



The Coupler: uma nova ferramenta bibliométrica para análises relacionais de citação, acoplamento bibliográfico e cocitação

Rafael Gutierrez Castanha¹

RESUMO

Introdução: O uso de linguagens de programação no contexto dos Estudos Métricos da Informação tem conquistado espaço junto à comunidade científica da área devido sua praticidade, gratuidade e possivelmente, baixo custo computacional. **Objetivos:** Apresenta uma ferramenta bibliométrica nova, gratuita e alternativa, voltada para análises relacionais de citação, com foco no acoplamento bibliográfico, construída sobre a linguagem de programação R, intitulada *The Coupler*. **Metodologia:** Fundamenta as análises relacionais de citação, cocitação e acoplamento bibliográfico sob a ótica matemática e apresenta o código fonte da ferramenta e as condições em que foi construída e disponibilizada ao público. Para testagem da ferramenta, utiliza dados genéricos e expõe todas suas funcionalidades. Para demonstração de uso com dados reais, opera dados bibliométricos, patentométricos, altométricos e linguagem natural. **Resultados:** A ferramenta é capaz de construir matrizes de citação, acoplamento e cocitação além de calcular os valores normalizados via Cosseno de Salton e Índice de Jaccard das frequências de acoplamento bibliográfico. Ainda, constrói a rede de acoplamento bibliográfico e identifica as unidades de acoplamento responsáveis por cada par de acoplamento. Este último, resultado incomum aos tradicionais softwares bibliométricos. **Conclusão:** Conclui que a ferramenta se comporta de maneira esperada e satisfatória ao processar tanto os dados genéricos quanto os dados bibliométricos, patentométricos, altométricos e de linguagem natural. Dentre os resultados, constata como principal resultado, e diferencial dos demais softwares, a identificação das unidades de acoplamento, bem como os cálculos das frequências de acoplamento normalizadas via Cosseno de Salton e Índice de Jaccard.

PALAVRAS-CHAVE

Acoplamento bibliográfico. Análise de citação. Cocitação. Bibliometria.

Correspondência do autor

¹ Universidade Estadual Paulista,
Marília, SP – Brasil
e-mail: r.castanha@gmail.com

The Coupler: a new bibliometric tool for relational citation, bibliographic coupling and co-citation analysis

ABSTRACT

Introduction: The use of programming languages in the context of Metric Information Studies has been gaining ground with the scientific community in the area due to its practicality,

gratuitousness and possibly low computational cost. **Objective:** Presents a new, free and alternative, bibliometric tool, focused on relational citation analysis, focusing on bibliographic coupling, built on the R programming language, entitled *The Coupler*. It also tests the tool using different data commonly used in bibliometric analyzes and demonstrates its use. **Methodology:** It bases the relational analysis of citation, co-citation and bibliographic coupling from a mathematical point of view and presents the source code of the tool and the conditions in which it was built and made available to the public. For testing the tool, it uses generic data and exposes all its features. For demonstration of use with real data, it operates bibliometric, patentometric, altmetric and natural language data. **Results:** The tool is able to build citation, coupling and co-citation matrices in addition to calculating normalized values via Salton's Cosine and Jaccard Index of bibliographic coupling frequencies. Furthermore, it builds the bibliographic coupling network and identifies the coupling units responsible for each coupling pair. The latter, an unusual result to traditional bibliometric softwares. **Conclusion:** It concludes that the tool behaves in an expected and satisfactory way when processing both generic data and bibliometric, patentometric, altmetric and natural language data. Among the results, the main result, and differential from the other software, is the identification of coupling units, as well as the calculation of normalized coupling frequencies via Salton's Cosine and Jaccard Index.

KEYWORDS

Bibliographic coupling. Citation analysis. Co-citation analysis. Bibliometrics.

CRediT

- **Reconhecimentos:** Não é aplicável.
- **Financiamento:** Este estudo foi parcialmente financiado pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES), Código financeiro: 88887.678240/2022-00.
- **Conflitos de interesse:** O autor certifica que não tem interesse comercial ou associativo que represente um conflito de interesses em relação ao manuscrito.
- **Aprovação ética:** Não é aplicável.
- **Disponibilidade de dados e material:** Todos os dados gerados e analisados durante o presente estudo estão disponíveis no próprio corpo do texto original e nos seus anexos.
- **Contribuições dos autores:** Conceituação, Curadoria de dados, Análise formal, Aquisição de Financiamento, Investigação, Metodologia, Recursos, Supervisão, Escrita - Rascunho original, Escrita - revisão & edição: CASTANHA, R. G.



JITA: BB. Bibliometric methods.

Artigo submetido ao sistema de similaridade



Submetido em: 11/10/2021 – Aceito em: 24/11/2022 – Publicado em: 01/12/2022

Editor: Gildenir Carolino Santos

1 INTRODUÇÃO

No âmbito dos estudos métricos da informação (EMI) é possível encontrar diversos softwares para análises bibliométrica como: Bibexcel, Bibliometrix, CiteSpace, CoPalRed, IN-SPIRE™, InCites, Leydesdorff's Softwares, Metaknowledge, Network Workbench Tool, Publish or Perish, Science of Science (Sci²) Tool, SCImago Graphica, SciMAT, SciVal, VantagePoint, VosViewer, HitsCite, CRExplorer, ScientoPyUI, Bibliomaps (MORAL-MUÑOZ, 2020; MOREIRA; GUIMARÃES; TSUNODA, 2020; FARIA; PERISINI 2021).

Existem outros softwares, como Pajek, Ucinet e Gephi, que tem como foco a construção de grafos (redes) mas que não necessariamente são concebidos no cerne dos EMI. Contudo, são altamente difundidos na literatura especializada da área por se tratar de softwares grátis e pela interface amigável.

Ainda, o uso de linguagens de programação no contexto das Ciências Sociais Aplicadas e conseqüentemente nos EMI também vem ganhando espaço junto à comunidade científica da área. É possível notar este fenômeno dentro da linguagem de programação R, ao constatar a existência de pacotes e bibliotecas voltadas para análises bibliométricas como por exemplo, o *scholar packages*, *wosr package*, *package rScopus*, em que é possível realizar análise bibliométricas extraíndo dados das bases de dados Google Scholar, *Web Of Science* e *Scopus*, respectivamente.

Nesse sentido, tem-se o software Bibliometrix/Biblioshiny, também construído sobre a linguagem de programação R que possibilita análises bibliométricas diversas como análises da produção científica por autor, análises de impacto, leis de Lotka e Bradford, análises relacionais de citação, cocitação e acoplamento bibliográfico, análises fatoriais, dentre outras diversas funcionalidades. O Bibliometrix utiliza o pacote *shiny* para gerar sua interface gráfica, esta, construída em *html*, o que possibilita a utilização da ferramenta por meio do navegador de internet do usuário.

Dado a importância e a relevância adquirida pelo Bibliometrix, análises que utilizam a linguagem R estão cada vez mais inseridas nos estudos bibliométricos. A possibilidade de trabalhar com alto volume de dados aliada à praticidade da linguagem, a gratuidade e a um custo computacional relativamente baixo são alguns dos motivos que explicam essa expansão. A volúpia de dados sugere o uso recorrente de linguagens de programação por serem capazes de processar grandes quantidades de dados.

Diante do exposto, esta pesquisa tem como objetivo apresentar uma nova, gratuita e alternativa, ferramenta bibliométrica, voltada para análises relacionais de citação, com foco no acoplamento bibliográfico, construída sobre a linguagem de programação R, intitulada *The Coupler*. De maneira específica, testar a ferramenta utilizando dados diversos comumente utilizados em análises bibliométricas. Para isso, serão apresentadas as motivações para construção da ferramenta bem como as bases teóricas das análises que sustentam a construção do código da por trás do software.

Esta nova ferramenta tem foco nas análises relacionais de citação com atenção especial ao acoplamento bibliográfico, podendo ser compreendida, em sua essência, como uma ferramenta de análises de similaridade entre quaisquer tipos de unidade de análise. A concepção de novos recursos computacionais, sobretudo voltados ao campo dos EMI, é parte fundamental do progresso do campo como um todo e fortalece o desenvolvimento de novos estudos, contribuindo diretamente para evolução científica e tecnológica da bibliometria.

2 ANÁLISES RELACIONAIS DE CITAÇÃO

Dentro do escopo da bibliometria é possível afirmar que há uma gama de indicadores que variam, essencialmente, a partir dos elementos que se referem, porém, indicadores voltados as análises de citação, podem ser classificados, segundo Grácio (2020), em univariados e relacionais de citação.

Análises univariadas de citação remetem ao impacto e buscam aferir o reconhecimento da comunidade científica a partir das citações recebidas por determinado conjunto de documentos, autores, instituições, ou qualquer outra unidade de análise. Neste quesito, encontram-se, por exemplo, taxas de citação média por ano por artigo, o Fator de Impacto, ou *CiteScore*. Em relação ao segundo item, busca estabelecer agrupamentos entre unidades de análise diversas. Neste ponto, encontram-se as análises de acoplamento bibliográfico e cocitações.

Em relação a análise de acoplamento bibliográfico (AB), o método foi proposto por Kessler (1963a; 1963b) a fim de identificar como artigos se agrupam entre si por meio do compartilhamento de referências citadas. Ou seja, se dois artigos compartilham pelo menos uma mesma referência (citam a mesma referência), estes artigos estão bibliograficamente acoplados. Kessler (1963a) buscou avaliar o AB segundo dois tipos de agrupamentos entre artigos denominados GA e GB: GA) Dado um conjunto de artigos, todos estão acoplados entre si (citam pelo menos uma referência em comum); GB) todos os artigos estão acoplados a um determinado artigo.

A quantidade de referências citadas em comum por dois artigos foi denominada força de acoplamento bibliográfico (ou frequência de acoplamento bibliográfico) e as referências responsáveis por unir estes dois artigos são definidas como unidades de acoplamento. Desta forma, o AB configura-se como uma relação entre artigos citantes (citante-citante)

Anos depois, Small (1973) cunhou a análise de cocitação (AC) pautada na coocorrência de dois documentos citados conjuntamente em pelo menos uma lista de referências, isto é, a AC mensura em quantos documentos outros dois documentos são cocitados (citados concomitantemente). Assim, esta análise explicita a relação entre artigos citados (citado-citado). Em resumo, enquanto o AB quantifica a similaridade de referências entre dois documentos, a AC expõe a recorrência em que duas referências são citadas de maneira conjunta. Esta quantificação pode ser entendida como frequência de cocitação entre dois artigos.

As análises relacionais de acoplamento bibliográfico e cocitação evoluíram ao longo do tempo de modo que White (1981) e White Griffith (1981) propuseram a análise de cocitação entre autores (ACA). Nesta perspectiva, verifica-se em quantos artigos dois autores foram citados simultaneamente. Para isso, considera-se a obra do autor citado como única e não mais a referência do artigo citado.

Já em 2008, Zhao e Strotmann (2008) enunciaram o acoplamento bibliográfico de autores (ABA). Esta nova análise pretende acoplar dois pesquisadores sob duas perspectivas: utilizando documentos citados ou autores citados. No primeiro caso, é computado quantas referências em comum dois pesquisadores compartilham. Já no segundo, considera-se a obra do autor citado como única e calcula-se a quantidade de autores citados em comum por dois pesquisadores.

Desde os primeiros estudos propostos por Kessler (1963a; 1963b) as análises relacionais ganharam espaço no campo da bibliometria e diversos métodos de normalização das frequências de acoplamento, como por exemplo o Índice de Jaccard (IJ) e o Cosseno de Salton (CS), vem sendo debatidos desde Sen e Gan (1983). Estes métodos podem ser descritos nas Equações 1 e 2 como:

$$IJ_{(A,B)} = \frac{(A \cap B)}{(A \cup B)} = \frac{Coupling(A,B)}{(A) + (B) - (A \cap B)} \quad [1]$$

$$CS_{(A,B)} = \frac{(A \cap B)}{\sqrt{(A) \times (B)}} = \frac{Coupling(A,B)}{\sqrt{(A) \times (B)}} \quad [2]$$

Em que, A e B são duas unidades de análise quaisquer a serem acopladas e *Coupling* (A, B) representa quantas unidades de acoplamento A e B compartilham (citam em comum).

Tradicionalmente, as análises relacionais de citação são basicamente o resultado de interações matriciais concebidas a partir da matriz adjacência de citação da situação analisada. Ou seja, dada uma matriz de citação $C_{n \times m}$, a matriz contendo as frequências de acoplamento bibliográfico (F_{ab}) é dada por: $F_{ab} = C \times C^T$, enquanto a matriz contendo as frequências de cocitação (F_c) é dada por $F_c = C^T \times C$. Exemplifica-se na Equação 3:

$$C = \begin{array}{c|cccc} & i & j & k & l \\ \hline \mathbf{A} & 1 & 1 & 1 & 1 \\ \mathbf{B} & 1 & 0 & 0 & 1 \\ \mathbf{C} & 1 & 0 & 0 & 0 \end{array} \quad C \times C^T = \begin{array}{c|ccc} & \mathbf{A} & \mathbf{B} & \mathbf{C} \\ \hline \mathbf{A} & 4 & 2 & 1 \\ \mathbf{B} & 2 & 2 & 1 \\ \mathbf{C} & 1 & 1 & 1 \end{array} \quad C^T \times C = \begin{array}{c|cccc} & i & j & k & l \\ \hline i & 3 & 1 & 1 & 2 \\ j & 1 & 1 & 1 & 1 \\ k & 1 & 1 & 1 & 1 \\ l & 2 & 1 & 1 & 2 \end{array} \quad [3]$$

Em que A, B e C são unidades citantes e *i, j, k, l* são unidades citadas. Além disso, a segunda matriz corresponde a matriz de acoplamento bibliográfico e a terceira matriz corresponde a matriz de cocitação. A título de exemplo, as unidades A e B possuem força de acoplamento igual a 2 (as unidades de acoplamento são *i* e *l*); as unidades *i* e *l* possuem frequência de cocitação igual a 2 (foram citados simultaneamente por A e B). No que diz respeito à diagonal principal das matrizes de acoplamento e cocitação, estes elementos representam uma ligação reflexiva de um determinado elemento com ele mesmo. Em matrizes de cocitação, esta relação reflexiva representa a quantidade de elementos que o citaram. Por exemplo, o item *i* foi citado por A, B e C, logo, o elemento da diagonal principal relativo a *i*, é igual a 3. Já em matrizes de acoplamento, estes valores são pouco representativos. Nesse sentido, Grácio e Oliveira (2015) alertam esta situação e sugerem: eleger a maior frequência do elemento com os demais; adotar a média da frequência de cada elemento como valor da diagonal; registrar zeros na diagonal; ou definir a diagonal como um conjunto de valores ausentes (*missing*). Dentre elas, a última alternativa apresenta maior aceitação e uso na comunidade dos pesquisadores, ao ser mais facilmente executável e menor viés conceitual.

É possível observar ainda outro resultado, a cardinalidade da lista de unidades citadas de A é igual quatro (*i, j, k, l*), de B é igual a 2 (*i, l*), e de C é igual a 1 (*i*). A partir destas informações, é possível normalizar os valores das frequências de acoplamento via Cosseno de Salton e/ou Índice de Jaccard, como apresenta a Equação 4, respectivamente.

$$Acoplamento_{CS} = \begin{array}{c|ccc} & \mathbf{A} & \mathbf{B} & \mathbf{C} \\ \hline \mathbf{A} & 1 & 0,71 & 0,50 \\ \mathbf{B} & 0,71 & 1 & 0,71 \\ \mathbf{C} & 0,50 & 0,71 & 1 \end{array} \quad Acoplamento_{IJ} = \begin{array}{c|ccc} & \mathbf{A} & \mathbf{B} & \mathbf{C} \\ \hline \mathbf{A} & 1 & 0,50 & 0,25 \\ \mathbf{B} & 0,50 & 1 & 0,50 \\ \mathbf{C} & 0,25 & 0,50 & 1 \end{array} \quad [4]$$

As Equações 3 e 4 explicitam uma consequência do cálculo matricial entre a matriz de citação C e sua transposta: as matrizes de acoplamento (normalizada ou não) e a matriz de cocitação são obrigatoriamente quadradas e simétricas com relação a diagonal. Ou seja, os elementos da parte triangular superior são os mesmos da parte triangular inferior. Este resultado infere que tanto o AB (relação citante-citante) quanto a AC (relação citado-citado) não representam relações direcionadas, como a matriz de citação (relação citante-citado).

3 METODOLOGIA

A fim de apresentar o *The Coupler* foram organizadas três etapas. A primeira descreve as informações referentes aos seus aspectos computacionais como sua construção, hospedagem e funcionalidade. Nesta etapa foram apresentadas a concepção da ferramenta, sua interface gráfica, seu código, e como o usuário pode executar o *The Coupler* de maneira *online* ou *offline*, além de informações sobre os arquivos aceitos pela ferramenta (Figuras 1 e 2).

Na segunda etapa, todas as funcionalidades do *The Coupler* são demonstradas utilizando os dados presentes nas Equações 3 e 4. Todas as possibilidades de processamentos abrigadas na ferramenta são demonstradas, bem como a exportação de dados (Figuras 3 a 14).

Na terceira etapa, são realizados testes utilizando dados bibliométricos, patentométricos, altmétricos e de linguagem natural. Os dados bibliométricos utilizados foram obtidos após busca pelo termo “altmetria” em *todos os campos* na base de dados *Web Of Science*, gerando três resultados: Doc.1) Maricato e Martins (2017); Doc.2) Rocha e Silva (2020); Doc.3) Gouveia (2019). Para isso, os três documentos foram acoplados tomando como unidades de acoplamento os autores citados em cada documento (Figura 15a). Para extração dos autores citados, foi utilizada função *export selected authors* do software VosViewer (Apêndice 2).

Para o processamento de dados patentométricos, acoplou-se três patentes, presentes na base de dados *Derwent Innovation Index*, de códigos: CN113317206 (2021), CN112450208 (2021) e CN112229155 (2021). Neste caso, acoplou-se as três patentes via patentes citadas (Apêndice 3). Este resultado é apresentado na Figura 15b.

Já com relação aos dados altmétricos, foram utilizados os dados de Delbianco (2022, p. 114), referente a seguidores do *Twitter* de três perfis de revistas científicas da área de Ciência da Informação: Acervo: Revista do Arquivo Nacional; AtoZ: novas práticas em informação e conhecimento; Ciência da Informação em Revista. Desta maneira, acoplou-se os três perfis via seguidores em comum (verificou-se a quantidade de seguidores em comum). O acoplamento utilizando seguidores enquanto unidade de acoplamento está presente na Figura 15c.

Como última análise, foi realizado um processamento de linguagem natural comparando as palavras (em comum) presentes nos resumos dos três artigos supracitados (Doc.1: MARICATO E MARTINS (2017); Doc.2: ROCHA E SILVA (2020); Doc.3: GOUVEIA (2019)). Para extração das palavras, foi utilizada a função *export selected terms* do software VosViewer. Por padrão, o software sugere usar as 60% palavras mais relevantes do resumo processado. Para esta análise, utilizou-se o padrão sugerido pelo software e acoplou-se os resumos via palavras em comum (Apêndice 4). Esta análise está presente na Figura 15d.

Todos os testes presentes nas seções subsequentes foram realizados utilizando tanto a versão *online* quanto *offline* do *The Coupler*. O computador utilizado em todas

operações possui sistema operacional *Windows 10*, processador *Intel Core(TM) i7-8550UCPU @ 1.80GHz-1.99 GHz* e 8 Gigabytes de memória RAM.

4 THE COUPLER

A partir das bases teóricas das análises relacionais de citação, foi elaborado o *The Coupler*, uma aplicação *web* (*web-app*) voltada para análises relacionais de citação, com foco em análises de acoplamento. A ferramenta é capaz de construir matrizes de citação, acoplamento e cocitação. Além desta perspectiva matricial, a ferramenta calcula os valores normalizados via Cosseno de Salton e Índice de Jaccard das frequências de acoplamento bibliográfico admitindo qualquer tipo de unidade de análise e qualquer tipo de unidade de acoplamento¹. Ainda, constrói a rede de acoplamento bibliográfico e identifica as unidades de acoplamento responsáveis por cada par de acoplamento. Este último, resultado incomum aos tradicionais softwares bibliométricos. O código fonte está presente no Apêndice 1, podendo ser executado via linguagem de programação R (colando o código no R) ou acessado via *website* <https://rafaelcastanha.shinyapps.io/thecoupler>. Para hospedagem do *web-app*, foi utilizado o servidor privado da *shinyapps.io* do R-Studio que permite a implementação de aplicativos *shiny* no ambiente *web*.

No momento da submissão deste artigo, o *The Coupler* se encontra em processo de registro junto ao Instituto Nacional da Propriedade Industrial (INPI) por intermédio da Agência Unesp de Inovação (AUIN). O registro requisitado é de *registro de programa de computador*, embasado na Lei de Direito Autoral (Lei nº 9.610/1998) e na Lei de Software (Lei nº 9.609/1998). Conseqüentemente, após o registro, o autor garante a manutenção e assistência (aos usuários) do *The Coupler*.

A disponibilidade da ferramenta como um aplicativo *online* facilita a divulgação e o acesso de diversos dispositivos e sistemas operacionais como *Linux*, *Mac-OS* e *Windows*. Além de ser possível acessá-la via dispositivos móveis como tablets e celulares com diferentes sistemas como por exemplo *Android* ou *ios*. Também é possível acessar o *The Coupler* por meio de diferentes navegadores como Google Chrome, Microsoft Edge, Safari, dentre outros. Desta maneira, não necessita de instalação prévia da linguagem de programação R, somente do arquivo a ser processado.

O recurso, tanto em versão *online* quanto *offline* (acessando via R) admite somente arquivos de extensão *.txt* organizado em colunas com cabeçalho que esteja tabulado, separado por vírgulas ou por ponto e vírgula. Tomando como exemplo as unidades A, B e C presentes na Equação 2, a Figura 1 apresenta estes três tipos de organização. Dessa forma, o *The Coupler* tem como requisito, a organização prévia, e possivelmente manual, dos arquivos processados. Alguns softwares, como o VosViewer por exemplo, fornecem a opção de extração de unidades (autores, artigos ou periódicos) citadas por meio de sua função *export selected authors*, *export selected cited references* ou *export selected sources*.

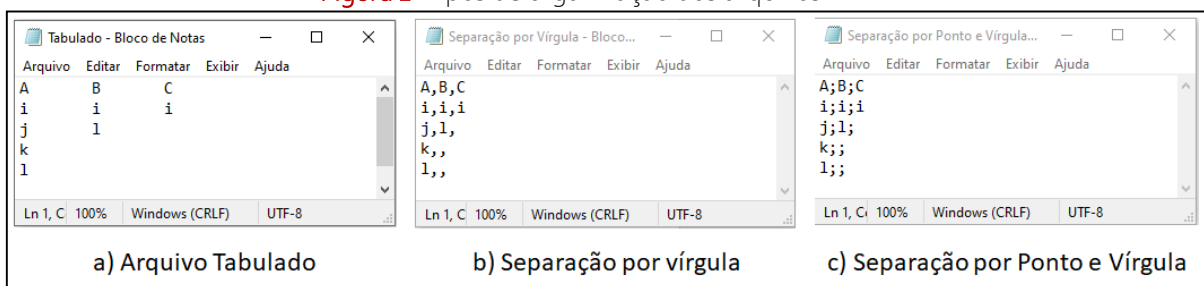
A organização manual dos itens citados pode ser entendida como uma etapa de pré-processamento do *The Coupler*. Esta etapa pode ser de grande serventia para obras que não estejam indexadas em bases de dados passíveis de extração automática de referências

¹ Distingue-se unidades de análise e de acoplamento. Unidades de análise são as unidades citantes que serão submetidas a acoplamento: artigos, periódicos, autores, entre outros. Unidades de acoplamento são as unidades citadas, que vincularão duas unidades de análise. Para exemplificar, em uma análise de acoplamento entre autores a partir de artigos citados em comum, os autores representam unidades de análise e os artigos, as unidades de acoplamento

citadas (como *Web Of Science*, *Socopus* e *Dimensions*), para obras não digitalizadas em que a extração obrigatoriamente será feita de maneira manual e para dados não bibliográficos, em que a organização ou coleta seria realizada manualmente.

Diferentemente de softwares como o VosViewer e o Bibliometrix que processam dados vindos diretamente das bases de dados em formatos do tipo *.csv*, *.xlms* ou *.ris*, o *The Coupler* possui esta função, porém, a ferramenta é capaz de processar qualquer tipo de unidade, como por exemplo listas de: DOI (*Digital Object Identifier*), ORCID, pesquisadores que compõem grupos de pesquisas ou departamentos, áreas de pesquisa de interesse, pesquisadores presentes em departamentos, referências sugeridas em planos de ensino de disciplinas, seguidores em redes sociais, entre outros. Dessa forma, as demais ferramentas supracitadas podem ser complementares e não excludentes entre si.

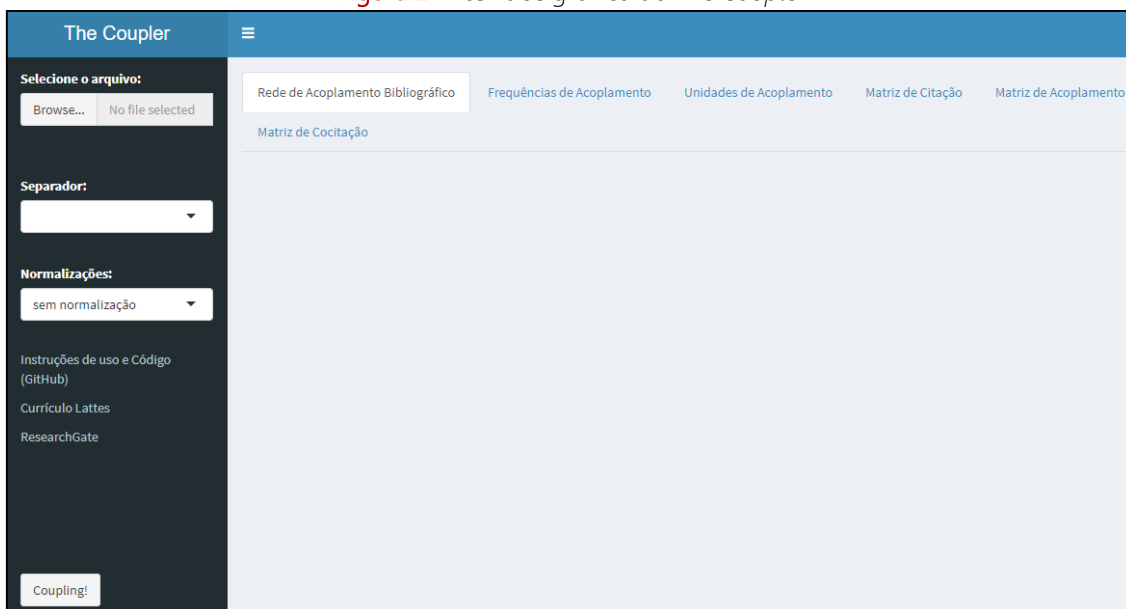
Figura 1. Tipos de organização dos arquivos



Fonte: Elaboração própria

A opção pela extensão *.txt* se dá pelo fato de a extensão ser leve e de fácil acesso em vários dispositivos. Desta maneira, o único requisito para a utilização do *The Coupler* é a existência de um arquivo previamente organizado. A Figura 2 ilustra a interface gráfica da ferramenta.

Figura 2. Interface gráfica do *The Coupler*



Fonte: Elaboração Própria

Ao inserir o arquivo tabulado em “Selecionar o arquivo”, o usuário deve escolher o tipo de separador existente em seu arquivo: separado por vírgula ou por ponto e vírgula. Após isso, ao clicar em “Coupling!” a ferramenta retornará seis resultados, que se encontram em formato de abas na Figura 2: i) a rede de acoplamento bibliográfico; ii) as

frequências de acoplamento entre cada par de unidade e suas normalizações via Cosseno de Salton e Índice de Jaccard (Equação 1) organizadas em formato de tabela; iii) as unidades de acoplamento responsáveis por acoplar cada par de unidade, organizados em formato de tabela; iv) a matriz de citação; v) a matriz de acoplamento entre as unidades citantes; vi) a matriz de cocitação entre as unidades citadas.

Diante destes seis resultados, o usuário poderá exportar todos os resultados obtidos, uma vez que a ferramenta permite salvar a imagem oriunda da rede de acoplamento bibliográfico e oferece a opção de download da tabela contendo as frequências de acoplamento bibliográfico entre cada unidade analisada, a tabela contendo todas as unidades de acoplamento e as matrizes de citação, acoplamento e cocitação. Com exceção da imagem, todos arquivos são salvos em formatos *.txt* tabulados e organizados em colunas.

Ainda, o usuário poderá optar por visualizar a rede sob três perspectivas ao clicar em “Normalizações”: i) sem normalização: retornará a rede de acoplamento valorada pelos valores absolutos das frequências de acoplamento; ii) Cosseno de Salton: retornará a rede de acoplamento valorada pelos valores normalizados das frequências de acoplamento via Cosseno de Salton; iii) Índice de Jaccard: retornará a rede de acoplamento valorada pelos valores normalizados das frequências de acoplamento via Índice de Jaccard.

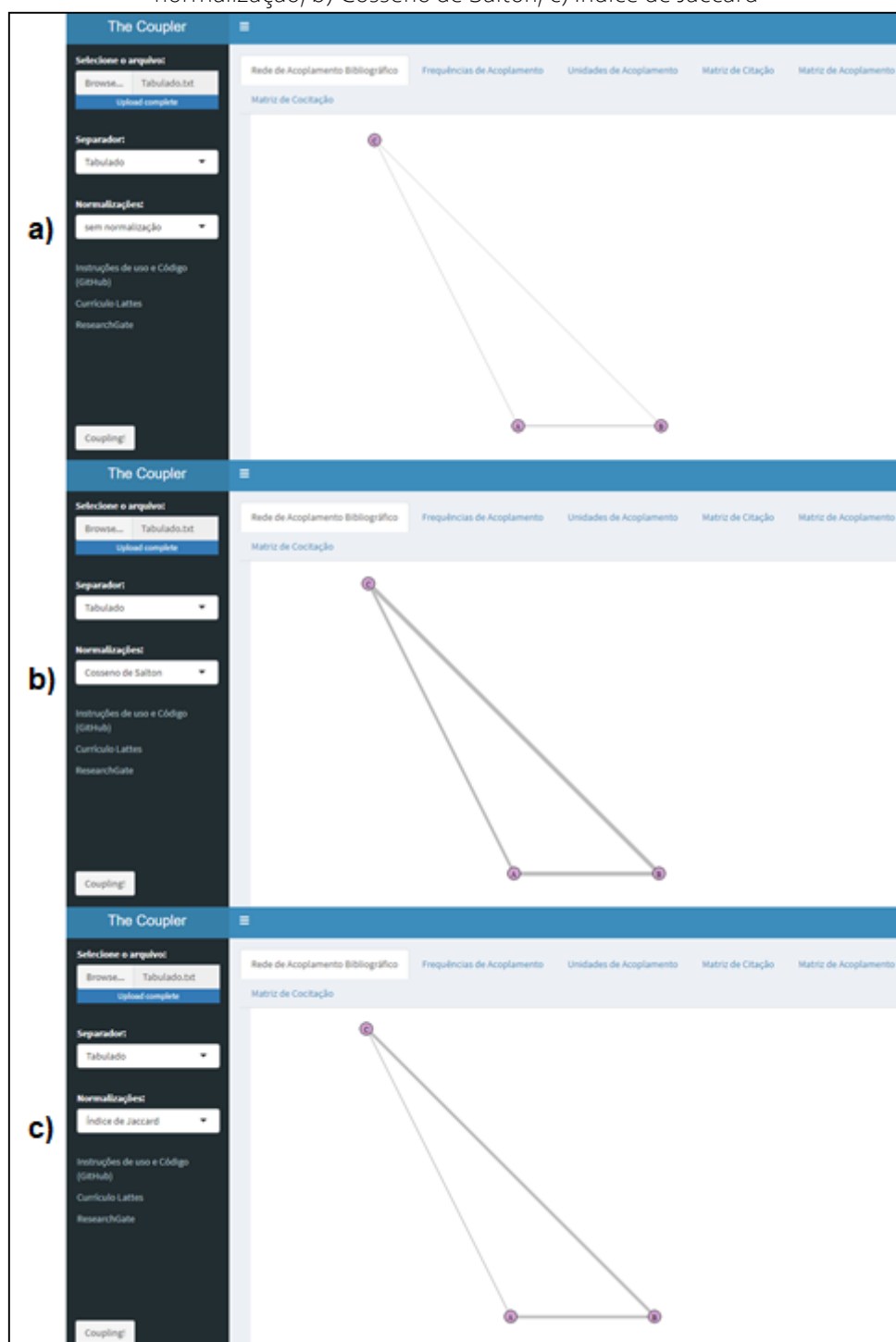
Por mais que a ideia central da ferramenta sejam as análises bibliométricas e consequentemente o uso de unidades bibliográficas como autores ou artigos, a ferramenta não é capaz de distinguir se os dados de entrada tratam de elementos bibliográficos ou não, trazendo maior diversidade para seu uso. Assim, as análises de similaridade (acoplamento) ou co-ocorrência (cocitação) podem ser estendidas para quaisquer unidades de análise como por exemplo, unidades altmétricas e patentométricas. Ou ainda, qualquer tipo de análise de intersecção (similaridade) entre conjuntos.

4 DEMONSTRAÇÃO DE ANÁLISES VIA THE COUPLER

Para apresentação de análises e testes do *The Coupler*, foi utilizado, inicialmente, o arquivo apresentado na Figura 1 para demonstração do funcionamento. Após este teste foram utilizados dados bibliométricos, patentométricos, altmétricos e linguagem natural. Desta maneira, os arquivos presentes na Figura 1 podem ser entendidos como unidades de análise qualquer (autores, documentos, instituições, dentre outros). Para o primeiro teste, a Figura 3 (a, b, c) apresenta a rede de acoplamento e as possibilidades de normalização via “sem normalização”, Cosseno de Salton ou Índice de Jaccard.

Ao escolher entre as três opções, as arestas variam espessura, de modo que, quanto maior a espessura, maior a proximidade entre as unidades analisadas ou ainda, quanto mais espessas as arestas, maior a frequência (força) de acoplamento bibliográfico.

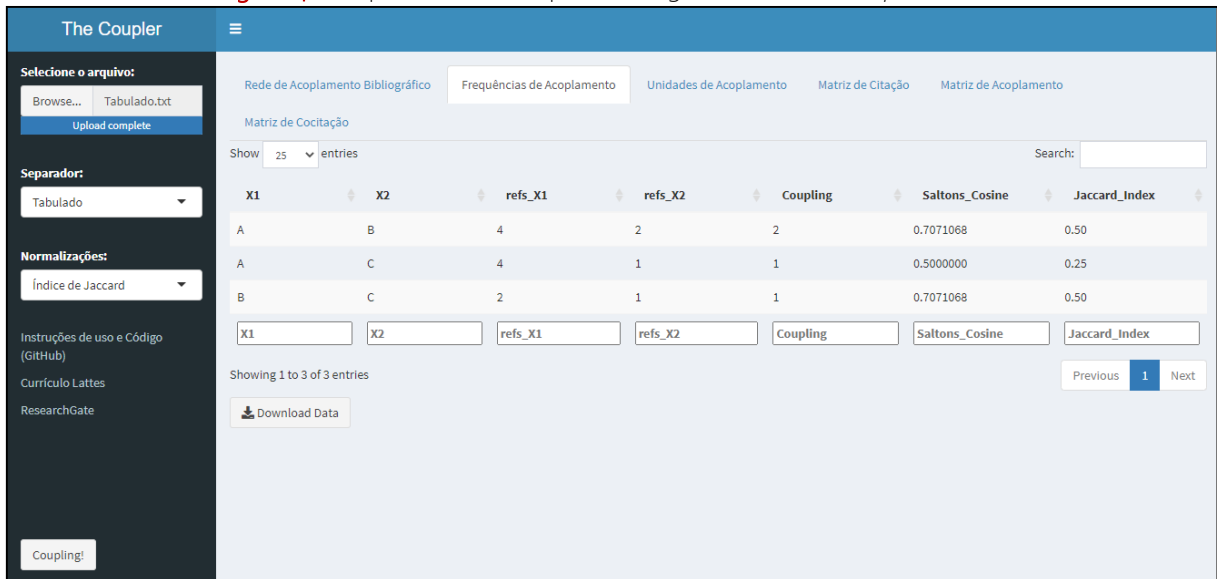
Figura 3. Rede de acoplamento bibliográfico gerada via *The Coupler* utilizando: a) sem normalização; b) Cosseno de Salton; c) Índice de Jaccard



Fonte: Elaboração via *The Coupler*

A visualização da rede de acoplamento bibliográfico utiliza a biblioteca de grafos *igraph* presente no R. É sabido que vários softwares são capazes de gerar redes diversas, e neste contexto, o foco do *The Coupler* não é propriamente a visualização do grafo, pois já existem softwares diversos para isso, mas sim os aspectos relacionais presentes por trás da representação e encontrados nas abas subjacentes a “Rede de Acoplamento Bibliográfico”. Assim, Figura 4, apresenta os resultados encontrados em “Frequências de Acoplamento”.

Figura 4. Frequências de Acoplamento gerada via *The Coupler*



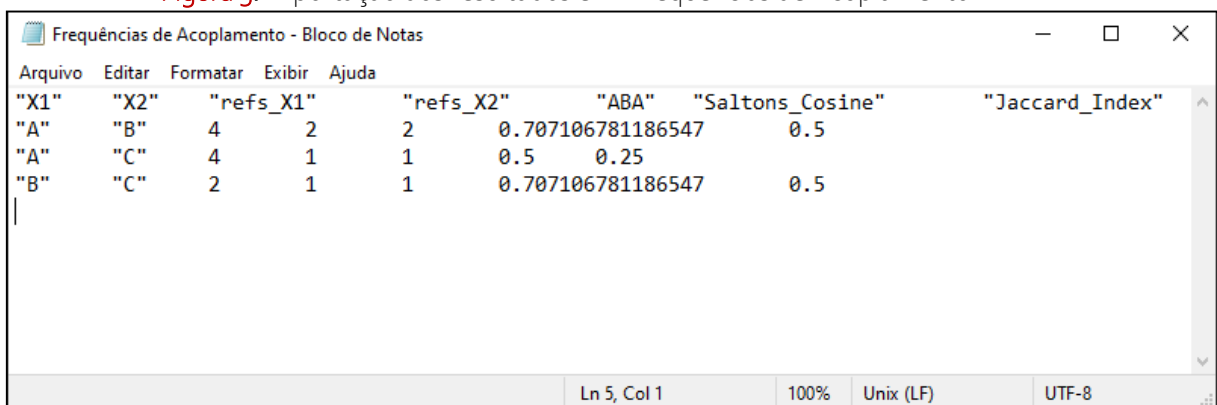
Fonte: Elaboração via *The Coupler*

A Figura 4 apresenta os elementos: i) “X1” e “X2”: representam as unidades de análise a serem comparadas (acopladas); ii) “refs_X1” e “refs_X2”: cardinalidade dos itens citados por “X1” e “X2” respectivamente; iii) “Coupling”: itens citados em comum por X1 e X2, respectivamente; iv) Saltons_Cosine: Valores de “Coupling” normalizados via Cosseno de Salton; v) Jaccard_Index: Valores de “Coupling” normalizados via Índice de Jaccard.

A título de exemplo, leia-se na primeira linha: a unidade A possui lista de referências composta por 4 elementos, a unidade B possui lista de referências composta por 2 elementos de modo que ambos citaram 2 elementos em comum. Este valor normalizado via Cosseno de Salton é igual a 0,7071068 e normalizado via Índice de Jaccard é igual a 0,5.

Nesta aba, é possível exportar estes resultados utilizando o botão “Download Data”. Ao exportar estes dados, é gerado um arquivo de extensão *.txt* nomeado “Frequências de Acoplamento.txt”, como apresenta a Figura 5.

Figura 5. Exportação dos resultados em “Frequências de Acoplamento”



Fonte: Elaboração via exportação do *The Coupler*

Os dados presentes na Figura 5 são tabulados e organizados em colunas. Tal fato facilita sua utilização em planilhas eletrônicas. Além disso, é possível utilizar o campo

“Search” para pesquisar unidades de análise que deseja dar enfoque além da possibilidade de navegar pelas páginas. Por padrão, a ferramenta admite os 25 resultados iniciais.

Ainda, o código por trás do *The Coupler* exclui colunas nulas, isto é, caso uma unidade não tenha citado nenhuma referência, esta unidade será excluída da análise, uma vez que não se acoplará com nenhuma outra unidade. Este caso é análogo a artigos que não possuem referências, ou seja, em que o autor do artigo não cita ninguém em seu estudo.

Em consequente, a aba “Unidades de Acoplamento” traz o resultado mais relevante e inédito dentre todos os presentes na ferramenta, que é a identificação das unidades de acoplamento, isto é, os elementos em comum por cada par de unidade analisada. A Figura 6 ilustra a identificação de todas as unidades de acoplamento.

Figura 6. Unidades de acoplamento identificadas via *The Coupler*

Unidades de Análise	Unidades de Acoplamento
A...B	i; l
A...C	i
B...C	i

| 12

Fonte: Elaboração via *The Coupler*

A Figura 6 identifica os elementos responsáveis pelo acoplamento entre as unidades A, B e C. Desta maneira, exemplifica-se com a primeira linha: As unidades A e B foram acopladas pelos elementos i e l. Ou seja, i e l são unidades de acoplamento. Analogamente ao exemplo anterior, é possível exportar os resultados encontrados clicando em “Download Data”. Os resultados também são organizados em colunas e tabulados como apresenta a Figura 7.

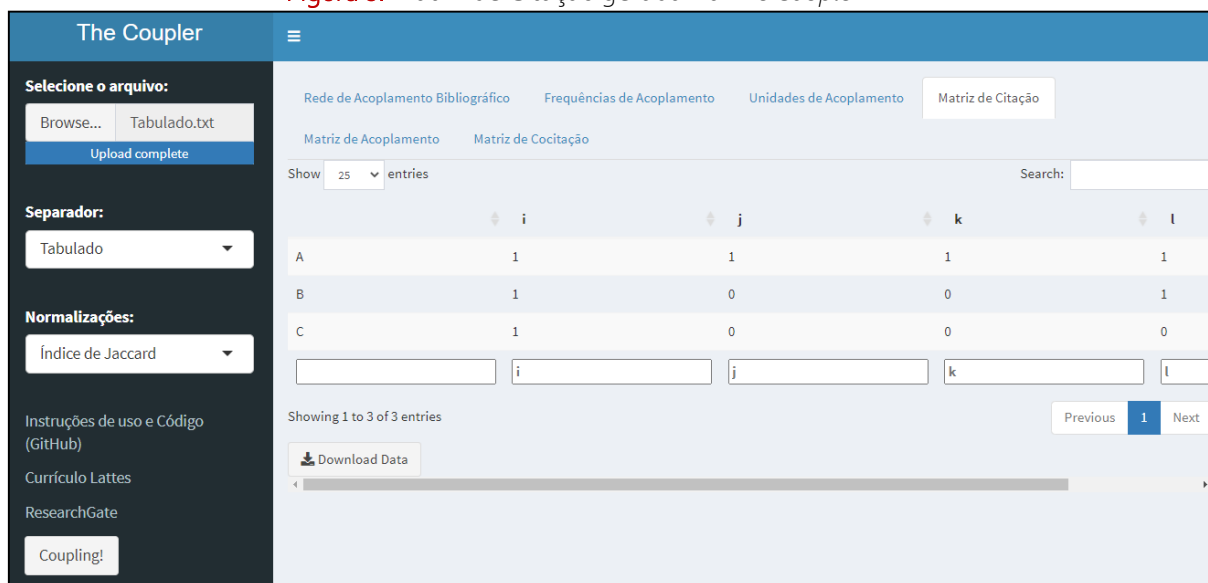
Figura 7. Exportação das unidades de acoplamento via *The Coupler*

```
"Unidades de Análise"  "Unidades de Acoplamento"
"A...B" "i; l"
"A...C" "i"
"B...C" "i"
```

Fonte: Exportação via *The Coupler*

A próxima aba a ser apresentada é a “Matriz de Citação”. Esta aba apresenta a matriz composta pelas unidades citantes e citadas, similar a matriz C da Equação 3. Ou seja, é apresentada a matriz adjacência de citação que dará origem as análises de acoplamento bibliográfico e cocitação. Esta matriz é apresentada na Figura 8.

Figura 8. Matriz de Citação gerada via *The Coupler*

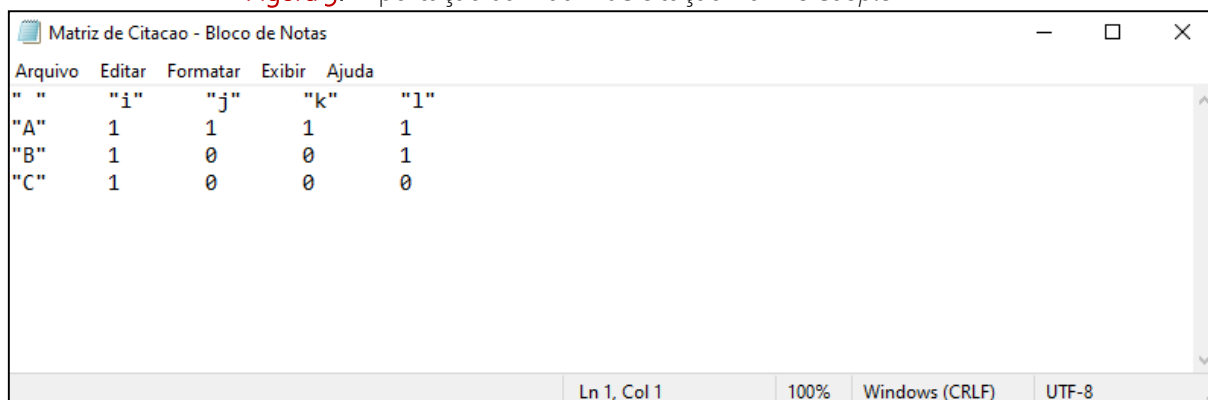


Fonte: Elaboração via *The Coupler*

A Matriz adjacência de citação presente na Figura 8 representa a relação assimétrica de citação entre as unidades A, B e C e os elementos i, j, k, l . Similar aos exemplos anteriores, é possível exportar esta matriz, como apresenta a Figura 9.

| 13

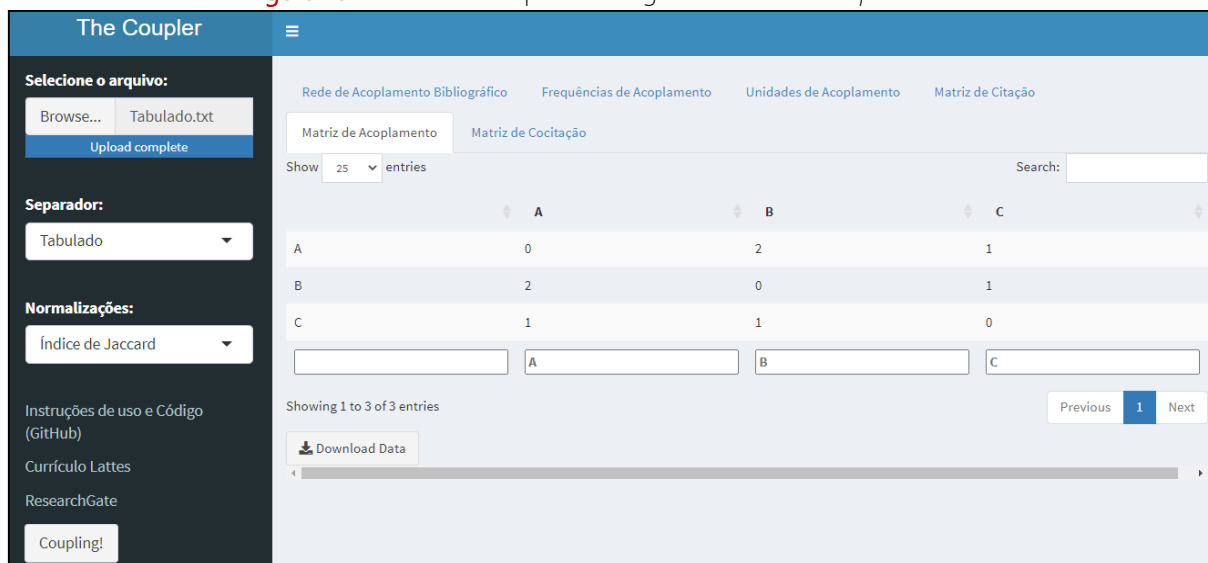
Figura 9. Exportação da matriz de citação via *The Coupler*



Fonte: Exportação via *The Coupler*

A partir da matriz de citação é possível obter as matrizes de acoplamento e cocitação por meio das operações matriciais entre a matriz de citação C e sua transposta C^T . Assim, na aba seguinte há a “Matriz de Acoplamento” que representa a relação $C \times C^T$. Esta relação está presente na Figura 10.

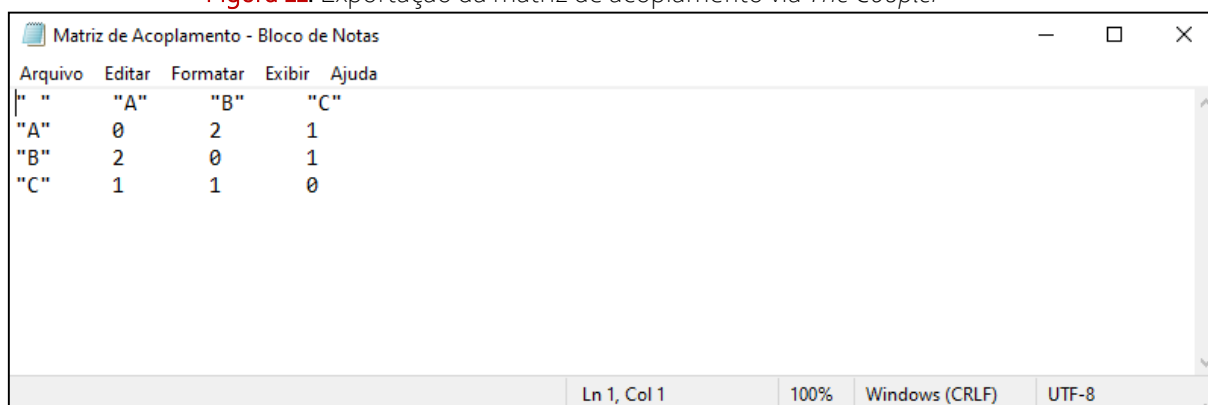
Figura 10. Matriz de Acoplamento gerada via *The Coupler*



Fonte: Elaboração via *The Coupler*

A Figura 10 explicita, em valores absolutos e em forma matricial, a relação de acoplamento bibliográfico entre as unidades A, B e C. Este resultado é similar ao encontrado na coluna “Coupling” da aba “Frequências de Acoplamento”. A opção pela redundância de resultados se dá pelo fato de, tradicionalmente, as análises relacionais serem tratadas do ponto de vista matricial, assim, é oferecido esta opção ao usuário. Ainda, a exportação destes dados é realizada de maneira análoga as demais e está presente na Figura 11.

Figura 11. Exportação da matriz de acoplamento via *The Coupler*

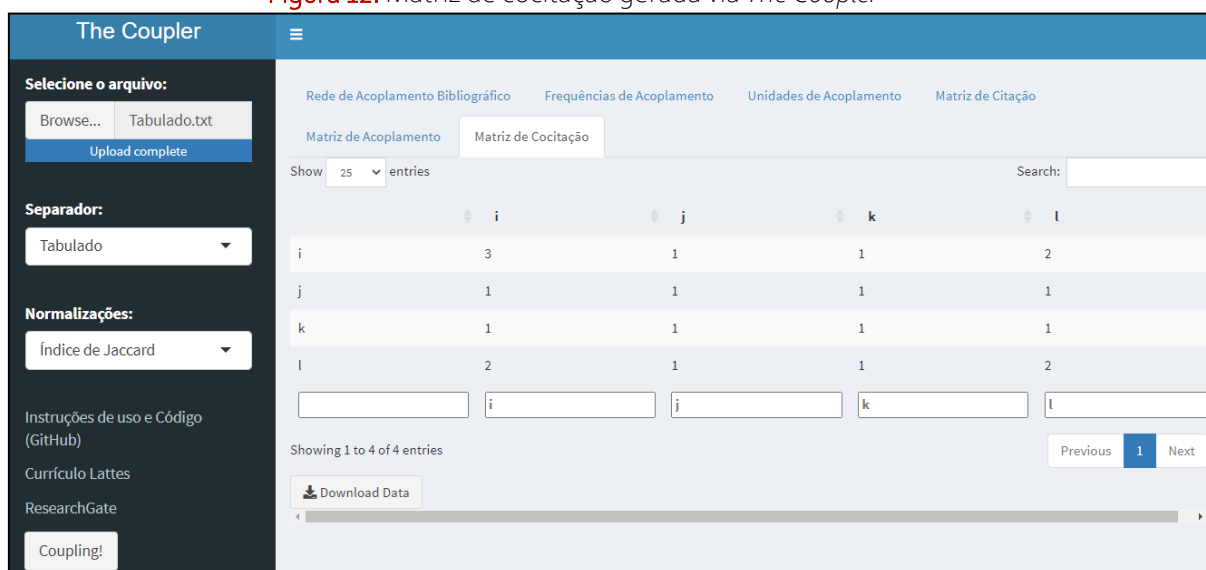


Fonte: Exportação via *The Coupler*

A matriz de acoplamento da Figura 11 gerada pelo *The Coupler* é composta por zeros em sua diagonal principal. Essencialmente, o uso de zeros é bem aceito. Notavelmente estes valores são diferentes dos encontrados na Equação 2, em que o produto entre as matrizes de citação e sua transposta retorna o valor do produto entre vetor A, B ou C consigo mesmo. Contudo, isso não afeta o resultado do acoplamento bibliográfico.

Como resultado final, tem-se a matriz de cocitação entre as unidades citadas i, j, k e l . Esta matriz este presente na Figura 12 e explicita a relação matricial entre a matriz de citação transposta e a própria matriz de citação ($C^T \times C$).

Figura 12. Matriz de cocitação gerada via *The Coupler*

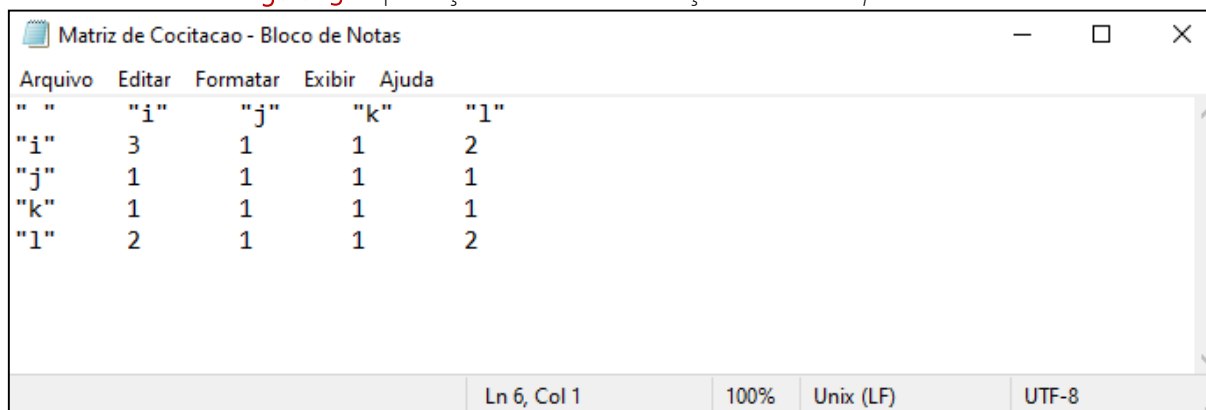


Fonte: Elaboração via *The Coupler*

A matriz da Figura 12 é similar a encontrada na Equação 3. Em sua essência, uma matriz de cocitação quantifica a frequência com que duas unidades foram citadas concomitantemente. A diagonal principal de uma matriz representa a frequência de citação recebida por cada unidade. Assim, os valores da diagonal principal 3, 1, 1 e 2, indicam que as unidades *i*, *j*, *k*, e *l* foram citadas respectivamente 3, 1, 1 e 2 vezes. Análoga as demais análises, é possível exportar a matriz de citação em extensão *.txt* como apresenta a Figura 13.

| 15

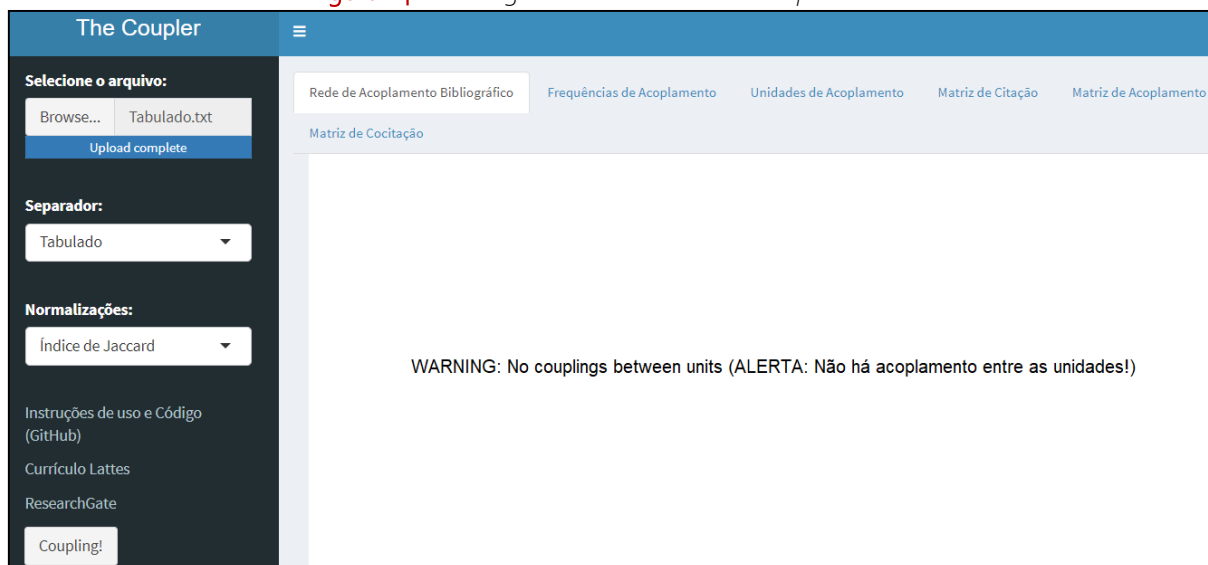
Figura 13. Exportação da matriz de citação via *The Coupler*



Fonte: Exportação via *The Coupler*

A exportação da matriz de cocitação pode ser considerada o último resultado fornecido pelo *The Coupler*. Assim, a partir do conjunto de dados obtidos por meio da ferramenta, o usuário pode de dar sequência a suas análises. Como última demonstração, caso um conjunto de unidades a ser analisada não apresente nenhum acoplamento entre si, o aplicativo retornará a mensagem de alerta: “WARNING: No couplings between units (ALERTA: Não há acoplamento entre as unidades!)” como apresenta a Figura 14.

Figura 14. Mensagem de alerta do *The Coupler*



Fonte: Elaboração via *The Coupler*

Caso nenhuma das unidades de análise sejam acopladas entre si, nenhuma das seis análises (Figuras 3, 4, 6, 8, 10, 12) será processada. Desta maneira, para que sejam processados os cálculos de frequência de acoplamento, as identificações das unidades de acoplamento além das matrizes de citação, acoplamento e cocitação, ao menos duas unidades devem estar acopladas entre si.

Caso haja necessidade, o usuário pode utilizar os dados exportados, oriundos das matrizes de citação, acoplamento e cocitação, em softwares como o Ucinet para construção das redes de citação, acoplamento e cocitação. Ressalta-se que, ao exportar os dados em formato tabulado, a utilização destes dados em planilhas eletrônicas como Microsoft Excel ou Google Planilhas é facilitada.

| 16

5 DADOS BIBLIOMÉTRICOS, PATENTOMÉTRICOS, ALTMÉTRICOS E LINGUAGEM NATURAL

A ideia central do *The Coupler* é promover análises de proximidades entre conjuntos quaisquer e independente de seu foco bibliométrico, permite qualquer tipo de análise, seja no âmbito dos EMI ou não. Assim, como demonstração final da ferramenta, foram processados dados bibliométricos, patentométricos, alométricos e linguagem natural. Após o processamento, foram exibidas as frequências de acoplamento (similaridades) e as unidades de acoplamento entre cada par analisado.

Figura 15. Processamentos via *The Coupler* com diferentes tipos unidades de análise

a) Dados bibliográficos	X1	X2	refs_X1	refs_X2	Coupling	Saltons_Cosine	Jaccard_Index	
	Doc1	Doc2	33	22	4	0.14845392	0.07843137	
	Doc1	Doc3	33	25	5	0.17407766	0.09433962	
	Doc2	Doc3	22	25	2	0.08528029	0.04444444	
	Unidades de Análise			Unidades de Acoplamento				
	Doc1...Doc2			bornmann, l; haustein, s; thelwall, m; torres-salinas, d				
	Doc1...Doc3			barros, m; gouveia, fc; neylon, c; priem, j; souza, ivp				
Doc2...Doc3			araujo, rf; nascimento, ag					
b) Dados de Patentes	X1	X2	refs_X1	refs_X2	Coupling	Saltons_Cosine	Jaccard_Index	
	CN112450208	CN112229155	16	10	6	0.47434165	0.30000000	
	CN113317206	CN112229155	12	10	0	0.00000000	0.00000000	
	CN113317206	CN112450208	12	16	1	0.07216878	0.03703704	
	Unidades de Análise			Unidades de Acoplamento				
	CN113317206...CN112450208			CN104920228				
CN112450208...CN112229155			CN101185567; CN104236272; CN105267248; CN106267252; CN111623601; US5174042					
c) Dados Altmétricos	X1	X2	refs_X1	refs_X2	Coupling	Saltons_Cosine	Jaccard_Index	
	ACERVO	AtoZ	23	12	1	0.06019293	0.02941176	
	ACERVO	Ciência.da.Informação	23	7	0	0.00000000	0.00000000	
	AtoZ	Ciência.da.Informação	12	7	1	0.10910895	0.05555556	
	Unidades de Análise			Unidades de Acoplamento				
ACERVO...AtoZ			@arq_csouza					
AtoZ...Ciência.da.Informação			@Kleberkafka1					
d) Processamento de Linguagem Natural	X1	X2	refs_X1	refs_X2	Coupling	Saltons_Cosine	Jaccard_Index	
	Doc1	Doc2	24	35	3	0.10350983	0.05357143	
	Doc1	Doc3	24	32	1	0.03608439	0.01818182	
	Doc2	Doc3	35	32	3	0.08964215	0.04687500	
	Unidades de Análise			Unidades de Acoplamento				
	Doc1...Doc2			area; information; use				
Doc1...Doc3			area					
Doc2...Doc3			area; article; twitter					

Fonte: Elaboração via *The Coupler*

Estas quatro análises pouco se diferenciam das exemplificadas anteriormente (Figuras 3 a 13), pois, o *The Coupler* não diferencia as unidades de análise. Esta característica favorece seu uso, uma vez que, ao admitir dados diversos, esta ferramenta torna-se extremamente versátil podendo ser aplicadas em qualquer tipo de análise de similaridade e/ou coocorrência. Exemplifica-se estas análises com a identificação de: coautores em comum entre pesquisadores (Por exemplo, tomando a Figura 1 como base, os pesquisadores seriam A, B e C e os coautores *i*, *j*, *k* e *l*), membros em comum por grupos de pesquisa, co-ocorrência e/ou similaridade de palavras-chaves, referências citadas em comum por planos de ensino de disciplinas, entre outros.

Todos os cálculos presentes na Figura 15 foram checados manualmente, e a ferramenta calculou corretamente os valores de *Coupling*, Cosseno de Salton e Índice de Jaccard, e identificou corretamente todas as unidades de análise e todas unidades de

acoplamento. Este último, funcionalidade de destaque do *The Coupler*. Em análises de publicações de determinada área, instituição ou temática específica, as unidades de acoplamento representam parte fundamental da estrutura intelectual do conjunto de obras analisados. Identificar os elementos responsáveis por acoplar dois itens poderá fornecer meios para identificação das principais influências de um domínio, a partir da recorrência de determinadas unidades nas diversas listas de referências analisadas.

6 CONSIDERAÇÕES FINAIS

Esta pesquisa apresentou e demonstrou o uso do *The Coupler*, nova ferramenta gratuita para análises relacionais com foco no acoplamento bibliográfico (ou análises de similaridade diversas). Primeiramente, fundamentou-se, sob o ponto de vista matemático e bibliométrico, as análises de citação, acoplamento bibliográfico e cocitação, e, a partir destas fundamentações, apresentou-se a nova ferramenta.

O *The Coupler* consiste em uma aplicação *web* capaz de gerar matrizes de citação, acoplamento e cocitação, além de calcular frequências de acoplamento normalizadas via Cosseno de Salton e Índice de Jaccard, construir a rede de acoplamento bibliográfico valorada pelas normalizações e identificar as unidades de acoplamento (elementos responsáveis por conectar duas unidades de análise).

Este último resultado pode ser considerado o item de maior destaque da ferramenta, uma vez que este tipo de identificação não é comum nas demais ferramentas bibliométricas. Em mesma proporção, o cálculo das frequências de acoplamento normalizadas via Cosseno de Salton e Índice de Jaccard também proporciona originalidade própria ao *The Coupler*, uma vez que as normalizações são extremamente úteis para que comparações entre diferentes contextos sejam realizadas.

Ademais, submeteu-se a ferramenta a uma demonstração utilizando dados genéricos (Figuras 3 a 13) bem como a testes utilizando dados reais, comumente utilizados em análise bibliométricas, patentométricas, altométricas e em processamento de linguagem natural. O *The Coupler* respondeu aos testes conforme o esperado, sendo capaz de calcular todas as frequências de acoplamento entre as unidades de análise, de maneira absoluta e normalizada, além de identificar todas as unidades de acoplamento.

Enquanto limitações encontradas, para versão *web-app*, a ferramenta encontrará problemas em processar grandes conjunto de dados, pois, atualmente, sua hospedagem no servidor *shinyapps.io* possui limite de *instance size* de 1 *gigabyte*. Em casos de grandes conjuntos, sugere-se a utilização do *The Coupler* em sua versão *offline* e o ajuste do limite de memória utilizada pelo *software R*. Para ajustes de memória, utiliza-se as funções *memory.size* e *memory.limit*. Em testes simulados, o *The Coupler*, inicializado de maneira *offline* via *R*, concluiu análise de um arquivo contendo 1001 artigos citantes e 10010 itens citados em aproximadamente 50 minutos e 3 segundos. Ou seja, ao processar 1001 itens, a ferramenta concluiu o cálculo de 500500 interações de acoplamento juntamente com seus respectivos valores normalizados, além das matrizes de citação, acoplamento e cocitação. Alerta-se que o custo computacional poderá variar de acordo com o computador utilizado. Este teste foi realizado no mesmo computador mencionado na seção de Metodologia.

Com isso, a ferramenta foi capaz de processar dados diversos e assim, pode ser considerada como instrumento alternativo a outros softwares bibliométricos voltados para análises relacionais. Como próximos passos, espera-se manter o *The Coupler* sempre atualizado, livre de eventuais problemas que possam ocorrer, além de realizar futuras implementações demandadas dos usuários como por exemplo, a aceitação direta de

arquivos exportados das bases de dados *Web Of Science* e *Scopus* e a conversão da ferramenta em arquivo executável. Por fim, convida-se toda comunidade a utilizar o *The Coupler*.

REFERÊNCIAS

CASTANHA, R. G. **THE COUPLER**. [S. l.]: Zenodo, 2022. Disponível em: <https://rafaelcastanha.shinyapps.io/thecoupler/>. DOI: <https://zenodo.org/record/7222122>

DELBIANCO, N. R. A comunicação científica no *Twitter*: um estudo altmétrico com periódicos brasileiros da ciência da informação. Dissertação (Mestrado) - Universidade Estadual Paulista, 2022. Disponível em: <http://hdl.handle.net/11449/235088>. Acesso em: 6 set. 2022.

MARICATO, J. de M.; LOPES, D. M. Altmetrics: complexities, challenges and new forms of measuring and comprehending scientific communication in the social. **Biblios**, Lima, n. 68, 2017. DOI: <https://doi.org/10.5195/biblios.2017.358>

MOREIRA, P. S. C.; GUIMARÃES, A. J. R.; TSUNODA, D. F. Qual ferramenta bibliométrica escolher? um estudo comparativo entre softwares. **P2P e Inovação**, Rio de Janeiro, v. 6, p. 140-158, 2020. DOI: <https://doi.org/10.21721/p2p.2020v6n2.p140-158>

PERISSINI, R. C.; FARIA, L. I. L. A presença de ferramentas analíticas em estudos bibliométricos. *In*: ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO, 21., 2021, Rio de Janeiro. **Anais do [...]**. Rio de Janeiro: ENANCIB, 2021. Disponível em: <https://bit.ly/3CnASvn> Acesso em: 10 set. 2022.

ROCHA, E. S. S.; SILVA, M. R. da. Métricas alternativas de periódicos da Ciência da Informação. **Perspectivas em Ciência da Informação**, Belo horizonte, v. 25, p. 118-139, 2021. DOI: <https://doi.org/10.1590/1981-5344/3740>. Acesso em 10 de setembro de 2022.

GOUVEIA, F. C. Estudos altmétricos no Brasil: uma análise a partir dos currículos da Plataforma Lattes-CNPq. **Transinformação**, Campinas, SP, v. 31, 2019. DOI: <https://doi.org/10.1590/2318-0889201931e190027>. Acesso em: 10 set. 2022.

GRÁCIO, M. C. C.; OLIVEIRA, E. F. T. Indicadores de proximidades em Análise de Cocitação de Autores: um estudo comparativo entre Coeficiente de Correlação de Pearson e Cosseno de Salton. **Informação & Sociedade**, João Pessoa, v. 25, n. 2, 2015. Disponível em: <https://bit.ly/3rO2ZyY> Acesso em: 10 de setembro de 2022.

GRÁCIO, M. C. C. **Análises relacionais de citação para a identificação de domínios científicos**: uma aplicação no campo dos Estudos Métricos da Informação no Brasil. Marília: Editora Oficina Universitária, 2020. DOI: <https://doi.org/10.36311/2020.978-65-86546-12-5>. Acesso em: 12 set. 2022

KESSLER, M. M. Bibliographic coupling between scientific papers. **American documentation**, New York, v. 14, n. 1, p. 10-25, 1963a. DOI: <https://doi.org/10.1002/asi.5090140103>. Acesso em: 13 set. 2022.

KESSLER, M. M. An experimental study of bibliographic coupling between technical papers. **IEEE Transactions on Information Theory**, San Francisco, CA, v. 9, Issue, 1, p. 49-51, Jan.1963b. <https://doi.org/10.1109/TIT.1963.1057800>. Acesso em: 13 set. 2022

MORAL-MUÑOZ, J. A. *et al.* Software tools for conducting bibliometric analysis in science: An up-to-date review. **Profesional de la Información**, Barcelona, v. 29, n. 1, 2020. DOI: <https://doi.org/10.3145/epi.2020.ene.03>. Acesso em: 9 set. 2022.

WHITE, Howard D. Cocited author retrieval *online*: an experiment with the social indicators literature. **Journal of the American Society for Information Science**, New York, v. 32, n. 1, p. 16-21, 1981. DOI: <https://doi.org/10.1002/asi.4630320103>. Acesso em: 4 set. 2022.

WHITE, H. D.; GRIFFITH, B. C. Author cocitation: a literature measure of intellectual structure. **Journal of the American Society for information Science**, New York, v. 32, n. 3, p. 163-171, 1981. DOI: <https://doi.org/10.1002/asi.4630320302>. Acesso em: 16 set. 2022.

SMALL, H. Co-citation in the scientific literature: A new measure of the relationship between two documents. **Journal of the American Society for information Science**, New York, v. 24, n. 4, p. 265-269, 1973. DOI: <https://doi.org/10.1002/asi.4630240406>. Acesso em: 13 set. 2022.

ZHAO, D.; STROTMANN, A. Evolution of research activities and intellectual influences in information science 1996–2005: Introducing author bibliographic-coupling analysis. **Journal of the American Society for Information Science and Technology**, Hoboken NJ, v. 59, n. 13, p. 2070-2086, 2008. DOI: <https://doi.org/10.1002/asi.20910>. Acesso em: 14 set. 2022.

APÊNDICE 1 – Código completo do *The Coupler*

```
#PACOTES

library(dplyr)
library(RVenn)
library(igraph)
library(shiny)
library(shinydashboard)
library(flexdashboard)

#Ajuste de memória

memory.limit(size=56000)

# UI e Server R

ui <- fluidPage(dashboardPage(dashboardHeader(title="The Coupler"), dashboardSidebar(

fileInput("file1", "Selecione o arquivo:", accept = ".txt"),

selectInput("sep", "Separador:", c(" ", Vírgula = ",", Ponto_Vírgula = ";", Tabulado = "\t")),

selectInput("normalization", "Normalizações:", c("sem normalização", "Cosseno de Salton",
"Índice de Jaccard")),

fluidPage(h5(" ")),

fluidPage(tags$a(href="https://github.com/rafaelcastanha/The-Coupler-Shiny-App", "Instruções e
Código (GitHub)")),

fluidPage(h5(" ")),

fluidPage(tags$a(href="https://zenodo.org/record/7130614#.YzdGCnbMLIU", "Arquivos para
testes")),

fluidPage(h5(" ")),

fluidPage(tags$a(href="http://lattes.cnpq.br/4834832439175113", "Currículo Lattes")),

fluidPage(h5(" ")),

fluidPage(tags$a(href="https://www.researchgate.net/profile/Rafael-Gutierrez-Castanha-2",
"ResearchGate")),

fluidPage(h5(" ")),

out = fluidPage(h5("Desenvolvido por:")), out = fluidPage(h5("Rafael Gutierres Castanha")),

fluidPage(h5(" ")),
```

```

out = fluidPage(h5("rafael.castanha@unesp.br")),
actionButton("runmodel", "Coupling!"),

dashboardBody(

tabsetPanel(type="tab",

tabPanel(title="Rede de Acoplamento Bibliográfico", column(textOutput("erro"), width = 12,
plotOutput(outputId = "PlotCoupling", width = "100%", heigh=680))),

tabPanel(title="Frequências de Acoplamento", dataTableOutput(outputId =
"DataFrameCoupling"),downloadButton("dlfreq", "Download Data")),

tabPanel(title="Unidades de Acoplamento",style='overflow-x: scroll', dataTableOutput(outputId =
"DataFrameUnits"),downloadButton("dlunits", "Download Data")),

tabPanel(title="Matriz de Citação", style='overflow-x: scroll', dataTableOutput(outputId =
"DataFrameCit"),downloadButton("dlcit", "Download Data")),

tabPanel(title="Matriz de Acoplamento", style='overflow-x: scroll', dataTableOutput(outputId =
"DataFrameMatrix"),downloadButton("dlaba", "Download Data")),

tabPanel(title="Matriz de Cocitação", style='overflow-x: scroll', dataTableOutput(outputId =
"DataFrameCocit"),downloadButton("dlcocit", "Download Data"))

)))

server <- function(input, output){

observe({

#Arquivo

input$runmodel

if (input$runmodel==0)
return()
else

isolate({

r<-reactive({input$file1})

req(r())

})

corpus<-isolate(read.table(r())$datapath, header = FALSE, sep = input$sep, quote="")

colnames(corpus)<-corpus[1,]
corpus<-corpus[(-1),]
hd<-gsub("\\.$", "",names(corpus))

```

```

colnames(corpus)<-hd

#Corpus para dataframe

corpus<-as.data.frame(corpus)

#remover espaços e vazios
corpus<-corpus
corpus[corpus==" "|corpus==" "]<-NA
empty_columns<-sapply(corpus, function(x) all(is.na(x)|x==""))
corpus<-corpus[!empty_columns]

#Contagem de itens citados por lista

citados<-function(x){return(length(which(!is.na(x))))}
itens_citados<-apply(X=corpus,FUN=citados,MARGIN=c(1,2))
df1<-as.data.frame(itens_citados, header=TRUE)
df2<-colSums(df1)
df2<-as.data.frame(df2, header=TRUE)
df2<-tibble::rownames_to_column(df2, "VALUE")
colnames(df2)[1]<-"units"
colnames(df2)[2]<-"refs"
references<-df2

#Transformação em objeto Venn

corpus_aba<-Venn(corpus)

#Intersecção Pareada: identificação das unidades de acoplamento

ABA<-overlap_pairs(corpus_aba)

#Unidades por acoplamento

unit_aba<-na.omit(stack(ABA))
units_coupling<-unit_aba %>% group_by(ind) %>% summarise(valeus=(paste(values, collapse="
")))
units_final<-as.data.frame(units_coupling)
colnames(units_final)[1]<-"Unidades de Análise"
colnames(units_final)[2]<-"Unidades de Acoplamento"

#Intensidades de ABA

df<-as.data.frame(table(stack(ABA)))
df<-as.data.frame(table(stack(ABA)))

if (nrow(df)==0) {
output$PlotCoupling<-renderPlot({plot(c(0, 1), c(0, 1), ann = F, bty = 'n', type = 'n', xaxt = 'n', yaxt =
'n')
text(x = 0.5, y = 0.5, paste("WARNING: No couplings between units (ALERTA: Não há
acoplamento entre as unidades!)", cex = 1.5, col = "black")
})
}

```

```

stop
}

else {

int_aba<-aggregate(Freq ~ ind, data = df, FUN = sum)
Freq_ABA<-data.frame(do.call("rbind",strsplit(as.character(int_abasind),"...",fixed=TRUE)))
Freq_ABA["Coupling"]<-int_aba$Freq
m2=merge(Freq_ABA,df2,by.x="X2",by.y="units",all.x=TRUE)
m1=merge(m2,df2,by.x="X1",by.y="units",all.x=TRUE)
colnames(m1)[4]<-"refs_X2"
colnames(m1)[5]<-"refs_X1"
Freq_ABA<-m1 %>% select(X1,X2,"refs_X1","refs_X2","Coupling")

#Normalizações

novacoluna<-c("Saltons_Cosine")
Freq_ABA[,novacoluna]<-Freq_ABA$Coupling/sqrt(Freq_ABA$refs_X1*Freq_ABA$refs_X2)
novacoluna_2<-c("Jaccard_Index")
Freq_ABA[,novacoluna_2]<-Freq_ABA$Coupling/(Freq_ABA$refs_X1+Freq_ABA$refs_X2-
Freq_ABA$Coupling)

#Rede de acoplamento bibliográfico

net_list<-filter(Freq_ABA, Coupling>0)
links<-data.frame(source=c(net_list$X1), target=c(net_list$X2))
network_ABA<-graph_from_data_frame(d=links, directed=F)

edge_ABA<-net_list$Coupling
edge_CS<-net_list$Saltons_Cosine
edge_IJ<-net_list$Jaccard_Index

#Matrizes de Adjacência

#Matriz de Acoplamento

mtx<-as_adjacency_matrix(network_ABA)
E(network_ABA)$weight<-net_list$Coupling
mtx_ad<-as_adjacency_matrix(network_ABA, attr="weight")
mtx_adj<-as.data.frame(as.matrix(mtx_ad))
mtx_adj<-tibble::rownames_to_column(mtx_adj, " ")

#Multiplicação de Matrizes

dt<-stack(corpus)
dt2<-table(dt$values[row(dt[-1])], unlist(dt[-1]))
mtx_cit<-t(dt2)
mtx_cocit<-t(mtx_cit) %*% mtx_cit)
mtx_cocit<-as.table(mtx_cocit)
#Matriz Cocitação

mtx_cocit_df<-as.data.frame(mtx_cocit)

```



```

links_cocit<-data.frame(source=c(mtx_cocit_df$Var1), target=c(mtx_cocit_df$Var2))
network_cocit<-graph_from_data_frame(d=links_cocit, directed=T)
E(network_cocit)$weight<-mtx_cocit_df$Freq
mtx_adj_cocit<-as_adjacency_matrix(network_cocit, attr="weight")
mtx_adj_cocit_df<-as.data.frame(as.matrix(mtx_adj_cocit))
mtx_adj_cocit_df<-tibble::rownames_to_column(mtx_adj_cocit_df, "")

```

#Matriz de Citação

```

mtx_cit_df<-as.data.frame(mtx_cit)
links_cit<-data.frame(source=c(mtx_cit_df$Var1), target=c(mtx_cit_df$Var2))
network_cit<-graph_from_data_frame(d=links_cit, directed=T)
E(network_cit)$weight<-mtx_cit_df$Freq
mtx_adj_cit<-as_adjacency_matrix(network_cit, attr="weight")
mtx_adj_cit_df<-as.data.frame(as.matrix(mtx_adj_cit))
l<-length(references$units)
l1=l+1
l2<-length(unique(mtx_cit_df$Var2))+l
mtx_citation<-mtx_adj_cit_df[1:l,l1:l2]
mtx_citation<-tibble::rownames_to_column(mtx_citation, "")

```

```

output$erro<-renderText({

```

```

input$runmodel

```

```

if (nrow(df)!=0) {print(" ")}

```

```

})

```

```

output$PlotCoupling <- renderPlot({

```

```

input$runmodel

```

```

if ("sem normalização" %in% input$normalization)

```

```

plot(network_ABA, layout=layout_as_star, edge.width=c(net_list$Coupling), vertex.size=9,
vertex.color=rgb(0.8,0.6,0.8,0.9), vertex.label.color='black', edge.color='grey', vertex.label.cex=1)

```

```

if ("Cosseno de Salton" %in% input$normalization)

```

```

plot(network_ABA, layout=layout_as_star, edge.width=c(net_list$Saltons_Cosine*10),
vertex.size=9, vertex.color=rgb(0.8,0.6,0.8,0.9), vertex.label.color='black', edge.color='grey',
vertex.label.cex=1)

```

```

if ("Índice de Jaccard" %in% input$normalization)

```

```

plot(network_ABA, layout=layout_as_star, edge.width=c(net_list$Jaccard_Index*10),
vertex.size=9, vertex.color=rgb(0.8,0.6,0.8,0.9), vertex.label.color='black', edge.color='grey',
vertex.label.cex=1)

```

```

})

```

```

output$DataFrameCoupling <- renderDataTable(Freq_ABA)
output$DataFrameUnits <- renderDataTable(units_final)
output$DataFrameCit <- renderDataTable(mtx_citation)
output$DataFrameMatrix <- renderDataTable(mtx_adj)
output$DataFrameCocit <- renderDataTable(mtx_adj_cocit_df)

output$dlfreq <- downloadHandler(
  filename = function(){
    paste("Frequências de Acoplamento", "txt", sep=".")
  },
  content = function(file){
    write.table(Freq_ABA, file, sep="\t", row.names = F, col.names = TRUE)
  })

output$dlunits <- downloadHandler(
  filename = function(){
    paste("Unidades de Acoplamento", "txt", sep=".")
  },
  content = function(file){
    write.table(units_final, file, sep="\t", row.names = F, col.names = TRUE)
  })

output$dlcit <- downloadHandler(
  filename = function(){
    paste("Matriz de Citacao", "txt", sep=".")
  },
  content = function(file){
    write.table(mtx_citation, file, sep="\t", row.names = F, col.names = TRUE)
  })

output$dlaba <- downloadHandler(
  filename = function(){
    paste("Matriz de Acoplamento", "txt", sep=".")
  },
  content = function(file){
    write.table(mtx_adj, file, sep="\t", row.names = F, col.names = TRUE)
  })
output$dlcocit <- downloadHandler(
  filename = function(){
    paste("Matriz de Cocitacao", "txt", sep=".")
  },
  content = function(file){
    write.table(mtx_adj_cocit_df, file, sep="\t", row.names = F, col.names = TRUE)
  })
})

}

shinyApp(ui, server)

```

APÊNDICE 2 – Dados bibliométricos

Quadro 1. Dados bibliométricos processados

Doc1	Doc2	Doc3
ball, a	adie, e	almind, tc
barros, m	alperin, jp	araujo, rf
bjorneborn, l	araujo, rf	barcelos, j
bornmann, l	asnafi, ar	barros, m
brigham tara, j	bornmann, l	borba, vd
cronin, b	bueno, wc	bossy, mj
donato, h	burke, m	costa, bir
dutta, b	fausto, s	dos santos, fb
fenner, m	gasparyan, ay	empinotti, ml
gilbert, gn	haustein, s	fraumann, g
gouveia, fc	holmberg, k	garfield, e
groth, p	ke, q	gouveia, fc
haustein, s	luiz, oc	heberle, h
ingwersen, p	mohammadi, e	levy, p
khan, gf	nascimento, ag	maricato, jd
kumar, s	nassi-calo, l	nascimento, ag
lariviere	ravenscroft, j	neylon, c
lin, j	schmitt, m	priem, j
liu, j	silva m. r., da	reis, je
moed, hf	thelwall, m	schramm, s
neylon, c	torres-salinas, d	souza, ivp
piwowar, h	waltman, l	travieso-rodriguez, c
priem, j		van eck, nj
roemer, rc		vanti, n
rousseau, r		vaughan, l
shema, h		
souza, ivp		
sud, p		
taraborelli, d		
taylor, m		
thelwall, m		
torres-salinas, d		
zahedi, z		

Fonte: Elaboração própria com extração via software VosViewer

APÊNDICE 3. Dados patentométricos

Quadro 2. Dados patentométricos processados

CN113317206	CN112450208	CN112229155
CN102210266	CN101796925	CN1217280
CN104604676	CN104920216	CN1268903
CN104920228	CN104920228	CN101185567
CN110195077	CN108513909	CN104236272
CN112616663	CN112568125	CN105267248
US5403736	CN101796925	CN106267252
CN102210266	CN104920216	CN111623601
CN104604676	CN104920228	US5174042
CN104920228	CN108513909	CN1217280
CN110195077	CN112568125	CN1268903
CN112616663	CN101185567	
US5403736	CN104236272	
	CN105267248	
	CN106267252	
	CN111623601	
	US5174042	

Fonte: Elaboração própria

APÊNDICE 4 – Processamento de linguagem natural

Quadro 3. dados para o processamento de linguagem natural

Doc1	Doc2	Doc3
altmetria	altmetric indicator	altmetric
area	altmetric study	altmetrica
information	altmetric tool	altmetrico
institutionalization	Altmetrics	applied social science
international database	Área	area
metric	Article	article
metric study	Attention	brazil
order	capas	communication
own understanding	comparative analysis	curricula
possibility	complementary resource	data
production	contribution	delimitation
reflection	evaluation	doctor
reflexive analyze	exploratory type	doctorate degree
relation	favorable context	emphasis
scientometric	impact	end
set	information	event
subject	information science	evolution
subject altmetria	information science journal	existence
text	institutional site	expertise
theoretical approach	interest	exploratory bibliometric study
theoretical foundation	journal	facebook
uncertainty	newspaper	field
use	scientific communication	geographical distribution
webometric	scientific dissemination	hand
	significant number	history
	significant way	identification
	social media profile	Information metrics study
	social network	large area
	social web	researcher
	source	Twitter
	study	web
	Twitter	year
	use	
	variable	
	view	

Fonte: Elaboração própria com extração via software VosViewer