão de habilidades, interesses e demandas do campo científico (BELLO, 2013). Estas redes podem ser identificadas por meio de documentos coescritos. Mas, no caso da criação de redes voltadas para a identificação de possíveis colaborações, Boyack (2009) aponta a necessidade de realizar agrupamentos mais detalhados, pois é necessário ter uma maneira de identificar os pesquisadores que possam colaborar por trabalhar com os mesmos tópicos, definindo como “colaboração potencial”, aquelas por aproximação temática.

Uma possível maneira de criar redes visando identificar potenciais colaborações científicas é por meio das temáticas dos pesquisadores, identificadas segundo as palavras-chaves ou *keywords* informadas em suas publicações, que fornecem um breve resumo do conteúdo do documento publicado. Para Dias (2004, p. 35), as palavras-chaves “são particularmente úteis porque podem ser interpretadas individualmente e independentemente umas das outras”, porém é preciso ter cautela para realizar esse tipo de análise, uma vez que as palavras-chave normalmente são escolhidas de forma manual e a critério do próprio autor, assim nem sempre expressam corretamente os tópicos compreendidos pela publicação. Para este trabalho foram utilizadas as palavras-chaves que constam na Plataforma Lattes (PL) dos docentes da universidade foco do estudo. A PL é utilizada pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) com o objetivo de avaliar os pesquisadores e acompanhar os dados estatísticos da pesquisa científica brasileira, por meio dos currículos cadastrados, com acesso público. Conforme Corrêa et al. (2017) a criação da PL permitiu a padronização dos currículos acadêmicos no país, tornando-se estratégica na questão de planejamento, gestão e formulação de políticas públicas em Ciência e Tecnologia. Os autores ainda ressaltam a importância da PL quanto a recuperação de dados, pois há dificuldade em se analisar a produção científica de autores e instituições por meio das bases de dados de informação científica e tecnológica, uma vez que apresentam limitações quanto ao tipo

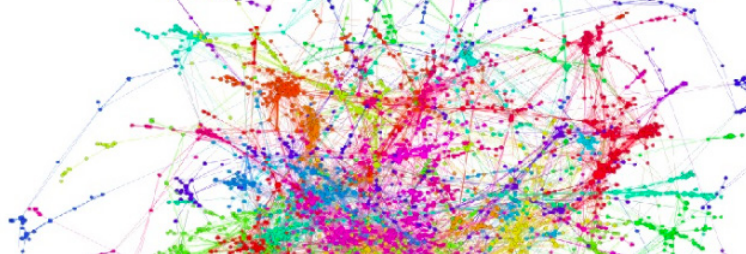


de documentos e parcela de publicações indexados. Deste modo, a PL torna-se uma fonte estratégica para analisar a produção científica brasileira, uma vez que conta com milhares de currículos “de todas as áreas do conhecimento, da maioria das instituições de ensino e pesquisa do país” (BASSOLI, 2017, p. 41). Como o pesquisador cadastra todas as suas atividades na Plataforma, é possível realizar uma análise mais completa de suas atividades científicas, além de possibilitar a compreensão de maneira ampla de todo o campo científico nacional.

Portanto, o presente trabalho tem como objetivo avaliar a colaboração científica dos docentes da Universidade foco do estudo, utilizando as métricas da análise de redes sociais e dados coletados junto a PL, com vista a dar suporte à gestão para fomentar possíveis colaborações científicas baseando-se na coocorrência de palavras-chave, tema pouco abordado nos trabalhos científicos.

2 DADOS E MÉTODOS

De acordo com Gil (2017, p. 25), a “classificação é uma característica da racionalidade humana”, pois facilita o entendimento e objetivos da pesquisa a ser desenvolvida. Esta é uma pesquisa descritiva, bibliográfica, quantitativa e aplicada com técnicas bibliométricas (análises estatísticas de publicações e atividades científicas). O objeto de pesquisa foi a Universidade Federal de São Carlos (UFSCar), fundada em 1968, instalada no interior do Estado de São Paulo, formada atualmente por quatro campi: São Carlos, Araras, Sorocaba e Lagoa do Sino. Como amostra foram utilizados dados extraídos da PL por meio da ferramenta *SyncLattes*, desenvolvida por Matias (2015), que trata-se de conjunto de *scripts* com a funcionalidade de extrair metadados da PL, tratar quanto a duplicidade e controle de autoria, sincronizando-os com o software *DSpace*. Para Bassoli (2017, p. 87), as funções apresentadas pelo *SyncLattes* tornam o processo ágil e consistente, tendo os dados extraídos por esta ferramenta um alto grau de confiabilidade, a autora ressalta que “[...] o Currículo Lattes viabiliza a leitura mais realista da trajetória científica da instituição e suas unidades organizacionais”. Justifica-se assim, a utilização dos



dados extraídos da PL para desenvolver este trabalho. Para esta pesquisa foram analisadas as produções científicas constantes nos currículos da PL de todos os docentes vinculados à UFSCar, desde sua fundação em 1968 até dezembro de 2016, num total de 1.598 currículos. A coleta de dados foi realizada em outubro de 2017, nos quais foram identificadas 27.999 palavras-chaves oriundas de um total de 27.107 artigos.

Para o tratamento e análise dos dados foi utilizado o *VantagePoint* (VP), ferramenta computacional de mineração e análise de dados, voltado para a construção de indicadores bibliométricos (SEARCH TECHNOLOGY INC., 2016). Para fazer as análises de rede foi utilizado o software UCINET, que possibilita a análise das métricas de relacionamentos, a criação e a manipulação de matrizes. Para a geração das redes gráficas foi utilizado o *Gephi*, software voltado para a exploração e manipulação de redes.

Para a construção das redes foram realizados recortes, visando reduzir a amostra para um conjunto de palavras passíveis de serem representadas e analisadas. O primeiro critério adotado foi o temporal, fazendo-se um recorte entre os anos de 2013 a 2016, que corresponde à última avaliação realizada pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), que é um sistema utilizado para classificar a produção científica dos programas de pós-graduação, em relação aos artigos publicados em periódicos científicos, que ocorre a cada três anos. Posteriormente, utilizando filtros no *VantagePoint*, foram selecionados apenas os artigos de docentes da UFSCar que possuíam ao menos uma relação de coautoria com outro docente vinculado a instituição. Após, foram selecionados apenas os artigos que possuíam ao menos uma palavra-chave, e, por fim, foi realizado o *List Cleanup* (recurso do VP que reúne palavras semelhantes em um único conjunto de registros), totalizando assim como amostra final 2.510 palavras, 635 docentes e 2.009 artigos.

Num primeiro momento foi elaborada uma rede com dois nós (autor x palavra-chave) no UCINET e, a partir desta matriz foi realizada a conversão para uma matriz de um nó, onde um par de docentes foi vinculado na medida em que compartilham afiliações, ou seja, citam a mesma palavra-chave, não indica, portanto, uma coautoria, mas sim,



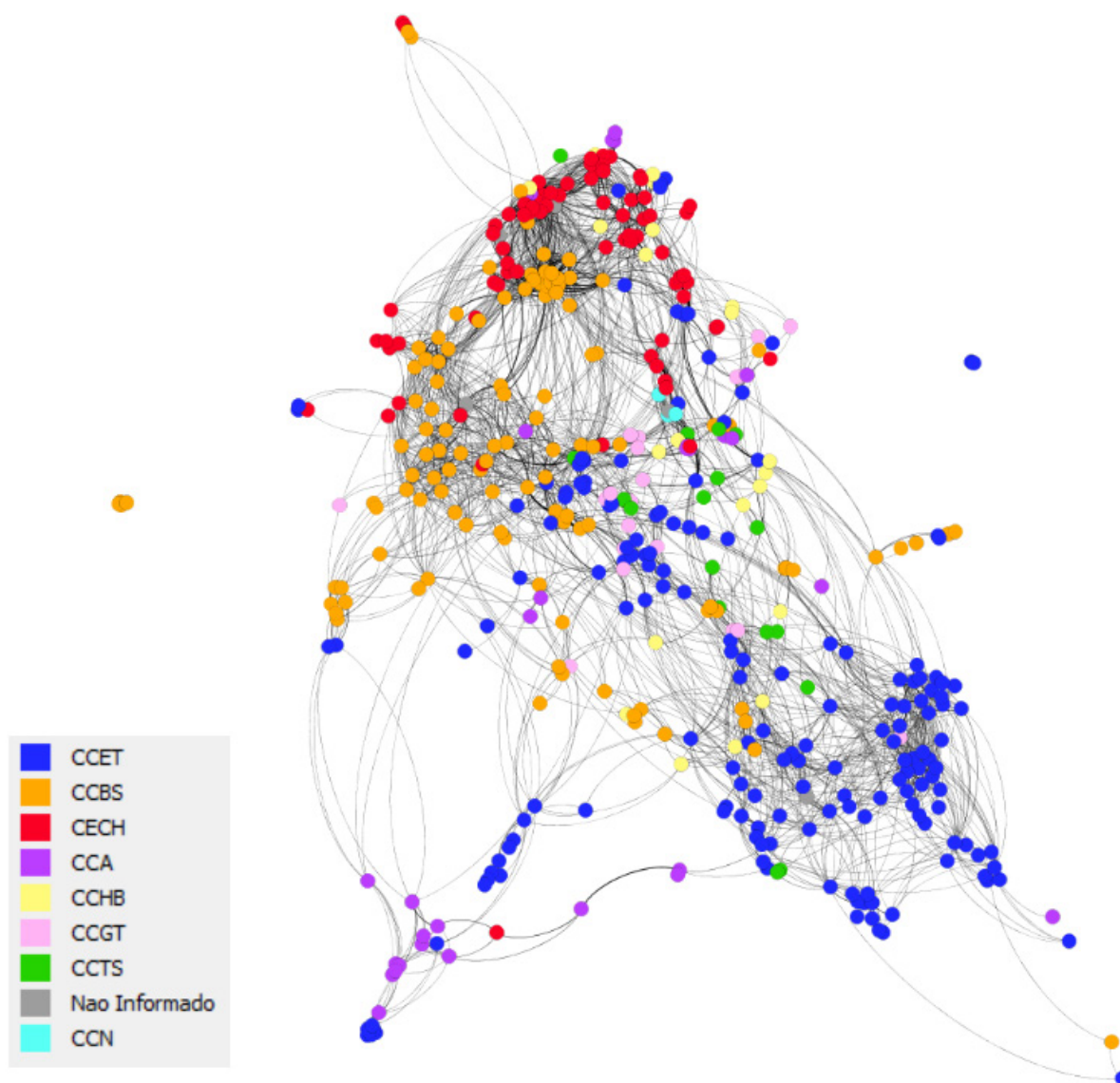
que ambos utilizaram a mesma palavra-chave em suas publicações. Borgatti, Everett e Johnson (2013) explicam como construir esta matriz de forma matemática: deve-se multiplicar a matriz de dois modos por sua transposição: , para cada par de linhas, olha-se para cada coluna e conta o número de vezes que ambos são iguais a 1. Quando um par de docentes tem uma palavra em comum conta-se 1, e quando nunca lançaram a mesma palavra na PL, eles não estão conectados, apresentando valores menores que zero. Quanto mais negativo, mais longe estão os temas entre os atores e quanto mais próximos a 1, maior a possibilidade de aproximação ou colaboração. Importando-se para o Microsoft Excel® as matrizes de afiliação por palavras-chaves e de coautoria, foi elaborada uma terceira matriz, onde apenas os docentes com afiliação por palavras-chave e sem coautoria foram apontados.

3 RESULTADOS E DISCUSSÃO

A partir da matriz gerada, tem-se o mapeamento das possíveis colaborações intraorganizacional, conforme a Figura 1. Na rede, o tamanho dos nós é igual, as cores estão relacionadas aos vínculos institucionais dos docentes por centros acadêmicos [sendo as siglas da legenda correspondente: Centro de Ciências Exatas e da Terra (CCET); Centro de Ciências Biológicas e da Saúde (CCBS); Centro de Educação e Ciências Humanas (CECH); Centro de Ciências Agrárias (CCA); Centro de Ciências Humanas e Biológicas (CCHB); Centro de Ciências em Gestão e Tecnologia (CCGT); Centro de Ciências e Tecnologias para a Sustentabilidade (CCTS); Centro de Ciências da Natureza (CCN)], que representam a unidade básica na estrutura organizacional da UFSCar e estão ligados com as áreas de conhecimento por ela pesquisadas. A distribuição foi a “Force Atlas 2”, que de acordo com Sampaio et al. (2015, p. 85) “[...] foca nos relacionamentos mais fortes e evidentes, demonstrando a coesão dos grupos”. Portanto, os nós mais próximos da Figura são os que têm maior afinidade de temas.

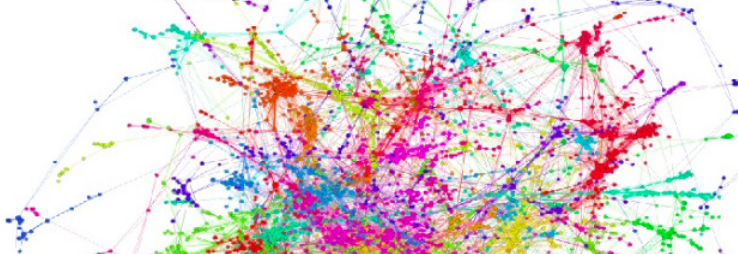


FIGURA 1 - REDE DE POSSÍVEIS COLABORAÇÕES ENTRE OS DOCENTES DA UFSCAR



Fonte: Elaborado pelos autores.

É possível verificar que os centros acadêmicos que trabalham com áreas similares acabam ficando mais próximos na rede, o que ocorre devido a abordarem temas em comum em suas publicações, mas existem diversos nós intercalados, o que propõe uma interdisciplinaridade conforme a representação por palavras-chaves em suas publicações. A partir daí podemos identificar possíveis relações de colaboração em potencial, como exemplo, os professores FAC (utilizando siglas para

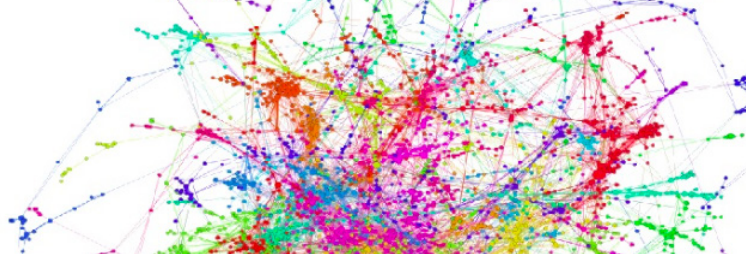


não identificação dos docentes), do Departamento de Medicina (DMed – CCBS), e LPA, do Departamento de Fisioterapia (DFisio – CCBS). Eles têm o maior peso de ligação desta rede, 0,9233, e trabalham com as temáticas: Alzheimer, envelhecimento, *aging*; entretanto, estes docentes nunca publicaram nenhum trabalho juntos, nem dentro do quadriênio analisado, nem em todo o período de vínculo com a universidade. Neste caso, esta possível colaboração demonstra que existem relações de colaboração em potencial a nível intraorganizacional ainda não explorado. Entretanto, não é por existir uma colaboração potencial, que ela deva ocorrer na prática, já que muitos outros fatores devem ser considerados como financiamentos, conjuntos de habilidades e relações pessoais (ou situações competitivas). Mas também existem muitas colaborações potenciais que podem não estar acontecendo.

A Figura I apresenta a situação da UFSCar como um todo, mas é possível selecionar cada docente e verificar quais são seus potenciais colaboradores. Por meio do *Gephi* é possível destacar os pesquisadores de interesse e identificar aqueles que trabalham com a mesma temática, possibilitando a identificação de novas parcerias em potencial, otimizando desse modo a formação de redes de colaboração intrainstitucional da universidade. Além dessa finalidade, análises de redes de colaboração e a identificação de potenciais colaborações segundo as áreas de interesse pesquisadas podem ser utilizadas por setores estratégicos da instituição, no âmbito do planejamento e desenvolvimento institucional, e setores de informação, que podem utilizar estes dados para incentivar e fomentar parcerias com alto grau de potencialidade.

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 (BELLO, 2013). Nesse caso, as colaborações potenciais identificadas são aquelas em que os autores trabalham com as mesmas palavras-chaves informadas na PL, mas esse processo pode ser expandido e aplicado para outros conjuntos de dados



e instituições, como também pode ser analisado em níveis mais específicos, como por departamentos, programas de pós-graduação ou temáticas. Vale ressaltar a potencialidade da análise de redes de coautoria para a gestão da instituição de pesquisa, em específico a identificação das análises de coautoria em potencial, uma vez que podem fomentar o incentivo para o aumento da colaboração intrainstitucional, tendo como consequência o fortalecimento do conhecimento dentro da Universidade e a formação de redes de colaboração com suas características próprias e determinantes, estratégicas para o avanço do campo científico.

REFERÊNCIAS

- BASSOLI, M. **Avaliação do Currículo Lattes como fonte de informação para construção de indicadores: o caso da UFSCAR**. 2017. 139 f. Dissertação (Mestrado em Ciência, Tecnologia e Sociedade) – Universidade Federal de São Carlos, São Carlos, 2017.
- BELLO, S. F. **Análise de redes de colaboração científica entre a educação especial e a fonoaudiologia**. 2013. 228 f. Tese (Doutorado em Educação Especial) – Universidade Federal de São Carlos, São Carlos, 2013.
- BORGATTI, S. P.; EVERETT, M. G.; JOHNSON, J. C. **Analyzing Social Networks**. SAGE: Los Angeles, 2013.
- BOYACK, K. W. Using detailed maps of science to identify potential collaborations. **Scientometrics**, v. 79, n. 1, p. 27-44, 2009.
- CORRÊA, T. S. et al. O fim do script Lattes? Uma análise de suas funcionalidades, alternativas para o presente e perspectivas para o futuro. **Revista do EDICC** (Encontro de Divulgação de Ciência e Cultura), v. 3, n. 3, p. 138-148, 2017.
- DIAS, M. A. L. **Extração automática de palavras-chave na língua portuguesa aplicada a dissertações e teses da área das engenharias**. 2004. 138 f. Dissertação (Mestrado em Engenharia Elétrica) – Faculdade de Engenharia Elétrica e de Computação, Universidade de Campinas, Campinas, 2004.
- GIL, A. C. **Como elaborar projetos de pesquisa**. 5. ed. São Paulo: Atlas, 2017.
- MATIAS, M. S. O. **Base referencial para o povoamento de repositórios institucionais: coleta automatizada de metadados da Plataforma Lattes**. 2015.



86 f. Dissertação (Mestrado em Gestão de Organizações e Sistemas Públicos) – Universidade Federal de São Carlos, 2015.

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

SEARCH TECHNOLOGY INC. **Vantage Point**. 2001. Disponível em: <<https://www.thevantagepoint.com/products/4-products/vantagepoint/I5-turn-information-into-knowledge.html>>. Acesso em: 08 jun. 2017.

VANZ, S. A. S. **As redes de colaboração científica no Brasil**. 2009. 204 f. Tese (Doutorado em Comunicação e Informação) – Universidade Federal do Rio Grande do Sul, Faculdade de Biblioteconomia e Comunicação, Porto Alegre, 2009.

WWF Brasil. **Uma introdução às dinâmicas da conectividade e da auto-organização**. Editora: Cidade, 2003.