Article

# RDBCI |
Revista Digital de Biblioteconomia e Ciência da Informação
Digital Journal of Library and Information Science

Check for updates

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

Rafael Gutierres Castanha [1] iD

ABSTRACT

Introduction: The use of programming languages in the context of Metric Information Studies has gained ground among the scientific community in the area as they are practical, free or have low computational cost. Objectives: To present a new, free and alternative bibliometric tool, aimed at relational citation analysis, focusing on bibliographic coupling, built on the R programming language, entitled *The Coupler*. Methodology: We ground the relational analysis of citation, co-citation and bibliographic coupling from a mathematical perspective and present the source code of the tool and the conditions under which it was built with its access granted for all. We use generic data and expose all its features to test the tool; and to demonstrate the use with real data, we operate bibliometric, patentometric, altmetric and natural language data. Results: The tool builds citation, coupling and co-citation matrices, in addition to calculating normalized values via Salton's Cosine and Jaccard Index of the bibliographic coupling frequencies. Furthermore, *The Coupler* builds bibliographic coupling networks and identifies the coupling units responsible for each coupling pair; the latter feature, a distinct result for traditional bibliometric software. Conclusion: The paper concludes that the tool behaves as expected and satisfactorily when processing both generic data and bibliometric, patentometric, altmetric and natural language data. Among the results, especially and differentiating from other software, we highlight the identification of coupling units, and calculations of normalized coupling frequencies via Salton's Cosine and Jaccard Index.

KEYWORDS

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

| 1

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

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,

Author's Correspondence

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

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, altmétricos 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, altmétricos 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.

## CRediT

| 2

**Article submitted to the similarity system**

# 1 INTRODUCTION

In the field of metric information studies (MIS) several software programs for bibliometric analysis are found, such as: 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).

Other software applications, such as Pajek, Ucinet and Gephi, focus on the construction of graphs (networks) but which are not necessarily designed at the core of MIS. However, they are highly disseminated in the specialized literature in the area because they are free software and have user-friendly interfaces.

Moreover, the use of programming languages in the context of Applied Social Sciences and consequently in MIS has also been gaining ground among the scientific community in the area. This phenomenon is noted within the R programming language, with the existence of packages and libraries aimed at bibliometric analysis, such as, for example, *scholar packages, wosr package, package rScopus*, which perform bibliometric analysis by extracting data from Google Scholar, Web Of Science and Scopus database, respectively.

In this sense, the Bibliometrix/Biblioshiny software, also built on the R programming language, enables various bibliometric analyses such as analyses of scientific production by author, impact analysis, Lotka and Bradford laws, relational analysis of citation, co-citation and bibliographic coupling, factor analysis, among other diverse features. Bibliometrix uses the shiny package to generate its graphical interface, built in html, which allows use through the user's internet browser.

Given the importance and relevance the Bibliometrix has achieved, analyses using the R language are increasingly included in bibliometric studies. The possibility of working with a high volume of data combined with the convenience of the language, the free service and a relatively low computational cost are some of the reasons for this expansion. The volume of data suggests the recurring use of programming languages as they are capable of processing large volume of data.

In this context, this research aims to present a new, free and alternative, bibliometric tool, aimed at relational citation analysis, focusing on bibliographic coupling, built on the R programming language, entitled *The Coupler*. Specifically, testing the tool using various data commonly used in bibliometric analysis. For this, the motivations for building the tool will be presented as well as the theoretical bases of the analyses that support the construction of the code behind the software.

This new tool focuses on relational citation analyses with special attention to bibliographic coupling, and can be understood, in essence, as a similarity analysis tool between any type of analysis unit. The conception of new computational resources, mainly focused on the field of MIS, is a fundamental part of the progress of the field as a whole, and strengthens the development of new studies, directly contributing to the scientific and technological evolution of bibliometrics.

# 2 RELATIONAL CITATION ANALYSIS

Within the scope of bibliometrics, there is a range of indicators that vary, essentially, from the elements they refer to, however, indicators aimed at citation analysis can be classified, according to Grácio (2020), in univariate and relational.

Univariate citation analysis refers to the impact and measurement of the scientific community recognition based on the citations received by a certain set of documents, authors, institutions, or any other unit of analysis. In this regard, there are, for example, mean citation rates per year per article, the Impact Factor, or CiteScore. Regarding the second item, it seeks to establish groupings between different units of analysis. At this point, bibliographic coupling and co-citation analyses are found.

Regarding bibliographic coupling analysis (BC), the method was proposed by Kesller (1962; 1963) to identify how articles are grouped together through the sharing of cited references. That is, if two articles share at least one reference (they cite the same reference), these articles are bibliographically linked. Kesller (1963) evaluated BC according to two types of groupings between articles called GA and GB: GA) Given a set of articles, they are all coupled together (they cite at least one reference in common); GB) all articles are coupled to a certain article.

The number of cited references in common by two articles was called bibliographic coupling strength (or frequency of bibliographic coupling) and the references responsible for coupling these two articles are defined as coupling units. In this way, BC is configured as a relationship between citing articles (citing-citing)

Years later, Small (1973) coined the term co-citation analysis (CA) based on the co-occurrence of two documents cited together in at least one list of references, that is, CA measures in how many documents other two documents are co-cited (cited concomitantly). Thus, this analysis reveals the relationship between cited articles (cited-cited). In summary, while BC quantifies the similarity of references between two documents, CA exposes the recurrence in which two references are cited jointly. This quantification can be understood as the co-citation frequency between two articles.

Bibliographic coupling and co-citation relational analyses have evolved over time so that White (1981) and White Griffith (1981) proposed author co-citation analysis (ACA). In this perspective, ACA verifies in how many articles two authors were cited simultaneously. For this, the paper of the cited author is considered unique and no longer the reference of the cited article.

In 2008, Zhao and Strotmann (2008) enunciated author bibliographic coupling (ABA). This new analysis intends to couple two researchers from two perspectives: using cited documents or cited authors. In the first case, the number of common references two researchers share is computed. In the second one, the cited author's paper is considered unique and the number of authors cited in common by two researchers is calculated.

Since the first studies proposed by Kesller (1962; 1963), relational analyses have gained ground in the field of bibliometrics and various methods of normalization of coupling frequencies, such as the Jaccard Index (JI) and the Salton's Cosine (SC), have been debated since Sen and Gan (1983). These methods can be described in Equations 1 and 2 as:

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

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

Where A and B are any two analysis units to be coupled and $Coupling\ (A,B)$ represents how many coupling units A and B share (they cite in common).

Traditionally, relational citation analyses are basically the result of matrix interactions conceived from the citation adjacency matrix of the analyzed situation. That is, given a citation matrix $C_{n \times m}$, the matrix containing the bibliographic coupling frequencies ($F_{ab}$) is given by: $F_{ab} = C \times C^T$, while the matrix containing the co-citation frequencies ($F_c$) is given by $F_c = C^T \times C$. This is exemplified in Equation 3:

$C =$

|   | $i$ | $j$ | $k$ | $l$ |
|---|---|---|---|---|
| **A** | 1 | 1 | 1 | 1 |
| **B** | 1 | 0 | 0 | 1 |
| **C** | 1 | 0 | 0 | 0 |

$C \times C^T$

|   | **A** | **B** | **C** |
|---|---|---|---|
| **A** | 4 | 2 | 1 |
| **B** | 2 | 2 | 1 |
| **C** | 1 | 1 | 1 |

$C^T \times C =$

|   | $i$ | $j$ | $k$ | $l$ |
|---|---|---|---|---|
| $i$ | 3 | 1 | 1 | 2 |
| $j$ | 1 | 1 | 1 | 1 |
| $k$ | 1 | 1 | 1 | 1 |
| $l$ | 2 | 1 | 1 | 2 |

[3]

Where A, B and C are citing units and *i, j, k, l* are cited units. Furthermore, the second matrix corresponds to the bibliographic coupling matrix, and the third matrix corresponds to the co-citation matrix. To illustrate, units A and B have coupling strength equal to 2 (coupling units are *i* and *l*); units *i* and *l* have a co-citation frequency equal to 2 (they were simultaneously cited by A and B). As for the main diagonal of the coupling and co-citation matrices, these elements represent a reflexive connection of a certain element with itself. In co-citation matrices, this reflexive relationship represents the number of elements that cited it. For example, item *i* was cited by A, B and C, so the main diagonal element related to *i* is equal to 3. In coupling matrices, these values are little representative. In this sense, Grácio and Oliveira (2015) warn about this situation and suggest: electing the highest frequency of the element with the others; adopting the mean frequency of each element as the diagonal value; recording zeros diagonally; or defining the diagonal as a set of missing values. Among them, the last alternative has greater acceptance and use in the research community, as it is more easily executed and has less conceptual bias.

It is even possible to observe another result, the cardinality of the list of cited units of A is equal to four (*i,j,k,l*), of B is equal to 2 (*i, l*), and of C is equal to 1 (*i*). From this information, it is possible to normalize the coupling frequency values via Salton's Cosine and/or Jaccard Index, as shown in Equation 4, respectively.

$Coupling_{SC} =$

|   | **A** | **B** | **C** |
|---|---|---|---|
| **A** | 1 | 0,71 | 0,50 |
| **B** | 0,71 | 1 | 0,71 |
| **C** | 0,50 | 0,71 | 1 |

$Coupling_{JI} =$

|   | **A** | **B** | **C** |
|---|---|---|---|
| **A** | 1 | 0,50 | 0,25 |
| **B** | 0,50 | 1 | 0,50 |
| **C** | 0,25 | 0,50 | 1 |

[4]

Equations 3 and 4 explain a consequence of the matrix calculation between the citation $C$ matrix and its transpose: the coupling matrices (normalized or not) and the co-citation matrix must be square and symmetric in relation to the diagonal. That is, the elements of the upper triangular part are the same as those of the lower triangular part. This result infers that both AB (citing-citing relationship) and the CA (cited-cited relationship) do not represent directed relationships, such as the citation matrix (citing-cited relationship).

| 5

## 3 METHODOLOGY

Three steps were organized to efficiently present *The Coupler*. The first step describes the information regarding its computational aspects such as its construction, hosting and features. In this step, the tool's design, its graphical interface, its code, and how the user can run *The Coupler* online or offline are presented, in addition to information about the file formats compatible with the tool (Figures 1 and 2).

In the second step, all of *The Coupler*'s features are demonstrated using the data present in Equations 3 and 4. All processing possibilities supported by the tool are demonstrated, as well as data export (Figures 3 to 14).

In the third step, tests are performed using bibliometric, patentometric, altmetric and natural language data. The bibliometric data used in the tests were obtained after searching for the term "altmetrics" in all fields in the Web Of Science database, generating three results: Doc.1) Maricato and Martins (2017); Doc.2) Rocha and Silva (2020); Doc.3) Gouveia (2019). For this, the three documents were coupled using the authors cited in each document as coupling units (Figure 15a). To extract the cited authors, the *export selected authors* function of the VosViewer software was used (Appendix 2).

For processing the patentometric data, three patents, present in the Derwent Innovation Index database, with codes: CN113317206 (2021), CN112450208 (2021) and CN112229155 (2021) were combined. In this case, the three patents were coupled via the cited patents (Appendix 3). This result is shown in Figure 15b.

As for altmetric data, data from Delbianco (2022, p. 114) were used, referring to Twitter followers of three scientific journal profiles in the area of Information Science: 1) *Acervo: Revista do Arquivo Nacional; 2) AtoZ: novas práticas em informação e conhecimento; 3) Ciência da Informação em Revista*. In this way, the three profiles were coupled via common followers (the number of followers in common was verified). Coupling using followers as the coupling unit is shown in Figure 15c.

As a last analysis, a natural language processing was performed comparing the words (in common) present in the abstracts of the three aforementioned articles (Doc.1: Maricato and Martins (2017); Doc.2: Rocha and Silva (2020); Doc. 3: Gouveia (2019)). To extract the words, the *export selected terms* function of the VosViewer software was used. By default, the software suggests using the most relevant 60% of the processed abstract words. For this analysis, the pattern suggested by the software was used and the abstracts were coupled via words in common (Appendix 4). This analysis is shown in Figure 15d.

All tests in the subsequent sections were performed using both the online and offline versions of *The Coupler*. The computer used in all operations has a Windows 10 operating system, Intel Core (TM) i7-8550UCPU@1.80GHz-1.99 GHz processor and 8 Gigabytes of RAM memory.

## 4 THE COUPLER

*The Coupler* was developed based on the theoretical bases of relational citation analysis. A web application (web-app) focused on relational citation analysis, focused on coupling analyses. The tool is capable of building citation, coupling and co-citation matrices. In addition to this matrix perspective, the tool calculates normalized values via Salton's Cosine and Jaccard index of the bibliographic coupling frequencies admitting any type of analysis unit and any type of coupling unit. Furthermore, it builds the bibliographic coupling network and identifies the coupling units responsible for each

| 6

coupling pair. The latter feature is a distinct result for traditional bibliometric software. The source code is presented in Appendix 1, and can be executed via the R programming language (by pasting the code into R) or accessed via the website https://rafaelcastanha.shinyapps.io/thecoupler. For hosting the web-app, the private server of shinyapps.io by R-Studio was used, as it allows the implementation of shiny applications on the web environment.

At the time this paper was submitted, *The Coupler* was in the process of registration in the National Institute of Industrial Property (INPI) through the Unesp Innovation Agency (AUIN). The required record is a computer program record, based on the Copyright Law (Law No. 9.610/1998) and the Law of Software (Law No. 9.609/1998). Consequently, after registration, the author guarantees maintenance and assistance (to users) of *The Coupler*.

The availability of the tool as an online application facilitates dissemination and access from different devices and operating systems such as Linux, Mac-OS and Windows. *The Coupler* is also accessed via mobile devices such as tablets and cell phones with different systems such as Android or ios; in addition to different browsers such as Google Chrome, Microsoft Edge, Safari, among others. In this way, the app does not require prior installation of the R programming language, only the file to be processed.

The resource, both online and offline (accessed via R) only supports .txt files organized in columns with a header that is tabbed, separated by commas or semicolons. Taking as an example the units A, B and C from Equation 2, Figure 1 presents these three types of organization. Thus, *The Coupler* requires prior, and possibly manual, organization of the processed files. Some software, such as VosViewer for example, provide the option of extracting cited units (authors, articles or journals) through its features *export selected authors*, *export selected cited references* or *export selected sources*.

The manual organization of the mentioned items can be understood as a pre-processing step of *The Coupler*. This step can be very useful for papers that are not indexed in databases capable of automatically extracting cited references (such as Web Of Science, Scopus and Dimensions), for non-digitized papers in which the extraction must be done manually, and for non-bibliographic data, in which organization or collection would be carried out manually.

Unlike software such as VosViewer and Bibliometrix, which process data coming directly from databases in .csv, .xlms or .ris formats, *The Coupler* has this feature, however, the tool is capable of processing any type of unit , such as lists of: DOI (Digital Object Identifier), ORCID, researchers who make up research groups or departments, research areas of interest, researchers present in departments, references suggested in course teaching plans, followers on social networks, among others. In this way, the other tools mentioned above can be complementary, and not mutually exclusive.

Figure 1. Types of file organization



a) Tabbed File          b) Separation by comma          c) separation by semicolon
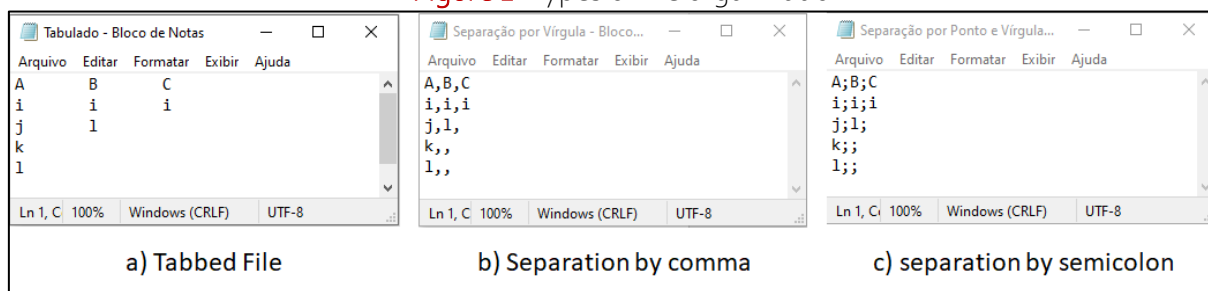
Source: by the author

The option for the .txt extension is justified as the extension is light and easy to access on several devices. In this way, the only requirement for using *The Coupler* is the existence of a previously organized file. Figure 2 illustrates the graphical interface of the tool.

Figure 2. *The Coupler* graphical interface
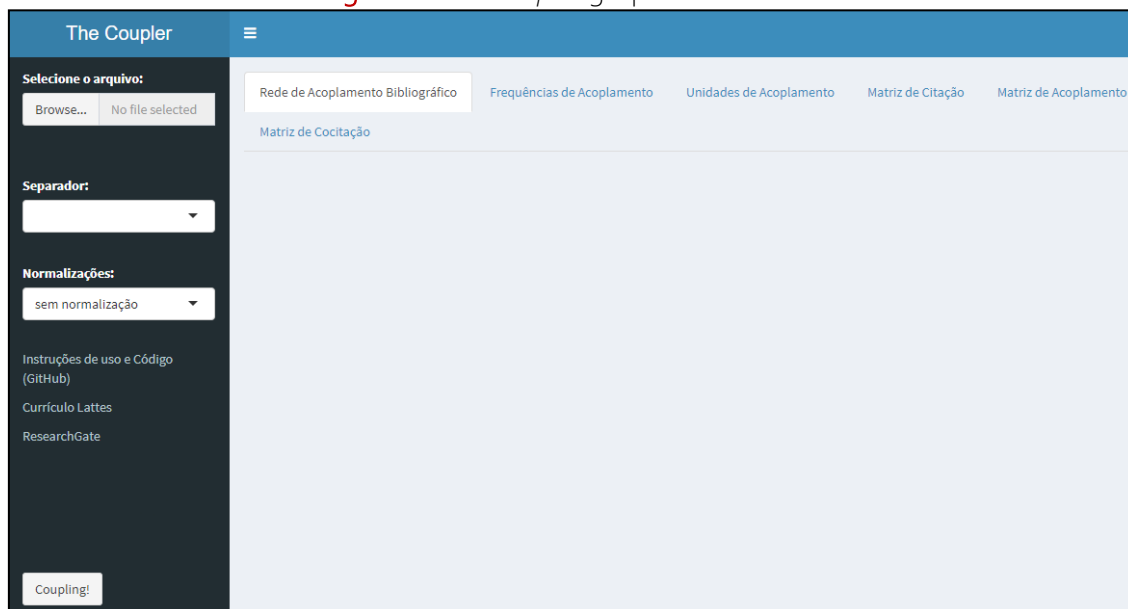
When inserting the tabbed file in "Select the file", the user must choose the type of separator in the file: separated by comma or semicolon. After that, when clicking on "Coupling!" the tool will return six results, which are shown in tab format in Figure 2: i) the bibliographic coupling network; ii) the coupling frequencies between each pair of units and their normalizations via Salton's Cosine and Jaccard Index (Equation 1) organized in table format; iii) the coupling units responsible for coupling each pair of units, organized in table format; iv) the citation matrix; v) the coupling matrix between the citing units; vi) the co-citation matrix between the cited units.

Given these six results, the user will be able to export all the obtained results, as the tool allows saving the image from the bibliographic coupling network and offers the option to download the table containing the bibliographic coupling frequencies between each analyzed unit, the table containing all the coupling units and the citation, coupling and co-citation matrices. Except for the image, all files are saved in tabulated .txt formats and organized in columns.

In addition, the user can choose to view the network from three perspectives by clicking on "Normalizations": i) without normalization: user will return the coupling network valued by the absolute values of the coupling frequencies; ii) Salton's Cosine: user will return the coupling network valued by the normalized values of the coupling frequencies via Salton's Cosine; iii) Jaccard Index: user will return the coupling network valued by the normalized values of the coupling frequencies via the Jaccard Index.

As much as the central idea of the tool is bibliometric analysis and, consequently, the use of bibliographic units such as authors or articles, the tool is not capable of distinguishing whether the input data deal with bibliographic elements or not, bringing greater diversity to its use. Thus, similarity (coupling) or co-occurrence (co-citation) analyses can be extended to any analysis units, such as altmetric and patentometric units, or even, any type of analysis of intersection (similarity) between sets.

| 8

## 4 DEMONSTRATION OF ANALYSIS VIA THE COUPLER

To present The Coupler's analyses and tests, the file shown in Figure 1 was initially used to demonstrate how it works. After this test, bibliometric, patentometric, altmetric and natural language data were used. In this way, the files present in Figure 1 can be understood as any analysis unit (authors, documents, institutions, among others). For the first test, Figure 3 (a, b, c) shows the coupling network and the normalization possibilities through "without normalization", Salton's Cosine or Jaccard Index.

When choosing between the three options, the edges vary in thickness, therefore the thicker the greater the proximity between the analyzed units, or even, the thicker the edges, the greater the frequency (strength) of bibliographic coupling.

| 9

Figure 3. Bibliographic coupling network generated via The Coupler using: a) without normalization; b) Salton's cosine; c) Jaccard index
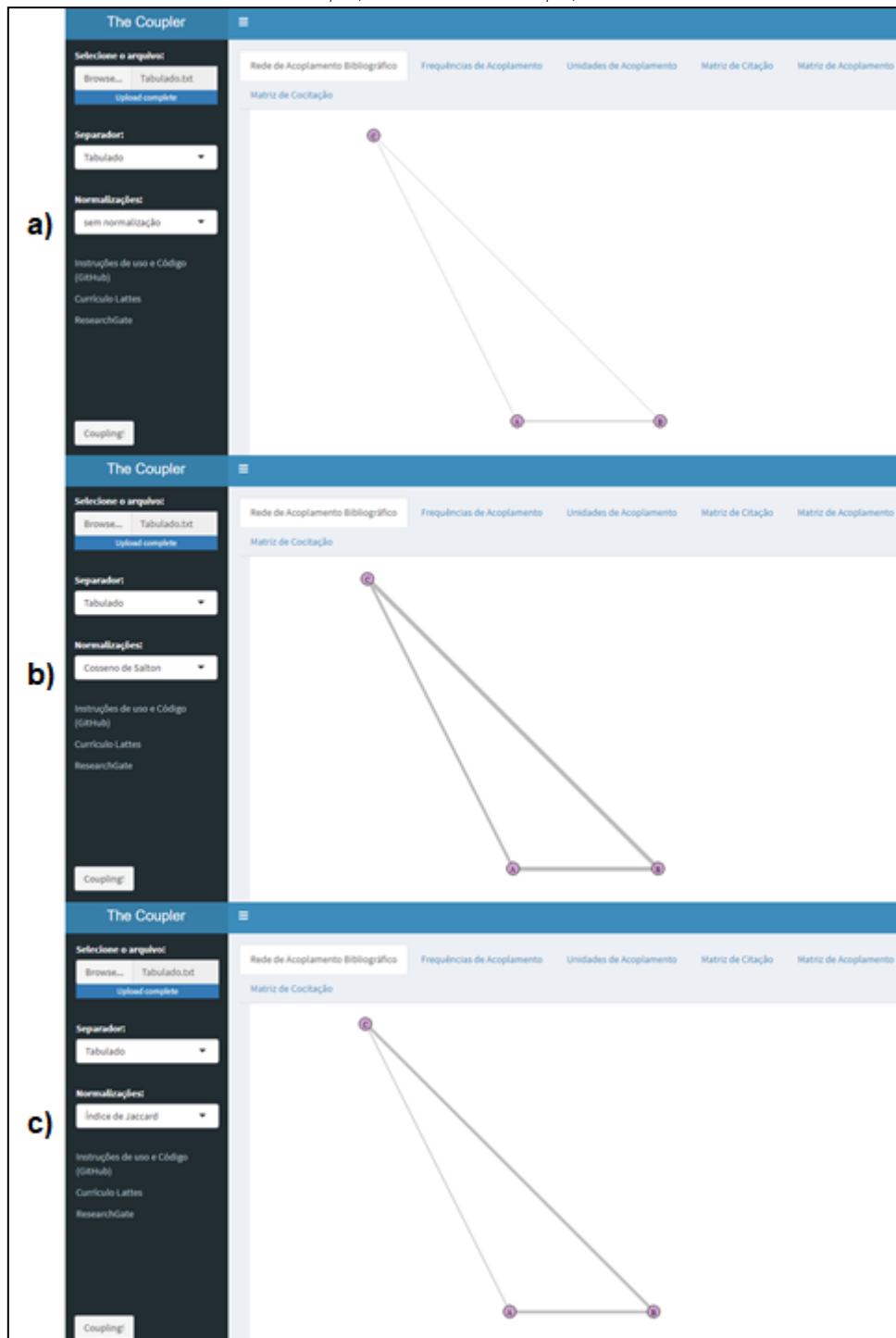


Source: Elaborated via The Coupler

The visualization of the bibliographic coupling network uses the igraph graph library present in R. It is known that several software programs are capable of generating different networks, and in this context, the focus of *The Coupler* is not exactly the visualization of the graph, since different software can do this, but rather the relational aspects present behind the representation and found in the tabs underlying the "Bibliographic Coupling Network". Thus, Figure 4 presents the results found in "Coupling Frequencies".

Figure 4. Coupling Frequencies generated via The Coupler
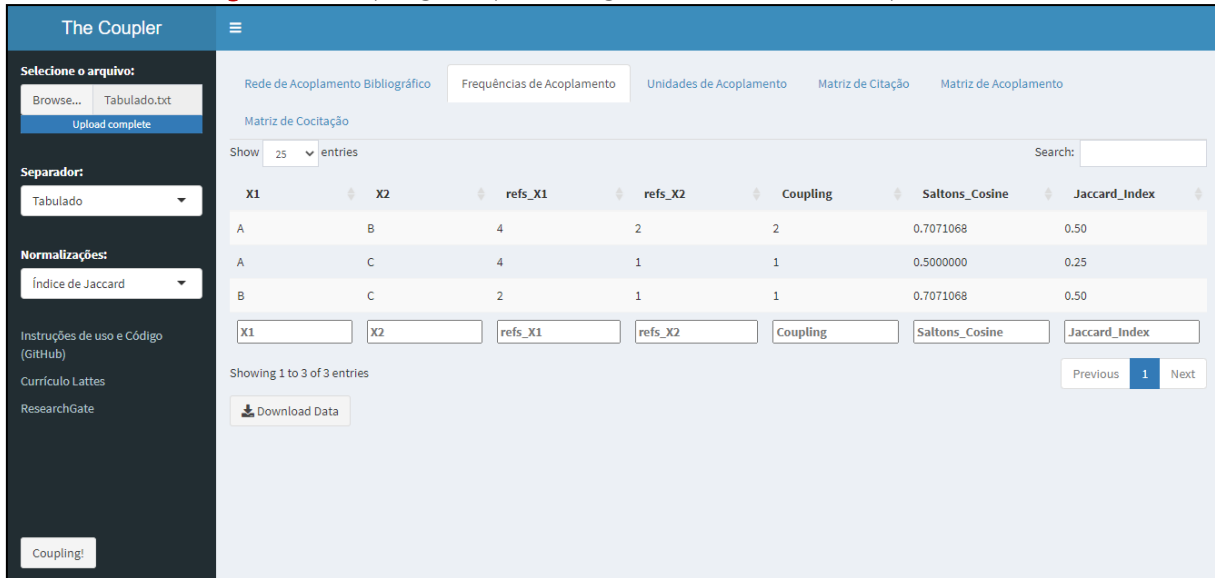


Source: Elaborated via The Coupler

Figure 4 presents the elements: i) "X1" and "X2": represent the units of analysis to be compared (coupled); ii) "refs_X1" and "refs_X2": cardinality of the items cited by "X1" and "X2", respectively; iii) "Coupling": items cited in common by X1 and X2, respectively; iv) Saltons_Cosine: Coupling values normalized via Salton's Cosine; v) Jaccard_Index: Coupling values normalized via Jaccard Index.

As an example, the first line reads: unit A has a list of references made up of 4 elements, unit B has a list of references made up of 2 elements, and both cited 2 elements in common. This value normalized via Salton's Cosine is equal to 0.7071068, and normalized via Jaccard Index is equal to 0.5.

In this tab, the "Download Data" button export these results. By exporting the data, a .txt file named "Coupling Frequencies.txt" is generated, as shown in Figure 5.

Figure 5. Export of results in "Coupling Frequencies"
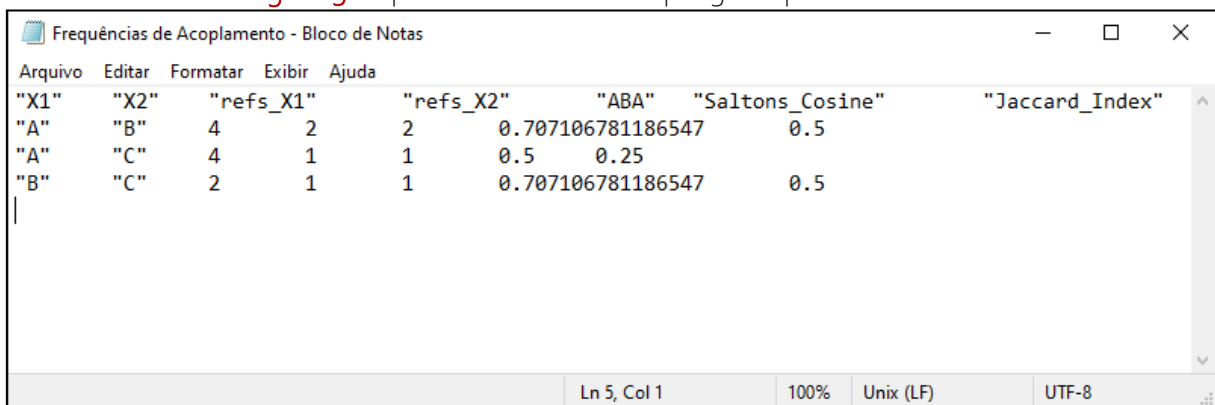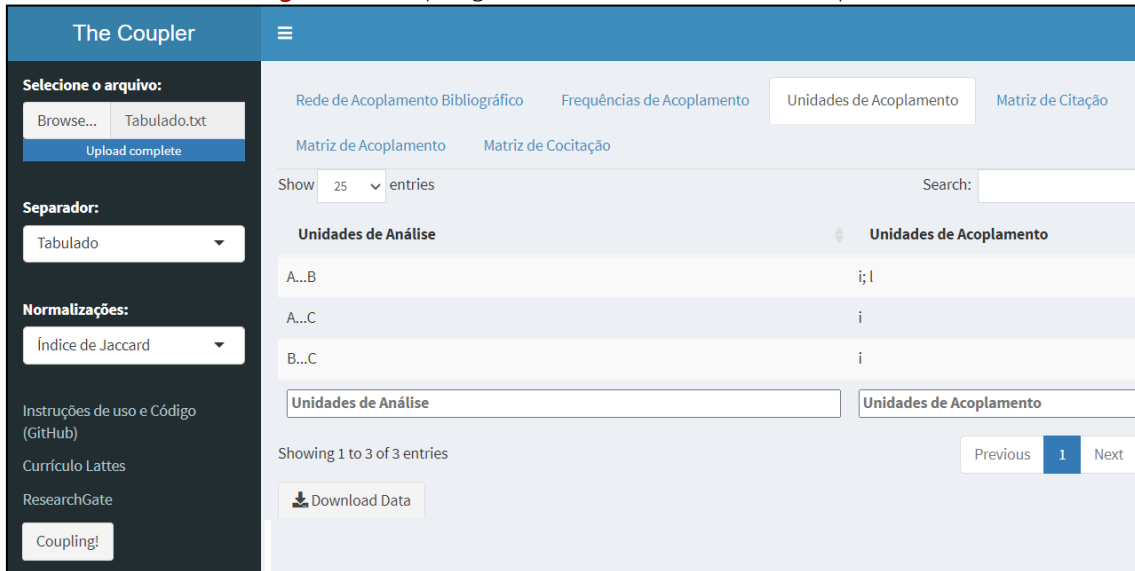


Source: Elaborated via The Coupler

The data in Figure 5 are tabulated and organized into columns. This feature facilitates its use in spreadsheets. In addition, it is possible to use the "Search" field to search for units of analysis the user wants to focus on in addition to the possibility of navigating through the pages. By default, the tool supports the initial 25 results.

| 11

Moreover, the code behind *The Coupler* excludes null columns, that is, in case a unit has not cited any reference, this unit will be excluded from the analysis, since it will not couple with any other unit. This case is analogous to articles that do not have references, that is, in which the author of the article does not cite anyone in his/her study.

Following, the "Coupling Units" tab brings the most relevant and unprecedented result among all those present in the tool, which is the identification of the coupling units, that is, the elements in common for each pair of analyzed units. Figure 6 illustrates the identification of all coupling units.

Figure 6. Coupling units identified via The Coupler



Source: Elaborated via The Coupler

Figure 6 identifies the elements responsible for coupling between units A, B and C. Thus, the first line is an example: Units A and B were coupled by elements *i* and *l*. That is, *i* and *l* are coupling units. Analogously to the previous example, user can export the results by clicking on "Download Data". The results are also organized in columns and tabulated as shown in Figure 7.

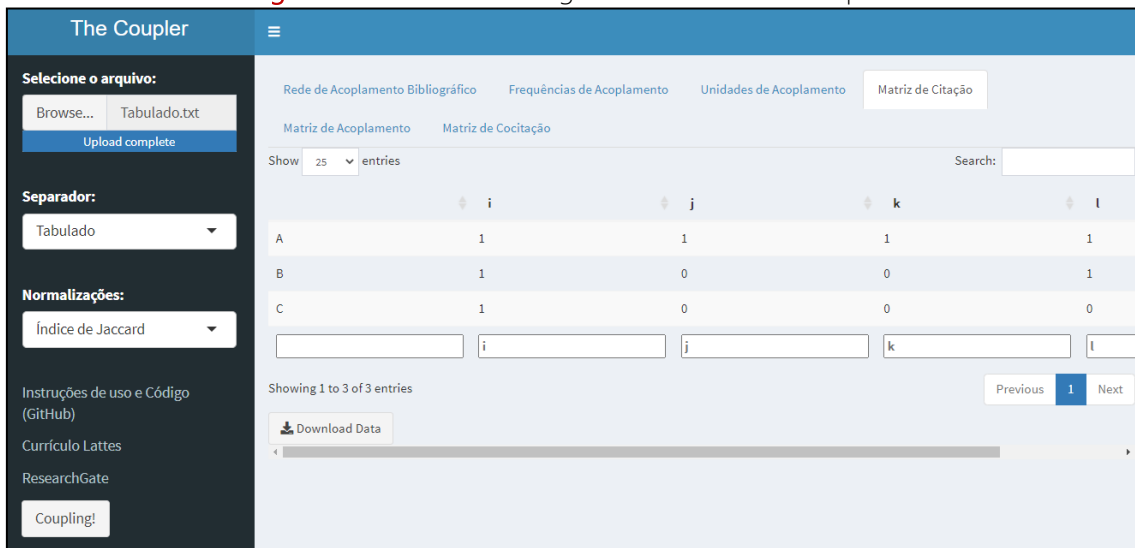Figure 7. Export coupling units via The Coupler
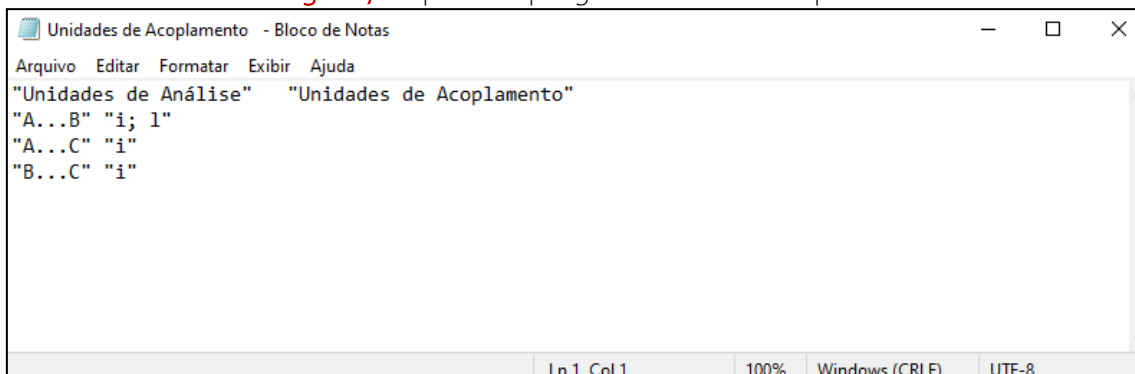


Source: Export via The Coupler

The next displayed tab is the "Citation Matrix". This tab displays the matrix composed of citing and cited units, similar to matrix $C$ in Equation 3. That is, the citation

adjacency matrix that will give rise to the analyses of bibliographic coupling and co-citation is displayed. This matrix is shown in Figure 8.

Figure 8. Citation Matrix generated via The Coupler

The citation adjacency matrix in Figure 8 represents the asymmetric citation relationship between units A, B and C and elements *i, j, k, l*. Similarly to the previous examples, this matrix can be exported, as shown in Figure 9.

Figure 9. Citation matrix export via The Coupler
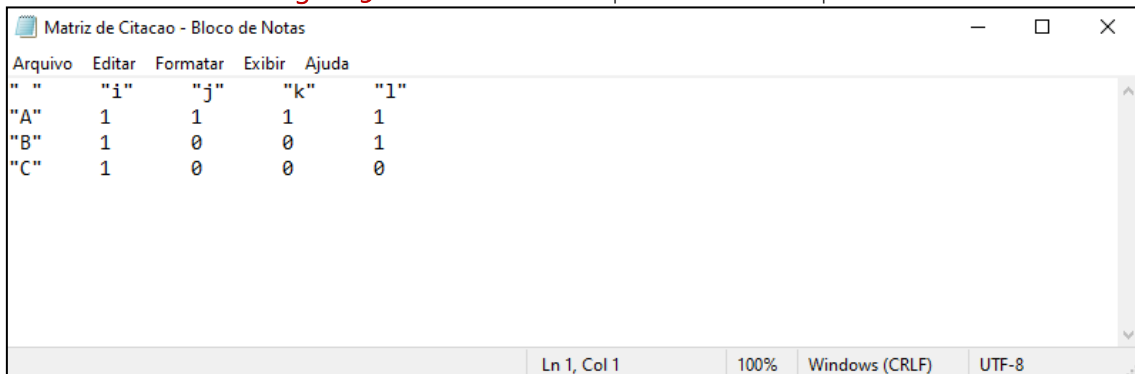
From the citation matrix, the coupling and co-citation matrices can be obtained through the matrix operations between the citation $C$ matrix and its transpose $C^T$. Thus, on the next tab, the "Coupling Matrix" which represents the $C \times C^T$ relation is found. This relationship is present in Figure 10.

| 13

Figure 10. Coupling Matrix generated via The Coupler
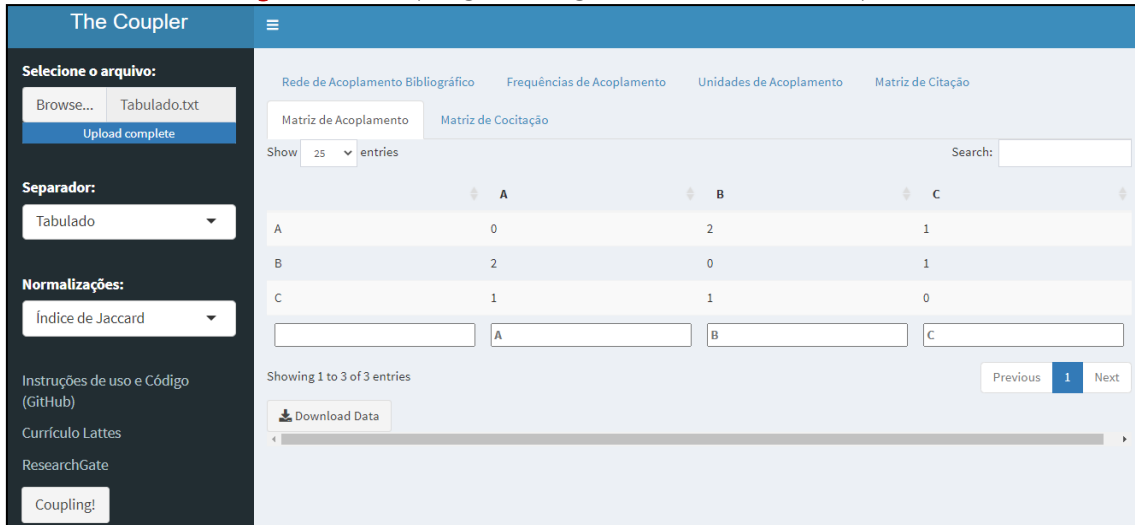


Source: Elaborated via The Coupler

Figure 10 shows, in absolute values and in matrix form, the bibliographic coupling relationship between units A, B and C. This result is similar to that found in the "Coupling" column of the "Coupling Frequencies" tab. The option for this redundancy of results is due to the fact that, traditionally, relational analyses are treated from a matrix point of view, thus, this option is offered to the user. The export of these data is similar to the others and is shown in Figure 11.

Figure 11. Export the coupling matrix via The Coupler
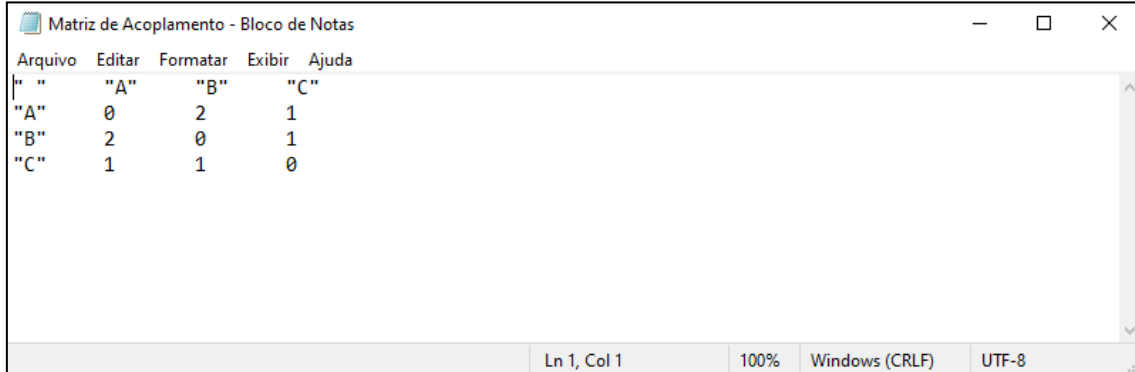


Source: Export via The Coupler

The coupling matrix in Figure 11 generated by The Coupler is composed of zeros on its main diagonal. Essentially, the use of zeros is well accepted. These values are notably different from those found in Equation 2, in which the product between the citation matrices and their transpose returns the value of the product between vector A, B or C with itself. However, this does not affect the result of the bibliographic coupling.

As a final result, we have the co-citation matrix between the cited units $i, j, k$ and $l$. This matrix is shown in Figure 12, and explains the matrix relationship between the transposed citation matrix and the citation matrix itself ($C^T \times C$).

Figure 12. Co-citation matrix generated via The Coupler
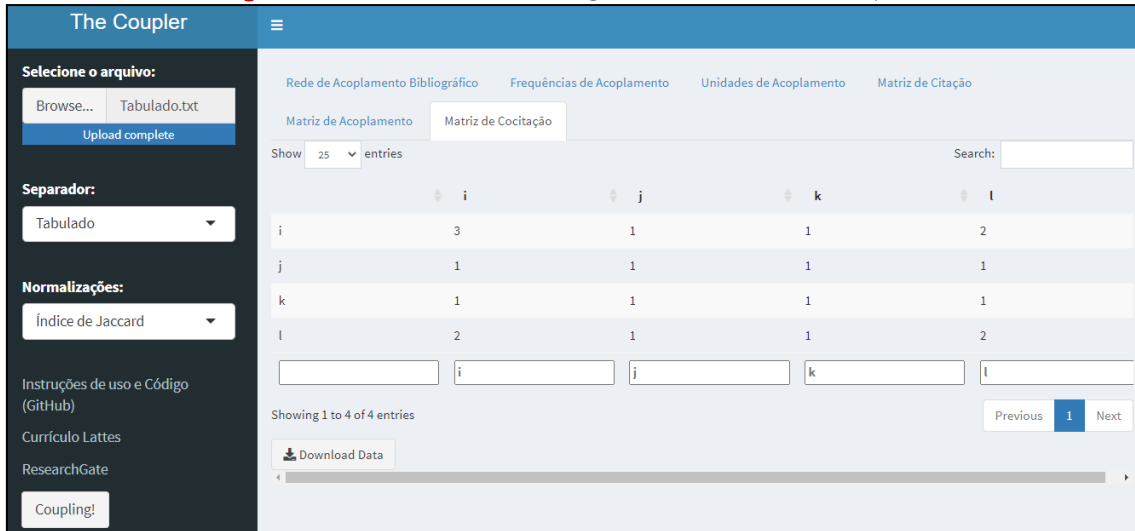
Source: Elaborated via The Coupler

The matrix in Figure 12 is similar to that found in Equation 3. In essence, a co-citation matrix quantifies the frequency with which two units were cited concomitantly. The main diagonal of a matrix represents the citation frequency received by each unit. Thus, the values of the main diagonal 3, 1, 1 and 2 indicate that the units $i, j, k,$ and $l$ were cited 3, 1, 1 and 2 times, respectively. It is possible to export the citation matrix in .txt extension, as shown in Figure 13.

Figure 13. Co-citation matrix export via The Coupler



Source: Export via The Coupler

| 15
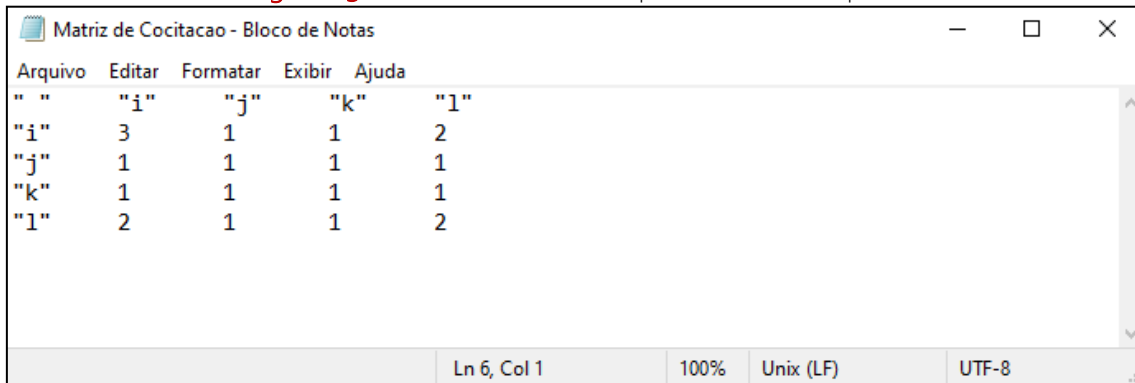
The co-citation matrix export can be considered the last result provided by *The Coupler*. Thus, from the set of data obtained through the tool, the user can proceed with the analyses. As a last demonstration, if a set of units to be analyzed does not show any coupling between them, the application will show the alert message: "WARNING: No couplings between units" as shown in Figure 14.

Figure 14. The Coupler warning message
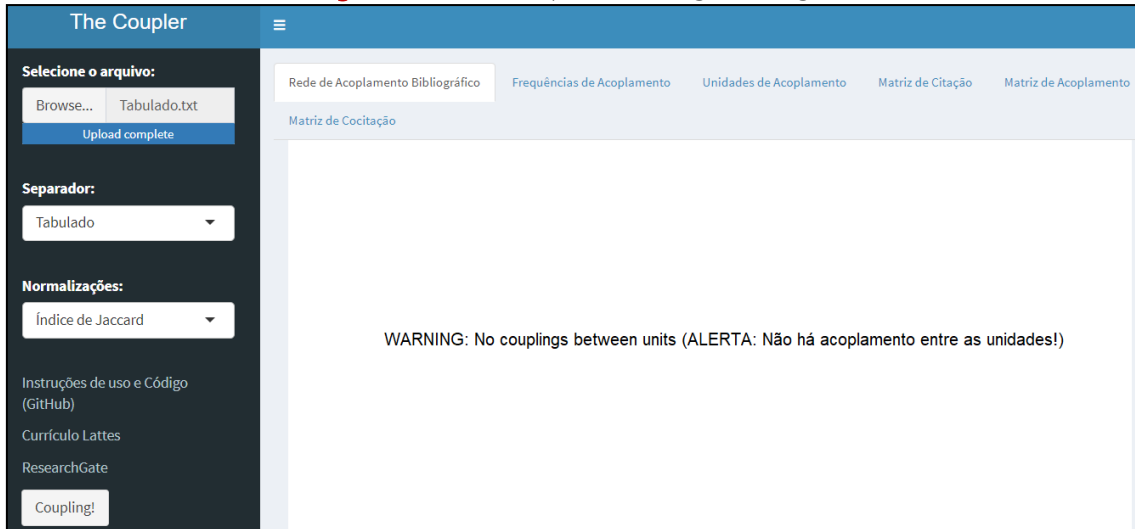
In case none of the analysis units are coupled together, none of the six analyses (Figures 3, 4, 6, 8, 10, 12) will be processed. In this way, in order for the coupling frequency calculations to be processed, the identification of the coupling units in addition to the citation, coupling and co-citation matrices, at least two units must be coupled to each other.

If necessary, the user can use the exported data from the citation, coupling and co-citation matrices in software such as Ucinet to build citation, coupling and co-citation networks. It is noteworthy that, when exporting data in tabulated format, the use of these data in electronic spreadsheets such as Microsoft Excel or Google Sheets is facilitated.

| 16

## 5 BIBLIOMETRIC, PATENTOMETRIC, ALTIMMETRIC DATA AND NATURAL LANGUAGE

The main idea of *The Coupler* is to promote analysis of proximities between any sets, and regardless of the bibliometric focus, it allows any type of analysis, whether within the scope of MIS or not. Thus, as a final demonstration of the tool, bibliometric, patentometric, altmetric and natural language data were processed. After processing, coupling frequencies (similarities) and coupling units between each analyzed pair were displayed.

Figure 15. Processing via The Coupler with different types of units of analysis

**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 |

Source: Elaborated via The Coupler

| 17

These four analyses differ little from those previously exemplified (Figures 3 to 13) since *The Coupler* does not differentiate the units of analysis. This characteristic favors its use, as by admitting diverse data, this tool becomes extremely versatile and can be applied in any type of similarity and/or co-occurrence analysis. These analyses are exemplified by identifying: co-authors in common between researchers (For example, taking Figure 1 as a basis, researchers would be A, B and C and co-authors *i, j, k* and *l*), common members of research groups, co-occurrence and/or similarity of keywords, references cited in common by course teaching plans, among others.

All calculations shown in Figure 15 were manually checked, and the tool correctly calculated Coupling, Salton's Cosine, and Jaccard Index values, and correctly identified all units of analysis and all coupling units; the latter, *The Coupler*'s standout feature. In analyses of publications in a particular area, institution or specific theme, the coupling units represent a fundamental part of the intellectual structure of the set of analyzed papers. Identifying the elements responsible for coupling two items may provide means

for identifying the main influences of a domain, based on the recurrence of certain units in the different lists of analyzed references.

## 6 FINAL CONSIDERATIONS

This research presented and demonstrated the use of the software *The Coupler*, a new free tool for relational analysis focused on bibliographic coupling (or several similarity analyses). Firstly, from a mathematical and bibliometric point of view, the analyses of citation, bibliographic coupling and co-citation was based, and, from these foundations, the new tool was presented.

*The Coupler* consists of a web application capable of generating citation, coupling and co-citation matrices, in addition to calculating normalized coupling frequencies via Salton's Cosine and Jaccard Index, building the bibliographic coupling network valued by the normalizations and identifying the coupling units (elements responsible for connecting two units of analysis).

This last result can be considered the tool's most prominent item, since this type of identification is not common in other bibliometric tools. In the same proportion, the calculation of normalized coupling frequencies via Salton's Cosine and Jaccard Index also provides *The Coupler* with its own originality, since normalizations are extremely useful for comparisons between different contexts to be carried out.

Furthermore, the tool was submitted to a demonstration using generic data (Figures 3 to 13) as well as tests using real data, commonly used in bibliometric, patentometric, altmetric analysis and in natural language processing. *The Coupler* responded to the tests as expected, calculating all coupling frequencies between the units of analysis, in an absolute and normalized way, in addition to identifying all coupling units.

While limitations are found, for the web-app version, the tool will find problems in processing large datasets, since, currently, its hosting on the shinyapps.io server has an instance size limit of 1 gigabyte. In cases of large sets, we suggest using *The Coupler* in its offline version and adjusting the memory limit used by the R software. For memory adjustments, the functions *memory.size* and *memory.limit* are used. In simulated tests, *The Coupler*, launched offline via R, completed analysis of a file containing 1001 citing articles and 10010 cited items in approximately 50 minutes and 3 seconds. That is, when processing 1001 items, the tool completed the calculation of 500500 coupling interactions together with their respective normalized values, in addition to the citation, coupling and co-citation matrices. It is noteworthy that the computational cost may vary according to the computer used. This test was performed on the computer mentioned in the Methodology section.

With this, the tool was able to process diverse data and thus, it can be considered as an alternative tool to other bibliometric software aimed at relational analysis. As next steps, we expect to keep *The Coupler* always up to date, free of any problems that may occur, in addition to carrying out future implementations required from users, such as, for example, the direct acceptance of files exported from the Web Of Science and Scopus databases and converting the tool into an executable file. Finally, we invite the entire community to use *The Coupler*.

| 18

# REFERENCES

CASTANHA, R. G. **THE COUPLER**. [*S. l.*]: Zenodo, 2022. DOI: https://zenodo.org/record/7222122Available at: https://rafaelcastanha.shinyapps.io/thecoupler/. Acesso on: 6 Sept. 2022.

DELBIANCO, N. R. A comunicação científica no *Twitter*: um estudo altmétrico com periódicos brasileiros da ciência da informação. Dissertação (Mestrado) - Universidade Estadual Paulista, 2022. Available at: http://hdl.handle.net/11449/235088. Acesso on: 6 Sept. 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. Available at: https://bit.ly/3CnASvn Acesso on: 10 Sept. 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 Sept. 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 on: 10 Sept. 2022.

GRÁCIO, M. C. C.; OLIVEIRA, E. F. T. Indicadores de proximidades em Analise 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. Available at: https://bit.ly/3rO2ZyY Acesso on: 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 on: 12 Sept. 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 on: 13 Sept. 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,

| 19

p. 49-51, Jan.1963b. https://doi.org/10.1109/TIT.1963.1057800. Acesso on: 13 Sept. 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 on: 9 Sept. 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 on: 4  Sept. 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 on: 16 Sept. 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 on: 13 Sept. 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 on: 14 Sept. 2022.

| 20

# APPENDIX 1 – *The Coupler* Complete Code

```
#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-Gutierres-Castanha-2",
"ResearchGate")),

fluidPage(h5(" ")),

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

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

| 21

```
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==" "|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_aba$ind),"...",fixed=TRUE)))
Freq_ABA["Coupling"]<-int_aba$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)
```

| **24**

```
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)
```

# APPENDIX 2 - Bibliometric data

Chart 1 – Processed bibliometric data

| 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 | | |

Source: by the author, extracted via VosViewer software

| 27

# APPENDIX 3 - Patentometric data

Chart 2 - Processed patentometric data

| 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 | |

Source: by the author

| 28

# APPENDIX 4 - Natural language processing

## Chart 3 – data for natural language processing

| 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 | capes | 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 |
| *web*ometric | 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 | |

Source: by the author, extracted via VosViewer software