

## Visualização de redes de coautoria como insumo biométrico às revistas científicas: uma proposta via *web scraping* para os periódicos *Em Questão* e *Encontros Bibli*

Co-authorship network visualization as bibliometric input for scientific journals: a *web scraping* proposal for the *Em Questão* and *Encontros Bibli* journals

Rafael Gutierrez Castanha

Universidade de Marília (UNIMAR), Departamento de Tecnologia da Informação,  
Marília, SP, Brasil  
r.castanha@gmail.com

<http://orcid.org/0000-0002-3117-1780> 

A lista completa com informações dos autores está no final do artigo 

### RESUMO

**Objetivo:** propor, a partir da raspagem de dados dos *web sites* dos periódicos, um meio de visualizar redes coautoriais entre os autores que publicaram artigos nestas revistas. De modo específico, gerar insumos para análises biométricas a partir do próprio portal eletrônico dos periódicos *Encontros Bibli* e *Em Questão*.

**Método:** implementa um código em linguagem de programação R para ler, identificar, extrair, organizar, estruturar e visualizar as redes de coautoria, aplicando-o nas revistas *Encontros Bibli* e *Em Questão*.

**Resultado:** demonstra o uso e a funcionalidade do código, assim como seu bom desempenho, ao gerar redes de coautoria e identificar os autores mais produtivos de cada uma das revistas analisadas. Sugere a inclusão das redes nos portais eletrônicos das revistas.

**Conclusões:** O método de *web scraping* pode ser utilizado pelas revistas para construir suas próprias redes de coautoria e fornecer mais informações passíveis de análises a pesquisadores. Além disso, pode ser utilizado por pesquisadores enquanto método de extração de dados de revistas científicas, sobretudo, aquelas construídas via *Open Journal System*, visto que periódicos que utilizam este sistema, possuem estruturas semelhantes. Ao propor e fornecer um código baseado em *web scraping*, este estudo releva o alto potencial biométrico desta técnica, contribuindo com novas perspectivas de técnicas computacionais aplicados à bibliometria.

**PALAVRAS-CHAVE:** Bibliometria. Coautoria. Periódicos Científicos. Produção Científica. *Web scraping*

### ABSTRACT

**Objective:** To propose, through *web scraping* of journal websites, a means to visualize co-authorship networks among authors who have published articles in these journals. Specifically, to generate inputs for bibliometric analyses directly from the electronic portal of the *Encontros Bibli* and *Em Questão* journals.

**Method:** Implement a code in the R programming language to read, identify, extract, organize, structure, and visualize co-authorship networks, applying it to the *Encontros Bibli* and *Em Questão* journals.

**Results:** Demonstrates the use and functionality of the code, as well as its strong performance, and suggests the inclusion of these networks on the journals' electronic portals. Concludes that *web scraping* can be used by both journals and researchers to create internal co-authorship networks and provide valuable information to researchers.

**Conclusions:** *Web scraping* can be employed by journals to construct their own co-authorship networks and provide additional analyzable information to researchers. Furthermore, it can be used by researchers as a data extraction method for scientific journals, especially those built using the *Open Journal System*, as journals using this system have similar structures. By proposing and providing a *web scraping*-based code, this study highlights the high bibliometric potential of this technique, contributing to new perspectives on computational techniques applied to bibliometrics.

**KEYWORDS:** Bibliometrics. Co-authorship. Scientific Journals. Scientific Production. *Web scraping*

## 1 INTRODUÇÃO

As redes de coautoria são frequentemente objeto de análises bibliométricas, pois oferecem uma maneira objetiva de visualizar as relações estabelecidas entre pesquisadores inseridos em contextos específicos por meio da redação conjunta de documentos científicos. Tradicionalmente, estas redes são analisadas para visualizar as relações de colaborações científicas entre pesquisadores. Embora a coautoria não seja sinônimo de colaboração, suas análises são amplamente utilizadas para identificar colaborações em diversos campos do conhecimento (Oliveira; Castanha; Grácio, 2021), visto que, a coautoria reflete um vínculo social entre autores, revelando redes sociais formadas pela produção científica (Lu; Wolfram, 2012).

Assim, a análise de redes de coautoria é um método importante para visualizar proximidades teórico-metodológicas em um campo específico (Bogado; Rosas; Grácio, 2022). Para representar visualmente essas redes e conexões, a análise de redes sociais e técnicas de mapeamento de dados são ferramentas úteis (Santin; Vanz; Caregnato, 2019). Embora a coautoria não seja perfeita quanto medida de colaboração, este indicador possui vantagens, por ser um método invariável e confiável, permitindo a reproduzibilidade dos resultados (Perianes-Rodriguez; Olmeda-Gomez; Moya-Anegón, 2010).

Dessa forma, o estudo das redes de coautoria se mostra fundamental para compreender as interações entre pesquisadores responsáveis pela produção científica de diferentes domínios científicos fornecendo formas para avaliar, mesmo que parcialmente, a dinâmica da colaboração científica no avanço do conhecimento. Entre os domínios passíveis de aplicação de análises de coautoria tem-se pequenos grupos de artigos, periódicos científicos, produções científicas de países, instituições ou até mesmo toda a produção de uma área conhecimento.

Especificamente com relação aos periódicos científicos, tais veículos podem fornecer valiosas informações a respeito da interação entre pesquisadores que publicam em periódicos específicos, considerando que a comunicação de resultados de pesquisas por meio de revistas acadêmicas é a forma mais utilizada por pesquisadores para disseminar os resultados de seus estudos (Urbizagastegui, 2016). Contudo, não se tem conhecimento de revistas que disponibilizem suas redes de coautoria em seus portais eletrônicos (*web sites*), ficando a cargo dos pesquisadores

extrair os dados de autoria de cada artigo e analisar as interações entre os pesquisadores que publicaram em determinadas revistas, e este procedimento nem sempre é feito de maneira automática, sendo quase sempre feito de maneira manual, dificultando tais análises.

Um procedimento que pode automatizar este tipo de extração é o método de *web scraping* (raspagem de dados da *web*) que, segundo Assis e Gomide (2021, p. 324), pode ser entendido “[...] como uma técnica de mineração de dados que pode ser realizada na internet. Essa técnica permite a estruturação de dados após a realização da busca, extração e manipulação dos dados desestruturados”. Porém, é de suma importância destacar que, anterior ao processo de *web scraping*, é necessário verificar se o *web site* a ser raspado permite que processos automatizados de coleta de dados e/ou robôs sejam implementados sobre sua plataforma. Desta maneira, por mais versátil que o método se apresente, alguns portais eletrônicos não autorizam que a raspagem seja realizada. Assim sendo, o método é aplicável a *web sites* diversos, independentemente de seu conteúdo, desde que não apresentem restrições para este método de mineração de dados.

Diante do exposto, esta pesquisa pretende responder ao seguinte problema: de que modo o método de raspagem de dados (*web scraping*) pode contribuir para visualização de redes de coautoria, enquanto insumo bibliométrico, em *web sites* de periódicos científicos? Para isso, é realizado uma análise a partir dos periódicos Encontros Bibli e Em Questão, da área de Ciência da Informação, com objetivo geral de propor, a partir da raspagem de dados dos *web sites* dos periódicos, um meio de visualizar redes coautorias entre os autores que publicaram artigos nestas revistas. De modo específico, gerar insumos para análises bibliométricas a partir do próprio portal eletrônico dos periódicos.

Dentre outras iniciativas e pesquisas que empregam o método de *web scraping* na área de Ciência da Informação brasileira, merece destaque o *ScraperCI* desenvolvido em Graciano (2022) e Graciano e Ramalho (2023). Ambos utilizam essa técnica com o objetivo de recuperar informações presentes em portais específicos da área, mais precisamente na Base de Dados em Ciência da Informação (BRAPCI). Vale ressaltar que essas pesquisas dão foco ao *web scraping* de maneira distinta da abordagem proposta neste estudo, uma vez que os autores concentram-se no

processo de recuperação da informação e esta pesquisa visa contribuir para a visualização de dados, especialmente no contexto bibliométrico das redes de coautoria.

Além disso, em relação às revistas construídas através do *Open Journal System* (OJS) da *Public Knowledge Project*, como é o caso dos periódicos mencionados anteriormente (Em Questão e Encontros Bibli), Ndungu (2020) destaca que o OJS dispõe de um conjunto de *plugins* rico e robusto conferindo interoperabilidade e permitindo a integração com diversos recursos e serviços que facilitam a descoberta de informações.

Nesse contexto, ao proporcionar uma abordagem inovadora na análise de informações bibliométricas provenientes das revistas, esta pesquisa contribui para o avanço de novas funcionalidades que podem ser implementados nos portais eletrônicos dos periódicos alinhando-se a proposta de Brito e Shintaku (2013), os quais sugerem a inclusão de um novo *plugin* no periódico Ciência da Informação do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT). Este *plugin* visa monitorar o processo editorial, quantificando os artigos submetidos, publicados e aprovados. Ainda, os autores discutem a utilização de funcionalidades adicionais, tais como a lista de artigos mais acessados e o mapeamento de termos de destaque nos artigos.

## 2 METODOLOGIA

Esta pesquisa pode ser classificada segundo sua abordagem, como qual-quantitativa, e de natureza aplicada, visto que parte da aplicação de métodos computacionais a um fenômeno social da ciência (coautorias). Desta forma, a fim de propor meios de visualização de redes de coautoria via *web scraping* para as revistas científicas Em Questão e Encontros Bibli, implementou-se, em linguagem R, um código capaz de raspar os dados destes periódicos. Previamente a implementação, foram verificados os termos de uso de ambas revistas a fim de apurar possíveis entraves à raspagem de dados e/ou coleta de dados em larga escala.

As duas revistas possuem termos de uso<sup>1</sup> e seus conteúdos estão licenciados sob a licença *Creative Commons 4.0 CC BY*, de modo que qualquer usuário tem direito de compartilhar (copiar, baixar, imprimir ou redistribuir o material em qualquer suporte ou formato) e adaptar (remixar, transformar e criar a partir do material para qualquer fim, mesmo que comercial). Além disso, foi realizada a verificação utilizando “/terms” e “/robots.txt” após as *urls* dos *web sites* das revistas para examinar se haviam conteúdos que as revistas não permitissem que fossem coletados, e ambas não acusaram nenhum tipo de problema.

Os periódicos *Em Questão* e *Encontros Bibli* foram selecionadas para esta pesquisa não só pela sua relevância da área, mas por possuírem *web sites* construídos (arquiteturas similares) via *Open Journal System (OJS) - Public Knowledge Project*, o que facilitou a codificação de raspagem de dados e a visualização das redes de coautoria. A revista *Em Questão* é vinculada ao programa de Pós-Graduação em Ciência da Informação da Faculdade de Biblioteconomia e Comunicação da Universidade Federal do Rio Grande do Sul, enquanto o periódico *Encontros Bibli* pertence ao Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Santa Catarina, e ambas estão classificadas como A2 no Qualis CAPES (referência 2017-2020), sendo duas das mais expressivas das área.

Em processos de raspagem de dados, comumente são utilizadas bibliotecas específicas das linguagens de programação como as famigeradas *selenium* e *beautifulsoup* da linguagem *Python*, ou a biblioteca *rvest*, utilizada pela linguagem R, e consequentemente, nesta pesquisa. Em suma, o processo de *web scraping* se deu a partir da: i) leitura da *url* do site pela linguagem de programação; ii) identificação da classe do css (*cascading style sheets*) dentro dos *web sites*; iii) extração dos elementos identificados anteriormente; iv) organização, estruturação e visualização dos dados.

Dessa forma, para leitura das *urls* dos sites, utilizou-se a função *read\_html* da biblioteca *rvest*. Para identificação das classes a serem usadas, foi utilizada a função “inspecionar” do navegador Google Chrome. Este recurso, permite explorar todos elementos de um determinado *web site* e favorece a identificação da classe css a

---

<sup>1</sup> Encontros bibli: [https://periodicos.ufsc.br/index.php/eb/politica\\_editorial](https://periodicos.ufsc.br/index.php/eb/politica_editorial);  
Em Questão: <https://seer.ufrcs.br/index.php/EmQuestao/about>

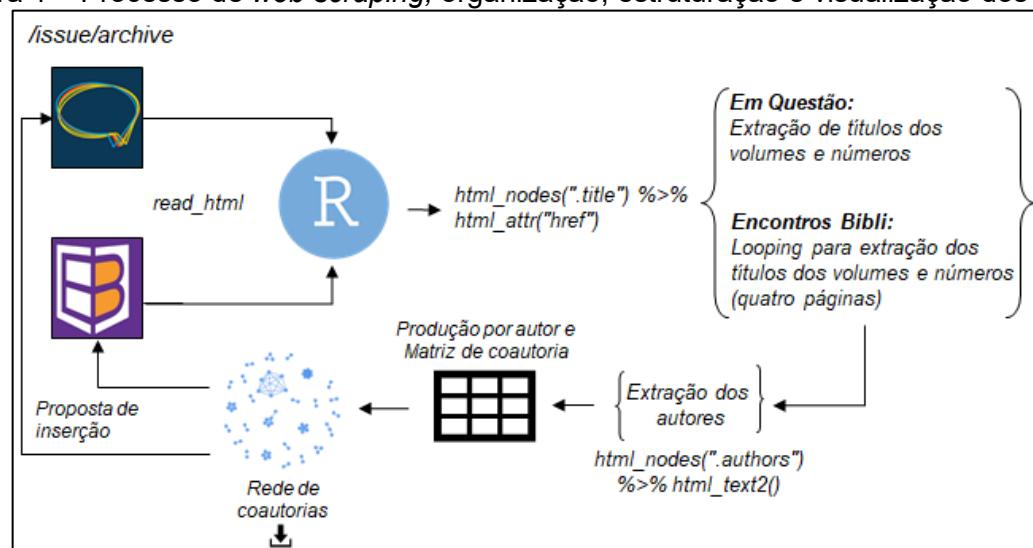
serem utilizadas. Para ambas revistas, inspecionou-se as páginas de “arquivo” (/issue/archive) por conterem os links de todos os títulos de todos volumes e números publicados pelos periódicos.

No *web site* da Em Questão, todos volumes e números estão dispostos em uma só página, totalizando 59 links de volumes e números, enquanto na Encontros Bibli, os volumes e números, estão distribuídas (atualmente) em quatro páginas. Assim, o código teve que ser capaz de perpassar pelas quatro páginas, encontrando todos os 77 links (disponibilizados até a data dos testes realizados). Este processo garante que, para os futuros artigos, números ou volumes a serem publicados pelas revistas, o código seja capaz de atualizar as redes e dados de coautoria raspados dos periódicos. A classe do css referente aos títulos dos volumes e números, em ambas revistas, foi o *.title* e as funções responsáveis por estas identificações foram *html\_nodes(".title") %>% html\_attr("href")*.

Após a identificação dos títulos dos volumes e números e dos links a serem acessados pelo código, foi identificada a classe do css dos autores de cada artigo, sendo esta a *.authors* nas duas revistas. Tendo estas informações, o código extraiu os dados de autoria de cada artigo pela combinação das funções *html\_nodes(".authors") %>% html\_text2()*. Tendo os dados de autoria de todos artigos, tais dados foram organizados de modo a gerar duas visualizações : i) a produção de artigos dos autores mais produtivos; ii) a rede de coautoria.

A produção de artigos por autor pode ser considerada uma etapa intermediária para construção da rede de coautoria, visto que a construção da rede depende da identificação de todos os autores e de suas publicações. A seguir, para construção da rede de coautoria, todos autores foram organizados em forma de matriz (adjacência) de coautorias e posteriormente esta matriz foi convertida para rede de coautoria. Para esta conversão foi utilizada a biblioteca *igraph* por meio da função *graph\_from\_data\_frame*, e para visualização da rede foi utilizada biblioteca *visNetwork* pela sua versatilidade. A biblioteca (*visNetwork*) dá opção ao usuário de interagir com a rede selecionando vértices específicos, aplicando *zoom*, além da possibilidade de arrastar os vértices pela tela de visualização. Todo processo descrito está presente na Figura 1.

Figura 1 – Processo de *web scraping*, organização, estruturação e visualização dos dados



Fonte: Elaboração do autor (2023).

A Figura 1 demonstra as quatro etapas descritas anteriormente, de modo que, a partir da página de arquivos de ambas as revistas, estabeleceu-se todo procedimento para obtenção da rede de coautoria mediante a raspagem de dados. Após a construção da rede de coautoria, propõe-se a inserção desta rede nos *web sites* das revistas, uma vez que a biblioteca *visNetwork* fornece a opção de salvar a rede em formato *html*, favorecendo a inserção nos respectivos *web sites*, tal como alguns *plugins* oferecidos pelo OJS e utilizados pelas revistas: a nuvem de palavras-chave e a lista de artigos mais lidos.

E mais, ao salvar a rede desta maneira, a visualização permanece interativa e em alta resolução de imagem. Enquanto elementos interativos, destacam-se a seleção de vértices específicos que é dada a partir de uma lista suspensa (*select by id*) contendo todos autores identificados em que, ao selecionar determinado pesquisador, são ressaltadas apenas as conexões que este autor estabelece, isto é, evidencia a sub-rede de coautoria ao qual ele pertence. Assim, a rede de coautoria pode complementar os *plugins* e fornecer um novo insumo bibliométrico às revistas. Após a execução do código, é possível salvar a rede gerada no diretório de trabalho do R. É necessário elucidar que todo este processo pode ser aplicado para todos artigos publicados pela revista ou simplesmente para um único volume ou número específico.

Para facilitar o uso, optou-se por tornar o código interativo a partir de três interações representadas por três perguntas a serem respondidas pelo usuário: 1) "Deseja analisar a revista Em Questão (1) ou Encontros Bibl (2) ?:;" 2) "Caso deseje analisar um número específico cole o link aqui, caso contrário digite 1":; 3) "Digite o número mínimo (maior que zero) de frequência de coautoria entre dois autores:". A Figura 2a apresenta a execução do código aplicado a revista Encontros Bibl em um número específico, representado pelo link inserido pelo usuário.

Neste caso, foi inserido o valor 1 para o número mínimo da frequência de coautoria entre dois autores. Isto é, a rede de autorias a ser visualizada será composta por todos autores que possuem pelo menos um artigo em coautoria e também por aqueles que redigiram artigos de maneira individual.

Figura 2 – Interação do código com o usuário: a) execução do código para um número específico; b) execução do código para toda revista

	<p>Console Background Jobs ×</p> <p>R 4.1.2</p> <p>Deseja analisar a revista Em Questão (1) ou Encontros Bibl (2) ?:2</p> <p>Caso queira analisar um número específico cole o link aqui, caso contrário digite 1:</p> <p><a href="https://periodicos.ufsc.br/index.php/eb/issue/view/3524">https://periodicos.ufsc.br/index.php/eb/issue/view/3524</a></p> <p>Digité o número mínimo (maior que zero) de frequência de coautoria entre dois autores:1</p>
b)	<p>Console Background Jobs ×</p> <p>R 4.1.2</p> <p>Deseja analisar a revista Em Questão (1) ou Encontros Bibl (2) ?:1</p> <p>Caso queira analisar um número específico cole o link aqui, caso contrário digite 1:1</p> <p>Digité o número mínimo (maior que zero) de frequência de coautoria entre dois autores:2</p>

Fonte: Elaboração do autor via RStudio (2023).

Já na Figura 2b, é demonstrada a execução do código aplicado a todas publicações do periódico Em Questão. Neste exemplo, tem-se o caso em que o usuário inseriria o valor 2 para frequência mínima de coautoria entre dois autores, ou seja, a rede de coautoria gerada pelo código será composta somente pelos autores que possuem pelo menos duas publicações em coautoria com outros autores.

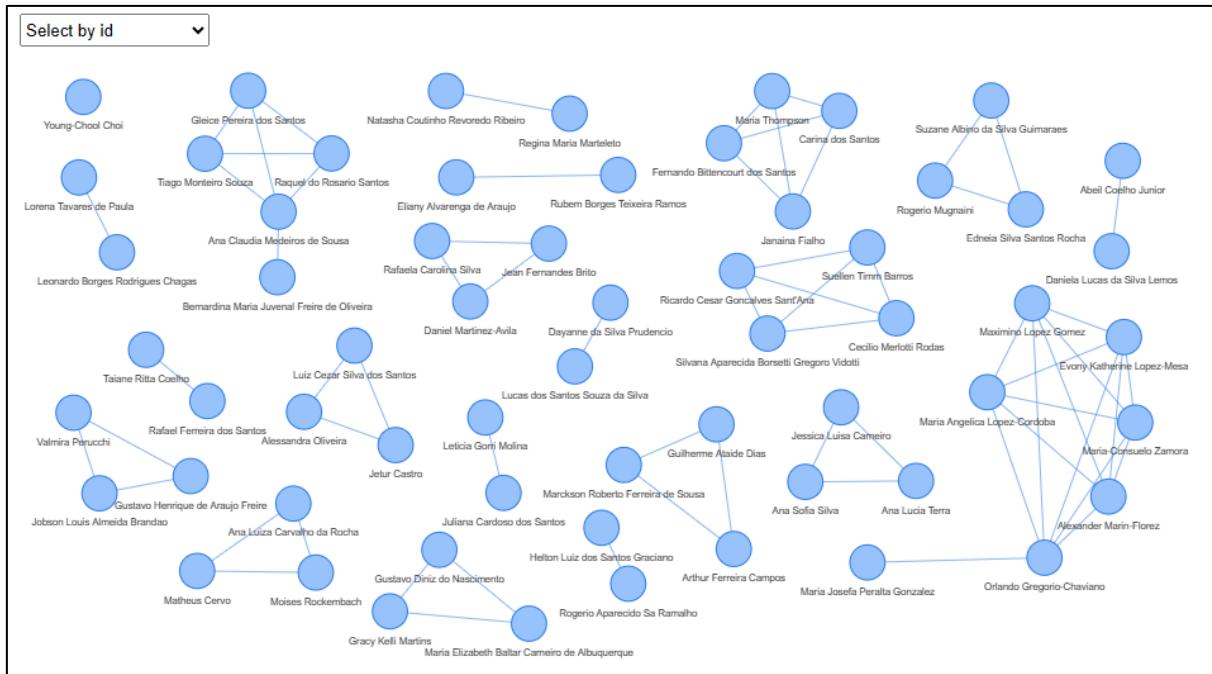
A motivação da criação desta interação (frequências mínimas de coautoria) se dá justamente para que o usuário possa explorar a rede de autorias em vários níveis, buscando as ligações (coautorias) mais intensas presentes nas revistas analisadas.

Pelo caráter propositivo desta pesquisa, os resultados serão compostos por demonstrações do uso do código. Serão apresentadas as redes de coautoria de um volume em específico para cada uma das revistas (Figuras 3 e 4), e para isso, serão utilizados os últimos volumes de cada revista: volumes 28 (v.28 - Inovação, Tecnologia e Sustentabilidade) e 29 (v.29), respectivamente das revistas Encontros Bibli e Em Questão. A seguir, serão apresentadas as redes de coautorias de ambas revistas contendo todos autores com frequência mínima de coautoria entre autores igual a 1 (Figuras 5a e 5b) e igual a 2 (Figuras 6a e 6b). Após isso, será apresentada a proposta de inserção das redes nos *web sites* das revistas (Figuras 7 e 8). O código implementado e todas as redes foram depositadas no repositório Zenodo para melhor visualização e está disponível em Castanha (2023). Para utilizar o código, basta copiar e executá-lo no R (ou RStudio). Foi utilizado o software *RStudio* em sua versão 4.1.2, para implementação e testes do código, em um computador com sistema operacional *Windows 10*, processador *Intel Core(TM) i7-8550U CPU @ 1.80GHz-1.99 GHz*, SSD *200Gb* e *8Gb* de memória *RAM*. Todos processamentos e testes foram realizados em 22 de junho de 2023.

### 3 RESULTADOS

Implementado o código, para demonstrar seu uso, tem-se as Figuras 3 e 4. Nestas figuras, são apresentadas a rede de coautoria entre os autores que publicaram artigos nos últimos volumes, v.28 e v.29, das revistas Encontros Bibli e Em Questão, respectivamente.

Figura 3 – Rede coautorias do volume 28 (Inovação, Tecnologia e Sustentabilidade) do periódico Encontros Bibli



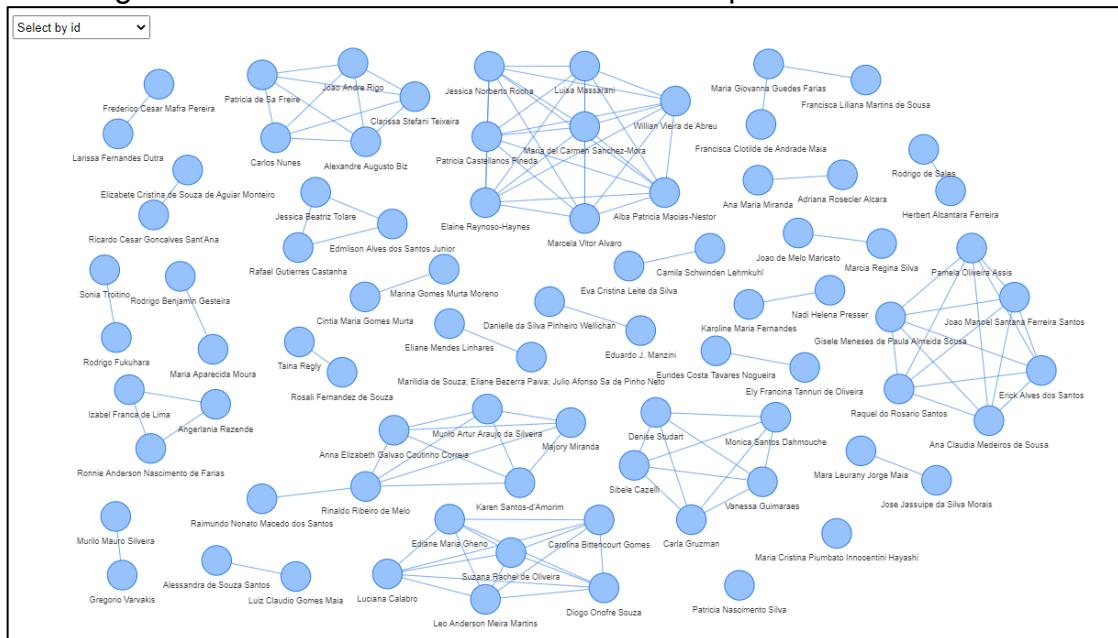
Fonte: Elaboração do autor via RStudio (2023).

A Figura 3 é constituída por 61 autores e o código demonstrou eficiência ao identificar com precisão todos os nomes dos autores sem apresentar qualquer tipo de inconsistência, de modo que, dois autores publicaram mais de um artigo (dois artigos cada), saber: Ana Claudia Medeiros de Sousa e Orlando Gregorio-Chaviano. Além disso, o código foi capaz de identificar o autor Young-Chool Choi, o único a publicar individualmente neste volume da revista. Essa validação foi realizada por meio de uma minuciosa conferência manual, realizada na data da coleta de dados, comparando os nomes extraídos pelo código com os nomes dos autores disponíveis no site correspondente a este volume.

No que diz respeito à Em Questão, a rede de coautoria entre os autores que publicaram em seu volume 29 é apresentada na Figura 4. Ao construir a rede de coautoria entre os pesquisadores que publicaram no 29º volume da Em Questão, o código foi capaz de identificar 81 dos 83 autores. Esta inconsistência entre o valor real e o valor extraído se deu pela não padronização dos metadados de autoria de um artigo em específico, em que as autoras são Eliane Mendes Linhares, Marilidia de Souza, Eliane Bezerra Paiva e Júlio Afonso Sá de Pinho Neto. Por padrão, a revista separa os nomes dos autores por vírgula e neste caso específico, a partir da segunda autora, foi utilizado ponto e vírgula, de tal forma que o web site apresenta esta coautoria da seguinte forma: *Eliane Mendes Linhares, Marilidia de Souza; Eliane*

Bezerra Paiva; Júlio Afonso Sá de Pinho Neto. Assim sendo, os nomes *Marilidia de Souza*; *Eliane Bezerra Paiva*; *Julio Afonso Sa de Pinho Neto* foram considerados como um autor.

Figura 4 – Rede coautoriais do volume 29 do periódico Em Questão



Fonte: Elaboração do autor via RStudio (2023).

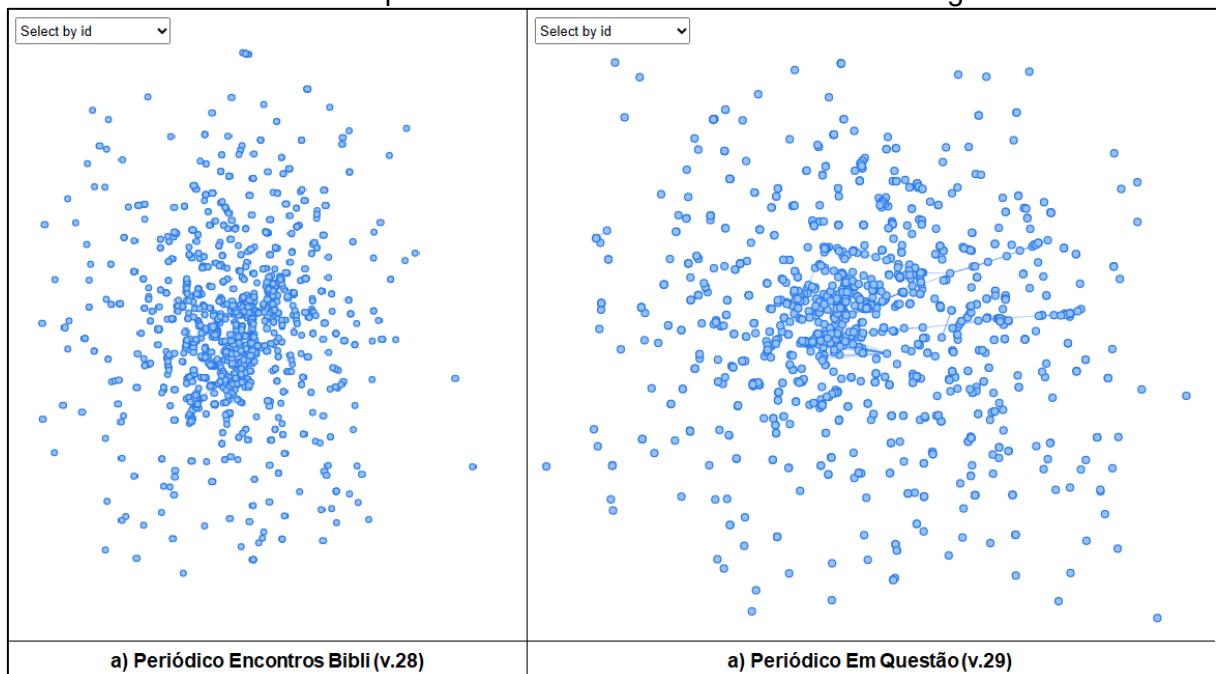
Ademais, após testes preliminares, optou-se por não utilizar a acentuação nos nomes dos autores, visto que, em algumas publicações foram encontrados nomes de autores digitados com e sem acento, prejudicando a visualização das redes e até mesmo o mecanismo de busca dos próprios *web sites* das revistas. A remoção dos acentos favorece a não duplidade de autores, como por exemplo: “Sonia Elisa Caregnato” e “Sônia Elisa Caregnato”, ou, “Samile Andréa de Souza Vanz” e “Samile Andrea de Souza Vanz”.

Similar a Encontros Bibli, o código identificou dois pesquisadores que mais publicaram neste volume da Em Questão (dois artigos cada), sendo eles: Maria Giovanna Guedes Farias e Rinaldo Ribeiro de Melo. Ademais, na rede presente na Figura 4 nota-se dois pesquisadores que não se ligaram com os demais, isto é, que não publicaram artigos em coautoria: Patrícia Nascimento Silva e Maria Cristina Piumbato Innocentini Hayashi.

Tendo a visualização de volumes específicos, apresenta-se na Figura 5, as redes de coautoria de todos os volumes e números de ambas as revistas. Ou seja,

em cada uma das redes, tem-se todos autores que publicaram ao menos uma vez em cada periódico. Foram identificados 1365 autores distintos que publicaram artigos no periódico *Encontros Bibli* e 1315 na *Em Questão*.

Figura 5 - redes de coautoria de ambas revistas contendo todos autores para ambas revistas com frequência mínima de coautoria entre autores igual a 1



Fonte: Elaboração do autor via RStudio (2023).

Neste caso, sugere-se uso do seletor de vértices (*select by id*) para que autores de interesse, e suas respectivas redes de autoria que representam sub-redes de toda a revista, sejam melhor exploradas. Tendo toda rede de coautoria das revistas, os autores mais produtivos estão dispostos na Tabela 1. O código fornece duas opções de análise de produtividade de pesquisadores: i) uma tabela contendo todos os autores e suas respectivas produções; ii) uma tabela contendo somente os dez autores mais produtivos (exemplificado na Tabela 1). Para obter estes resultados, basta o usuário, após rodar o código, executar as funções *prod\_total* ou *prod\_top10*, respectivamente.

A Tabela 1 identifica as pesquisadoras Sônia Elisa Caregnato e Leilah Santiago Bufrem em ambas revistas enquanto autoras mais produtivas. Ainda, é notório que alguns autores se apresenta como mais produtivos não só por produzirem muitos artigos mas também por exercerem, em algum momento, funções no corpo editorial

das revistas, é o caso de Francisco das C. de Souza, Adilson Luiz Pinto, Enrique Muriel-Torrado e Rafaela Paula Schmitz na revista Encontros Bibli e de Samile Andréa Vanz na Em Questão.

Tabela 1 – Autores mais produtivos de ambas revistas

<b>Encontros Bibli</b>	<b>Documentos</b>	<b>Em Questão</b>	<b>Documentos</b>
Francisco das C. de Souza	34	Samile Andrea de Souza Vanz	23
Adilson Luiz Pinto	26	Sonia Elisa Caregnato	12
Enrique Muriel-Torrado	21	Maria Claudia Cabrini Gracio	11
Rafaela Paula Schmitz	21	Jose Eduardo Santarem Segundo	8
Marcio Matias	17	Luciano Antonio Digiampietri	8
Leilah Santiago Bufrem	14	Valdir Jose Morigi	8
Ligia Maria Arruda Café	11	Elizete Vieira Vitorino	7
Guilherme Ataide Dias	10	Kelley Cristine G. Dias Gasque	7
Sonia Elisa Caregnato	10	Leandro Innocentini L. de Faria	7
Gustavo H. de A. Freire	9	Leilah Santiago Bufrem	7

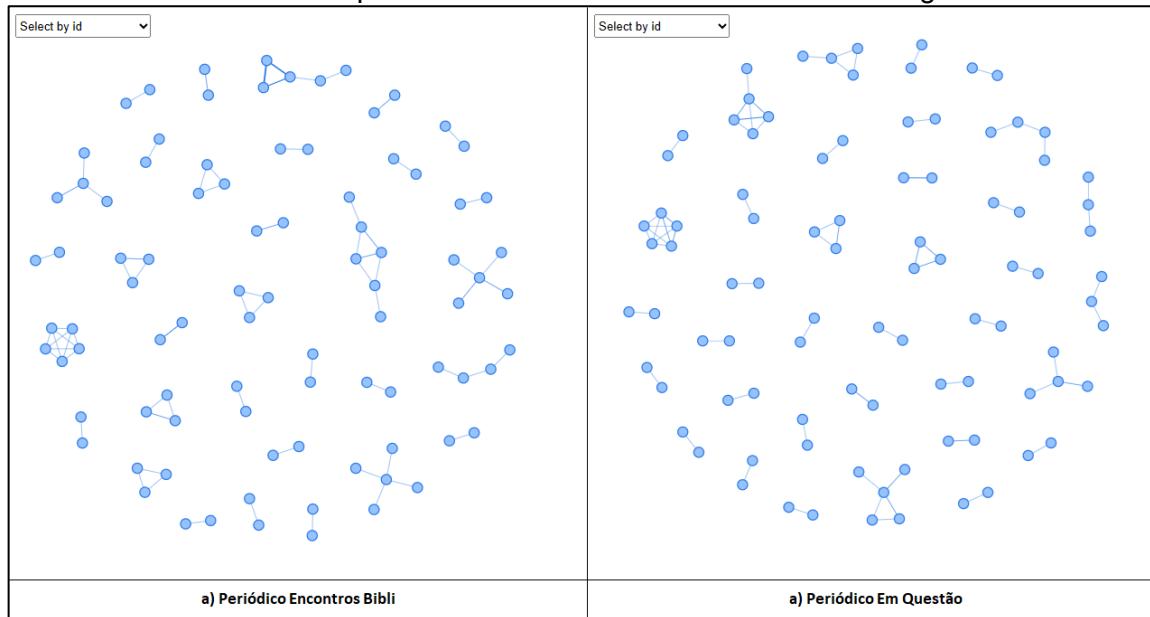
Fonte: Elaboração do autor (2023).

Cabe elucidar que o código ao extrair a classe identificada como “.authors” também irá extrair as informações de autoria dos editoriais publicados pelos periódicos, que por sua vez são assinados pelos editores da revista, incrementando o total de produção dos pesquisadores que também são editores. Assim sendo, no âmbito das análises bibliométricas, este resultado deve ser devidamente relativizado para que os documentos de cunho institucional não inflem a produção destes autores.

Como mencionado anteriormente, a identificação da produção de cada autor apresenta-se como etapa intermediária a criação das redes de coautoria e o fato de os editores apresentarem-se como altamente produtivos (perante toda a revista) pouco interfere nas redes de coautoria da revista Em Questão, uma vez que, não foram identificados editoriais em coautoria. E mais, várias publicações de cunho institucional da revista, como “expedientes” ou “normas de publicação” utilizam o autor institucional “Comissão Editorial Em Questão”. Se esclarece que este autor institucional também é identificado pela classe “.authors”, porém, foi excluído da análise, a fim de manter somente pesquisadores. Em contraponto, a revista Encontros Bibli possui diversos editoriais em coautoria, influenciado de alguma forma as redes de coautoria. Em algumas situações, a revista se utiliza também de um autor institucional denominado “Editor Encontros Bibli” que também foi excluído das análises. A fim de refinar as redes apresentadas nas Figuras 5a e 5b, executou-se o

código com o parâmetro de frequência mínima de coautoria entre dois autores igual a 2. Assim, gerou-se as redes, na Figura 6, das duas revistas contendo somente pesquisadores com ao menos dois artigos em coautoria entre si.

Figura 6 - redes de coautorias de ambas revistas contendo todos autores para ambas revistas com frequência mínima de coautoria entre autores igual a 2



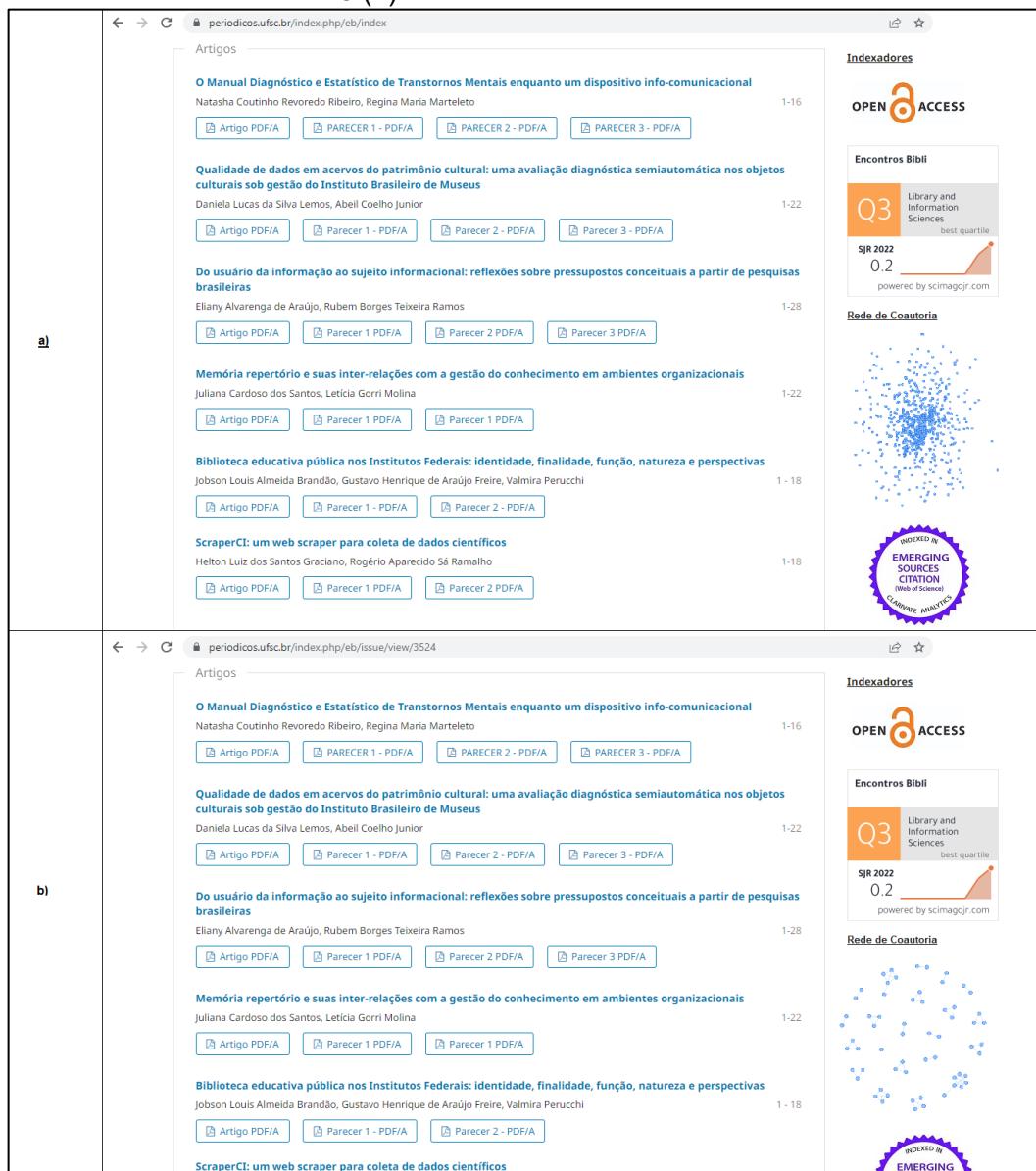
Fonte: Elaboração do autor via RStudio (2023).

A Figura 6 demonstra um refinamento opcional ao usuário de modo que o mesmo possa se aprofundar nas análises de coautorias observando os autores que estabeleceram ligações mais intensas entre si. Este refinamento encontrou 89 autores distintos do periódico Encontros Bibli que possuem ao menos dois artigos em coautoria com outros pesquisadores, enquanto que na revista Em Questão, foram identificados 91 autores distintos com dois ou mais artigos redigidos em coautorias com outros. Para acessar este conjunto de pesquisadores, basta o usuário executar a função *prod\_2*. Esta função retorna a lista de todos pesquisadores a estabelecerem a frequência mínima de coautoria informada juntamente à quantidade de coautores de cada um. Caso seja estabelecida uma frequência de coautorias maior do que as existentes, o sistema retornará erro e não executará a rede de coautoria.

Demonstrado o uso do código, é proposto a inserção das redes de coautoria nos respectivos *web sites* das revistas analisadas. A proposta de inserção das redes de coautoria nas revistas Encontros Bibli e Em Questão em suas páginas iniciais e em

seus respectivos volumes 28 (v.28) e volume 29 (v.29) estão presentes nas Figura 7 e 8, respectivamente.

Figura 7 – Proposta de inserção das redes de coautoria na página inicial (a) e no volume 28 (b) da revista Encontros Bibli

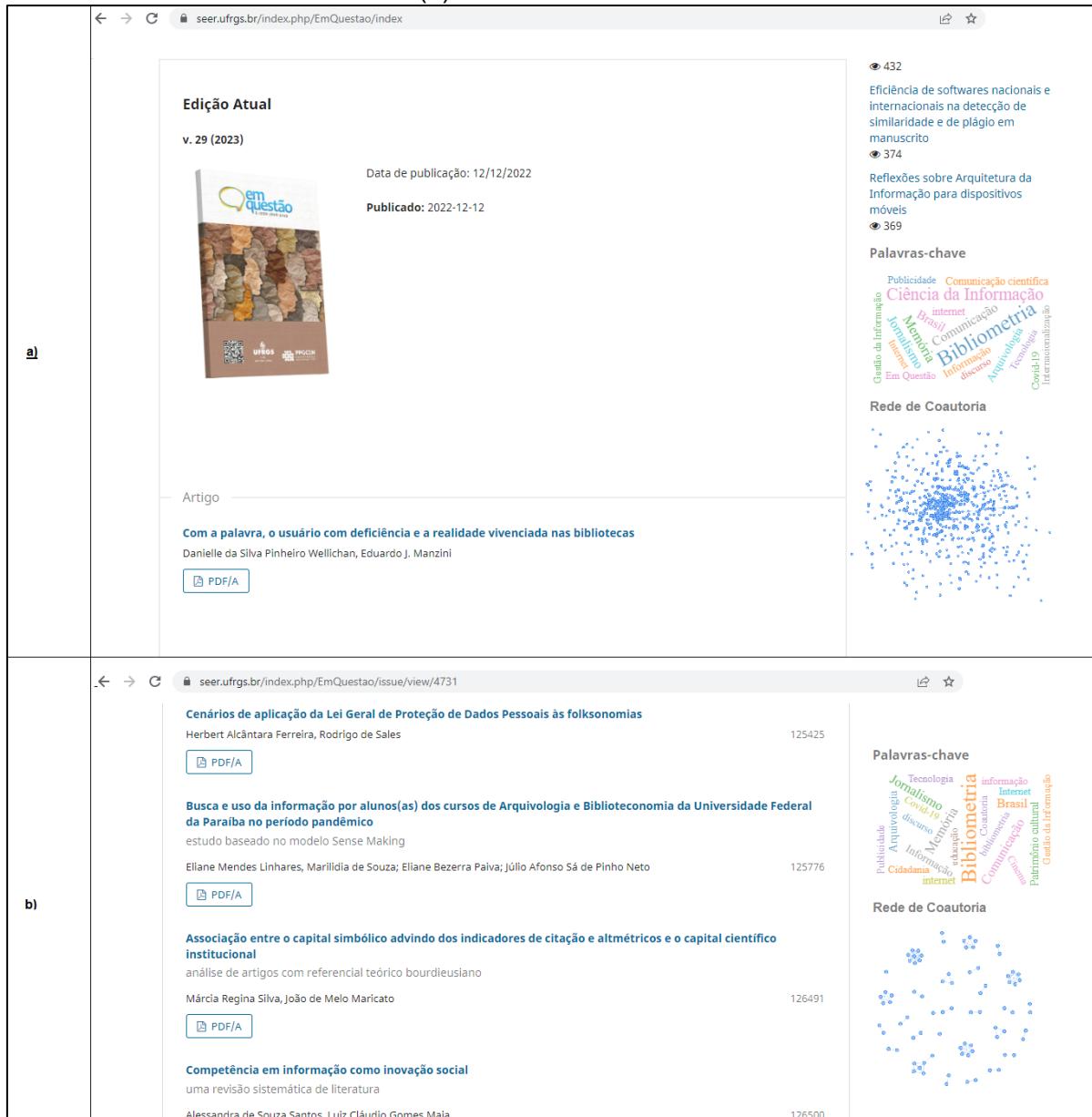


Fonte: Elaboração do autor (2023).

Para cada um dos periódicos, é sugerido que a rede de coautoria contendo todos pesquisadores que já publicaram nas revistas (Figuras 5a e 5b) sejam inseridas nas páginas iniciais, como *hiperlinks*, junto com os *plugins* ou informações adicionais de cada revista. Ou seja, é esperado que os usuários possam clicar na rede e ampliá-la em ambiente externo (nova aba ou janela do navegador), possibilitando a seleção

dos nós da rede e a identificação dos nomes de cada autor. Além do mais, propõe-se inserir uma rede de coautoria em cada um dos números e volumes publicados pelos periódicos, também junto com os plugins e informações adicionais das páginas.

Figura 8 – Proposta de inserção das redes de coautorias na página inicial (a) e no volume 29 (b) da revista Em Questão



Fonte: Elaboração do autor (2023).

A Figura 7a apresenta a incorporação da rede de coautoria de toda a revista em sua página inicial logo abaixo do campo “indexadores”. Se esclarece que a Encontros Bibli tem em sua página inicial (<https://periodicos.ufsc.br/index.php/eb/>) todos seus artigos publicados em seu volume atual e na coluna à direita, informações

relevantes aos usuários tais como: informações aos leitores, autores e bibliotecários, idioma, redes sociais, indexadores e agências de fomento. De maneira análoga, na Figura 7b, a rede de coautorias de um único volume da revista (volume 28), se apresenta da mesma forma que na página inicial, porém, contendo somente os autores que publicaram no referido volume.

Tanto na Figura 7a, quanto 7b, é proposto que as redes sejam *hiperlinks* para uma nova página (ou aba) de modo que, ao clicar na rede, o usuário tenha uma melhor visualização da rede e possa interagir com a rede por meio do seletor de vértices, zoom e pela possibilidade de arrastar os vértices.

No que diz respeito a proposta para revista Em Questão, a Figura 8a apresenta a rede de coautoria contendo todos autores em sua página inicial (<https://seer.ufrgs.br/index.php/EmQuestao/index>) logo abaixo dos *plugins* “Mais lidos” e “Palavras-chave” do OJS, ao lado direito do *web site*. Similar a Encontros Bibli, a revista Em Questão também apresenta todos os seus artigos do volume atual na página inicial do periódico. Já na Figura 8b, é apresentada a rede de coautorias para um volume (ou número) único, a ser inserida logo abaixo do *plugin* “Palavra-chave”. Da mesma maneira que na revista Encontros Bibli, propõe-se que as redes das Figuras 8a e 8b sejam *hiperlinks*, conduzindo o usuário a uma nova janela ou aba de seu navegador de internet, ampliando-as e proporcionando-lhe uma melhor visualização.

## 4 CONSIDERAÇÕES FINAIS

Esta pesquisa apresentou uma proposta de inserção de redes de coautoria nos *web sites* de revistas científicas, enquanto insumo para análises bibliométricas, a partir da raspagem de dados (*web scraping*) de duas revistas de grande relevância para área de Ciência da Informação, Encontros Bibli e Em Questão. Para isso foi implementado um código, em linguagem de programação R, contendo todo processo de leitura, identificação, extração, organização, estruturação e visualização dos dados, este último, elemento principal por se tratar da rede de coautoria. Durante a demonstração do código, construiu-se a rede de coautoria contendo todos autores de todos volumes e números de ambas revistas e também de um único volume ou

número, fornecendo ainda, a possibilidade de refinamento para visualização da rede a partir da frequência mínima de coautoria entre dois autores.

Após a demonstração do uso do código, foi proposto a inserção destas redes enquanto insumo bibliométrico para as revistas, de modo que usuários interessados neste tipo de informação tenha total acesso a todas relações estabelecidas pelos autores que publicaram estudos nas revistas, seja de maneira global (em todos números e volumes) ou de maneira particular (em um volume ou número específico).

Assim, ao responder o problema de pesquisa proposto, o método de *web scraping* pode ser utilizado pelas revistas para construir suas próprias redes de coautoria e fornecer mais informações, de forma automática, passíveis de análises a pesquisadores. Além disso, pode ser utilizado enquanto método de extração de dados de revistas científicas, sobretudo, aquelas construídas via OJS, visto que periódicos que utilizam este sistema, possuem estruturas semelhantes.

Enquanto limitações, o código engloba as produções de documentos institucionais, principalmente editoriais, às produções de artigos científicos, podendo inflar a produção dos editores das revistas, não filtrando as produções por tipologia documental, como visto nos casos das produções de Samile Andréa de Souza Vanz, editora da revista *Em Questão*, e, de Francisco das C. de Souza, Adilson Luiz Pinto, Enrique Muriel-Torrado e Rafael Paula Schmitz, na revista *Encontros Bibli*. Mesmo que essas tipologias documentais também sejam de natureza científica, inclusive sendo indexadas em bases de dados como produção científica dos editores, comprehende-se que essa limitação do código pode impactar diretamente a produção total desses pesquisadores. É sabido que, tradicionalmente no âmbito dos estudos métricos da informação (EMI), a coautoria utiliza como unidade de análise os artigos científicos e não produções tais como editoriais e expedientes. Tal fato poderia ser mitigado se todos esses documentos fossem de autoria estritamente institucional, como os identificados nas revistas: "Comissão Editorial *Em Questão*" e "Editor *Encontros Bibli*".

Enquanto implementações futuras, almeja-se a inserção de indicadores juntamente às redes, como o cálculo de densidade da rede, além da *clusterização* dos atores nas redes. Além disso, pretende-se estender a aplicação do *web scraping* a outras revistas desenvolvidas através do OJS, explorando a extração e visualização

de diferentes informações. Isso poderia incluir aspectos como, por exemplo, a extração, e consequentemente, a visualização de redes da coocorrência de palavras-chave, ampliando assim as possibilidades de análises e compreensão deste tipo de dados, além estender a aplicabilidade do método de *web scraping* no contexto dos EMI. Por fim, ao propor e fornecer um código baseado em *web scraping*, este estudo releva o alto potencial bibliométrico desta técnica, contribuindo com novas perspectivas de técnicas computacionais aplicadas à bibliometria.

## REFERÊNCIAS

- BRITO, R. F.; SHINTAKU, M. Relatório de acompanhamento editorial: plugin para o SEER/OJS. In: ENCONTRO NACIONAL DE EDITORES CIENTÍFICOS, 14., 2013, São Paulo. **Anais** [...]. São Paulo: ABEC, 2013. p. 1-3. Disponível em: [https://ridi.ibict.br/bitstream/123456789/1068/1/RelatoriodeAcompanhamentoEditorialPluginparaOJSSEER\\_ABEC\\_2013.pdf](https://ridi.ibict.br/bitstream/123456789/1068/1/RelatoriodeAcompanhamentoEditorialPluginparaOJSSEER_ABEC_2013.pdf). Acesso em: 15 jan. 2024.
- ASSIS, W. V.; GOMIDE, J. V. B. *Web scraping* em dados públicos: método para extração de dados dos gastos públicos dos vereadores da Câmara Municipal de Belo Horizonte. **Inf. & Inf.**, Londrina, v.24, n. 2, p. 341-319, 2021. DOI: <https://doi.org/10.5433/1981-8920.2021v26n4p319>. Acesso em: 10 set. 2023.
- BOGADO, A. C; ROSAS, F. S.; GRÁCIO, M. C. C. Coautoria institucional na produção científica brasileira sobre hanseníase: uma análise a partir da base de dados Web of Science. **RECIIS**, Rio de Janeiro, v. 16, n. 1, p. 28-47, 2022. DOI: <https://doi.org/10.29397/reciis.v16i1.2371>. Acesso em: 08 de set. 2023.
- CASTANHA, R. G. Conjunto de dados (código e imagens) utilizados em "Visualização de redes de coautoria como insumo bibliométrico às revistas científicas: uma proposta via *web scraping* para os periódicos Em Questão e Encontros Bíblicos". DOI: <https://doi.org/10.5281/zenodo.8087232>. Acesso em: 29 set. 2023.
- GRACIANO, H. L. S. **ScraperCI**: um protótipo de Web scraper para coleta de dados. 2009. 79 f. Dissertação (Mestrado em Ciência da Informação) – Programa de Pós-Graduação em Ciência da Informação, Universidade Federal de São Carlos, São Carlos, 2022. Disponível em: <https://repositorio.ufscar.br/handle/ufscar/17166>. Acesso em: 20 de jan. 2024.
- GRACIANO, H. L. S; RAMALHO, R. A. S. SCRAPERCI: UM WEB SCRAPER PARA COLETA DE DADOS CIENTÍFICOS. **Encontros Bibli**, Florianópolis, v. 28, p. e92471, 2023. DOI: <https://doi.org/10.5007/1518-2924.2023.e92471>. Acesso em 20 de jan. 2024.
- LU, K.; WOLFRAM, D. Measuring author research relatedness: A comparison of word-based, topic-based, and author cocitation approaches. **JASIST**, Nova York, v. 63, n. 10, p. 1973-1986, 2012. DOI: <https://doi.org/10.1002/asi.22628>. Acesso em: 11 de set. 2023.
- NDUNGU, M. W. *Publishing with open journal systems (OJS): a Librarian's perspective*. **Serials Review**, Londres, v. 46, n. 1, p. 21-25, 2020. DOI: <https://doi.org/10.1080/00987913.2020.1732717>. Acesso em 20 de jan. 2024.

OLIVEIRA, C. G.; CASTANHA, R. G.; GRÁCIO, M. C. C. Coautoria dupla nos artigos do campo da ciência da informação: análise dos periódicos brasileiros qualis A1 e A2 (2013-2017). In: MACHADO, R. N.; RODRIGUES, K. O.; BARROS, S. S. (Orgs.). **Diálogos sobre bibliometria e cientometria**. Salvador: EDUFBA, 2021. p. 91-108. Disponível em: <https://repositorio.ufba.br/bitstream/ri/33901/1/Dialogos%20sobre%20bibliometria-miolo-digital.pdf>. Acesso em: 21 mar. 2024.

PERIANES-RODRÍGUEZ, A.; OLMEDA-GÓMEZ, C.; MOYA-ANEGÓN, F. *Detecting, identifying and visualizing research groups in co-authorship networks*. **Scientometrics**, Berlim, v. 82, n. 2, p. 307-319, 2010. DOI: <https://doi.org/10.1007/s11192-009-0040-z>  
Acesso em: 11 de set. 2023.

SANTIN, D. M.; VANZ, S. A. S.; CAREGNATO, S. E. A análise de redes de colaboração científica com base em indicadores bibliométricos. In: FRANCO, S. R. K.; FRANCO, M. E. D. A.; LEITE, D. B. C. (Orgs.). **Educação superior e conhecimento no centenário da Reforma de Córdoba**: novos olhares em contextos emergentes. Porto Alegre: EDIPUCRS, 2019. p 189-207. Disponível em: <https://lume.ufrgs.br/handle/10183/195603>. Acesso em: 10 set. 2023.

URBIZAGASTEGUI, R. La bibliometría, informetría, cienciometría y otras “metrías” en el Brasil. **Encontros Bibli**, Florianópolis, v. 21, n. 47, p. 51-66, 2016.  
DOI: <https://doi.org/10.5007/1518-2924.2016v21n47p51>. Acesso em: 11 de set. 2023.

## NOTAS

### CONTRIBUIÇÃO DE AUTORIA

**Concepção e elaboração do manuscrito:** R. G. CASTANHA  
**Coleta de dados:** R. G. CASTANHA  
**Análise de dados:** R. G. CASTANHA  
**Discussão dos resultados:** R. G. CASTANHA  
**Revisão e aprovação:** R. G. CASTANHA

### CONJUNTO DE DADOS DE PESQUISA

Disponível em: <https://doi.org/10.5281/zenodo.8087232>

### FINANCIAMENTO

Não se aplica.

### CONSENTIMENTO DE USO DE IMAGEM

Não se aplica

### APROVAÇÃO DE COMITÊ DE ÉTICA EM PESQUISA

Não se aplica.

### CONFLITO DE INTERESSES

Não se aplica

### LICENÇA DE USO

Os autores cedem à **Encontros Bibli** os direitos exclusivos de primeira publicação, com o trabalho simultaneamente licenciado sob a [Licença Creative Commons Attribution](#) (CC BY) 4.0 International. Esta licença permite que **terceiros** remixem, adaptem e criem a partir do trabalho publicado, atribuindo o devido crédito de autoria e publicação inicial neste periódico. Os **autores** têm autorização para assumir contratos adicionais separadamente, para distribuição não exclusiva da versão do trabalho publicada neste periódico (ex.: publicar em repositório institucional, em site pessoal, publicar



uma tradução, ou como capítulo de livro), com reconhecimento de autoria e publicação inicial neste periódico.

**PUBLISHER**

Universidade Federal de Santa Catarina. Programa de Pós-graduação em Ciência da Informação. Publicação no [Portal de Periódicos UFSC](#). As ideias expressadas neste artigo são de responsabilidade de seus autores, não representando, necessariamente, a opinião dos editores ou da universidade.

**EDITORES**

Edgar Bisset Alvarez, Ana Clara Cândido, Patrícia Neubert, Genilson Geraldo, Jônatas Edison da Silva.

**HISTÓRICO**

Recebido em: 02-10-2023 – Aprovado em: 05-03-2024 – Publicado em: 15-04-2024