



BIBLIOBICLUSTER: a bicluster algorithm for Bibliometrics

Gloria Gheno

► To cite this version:

| Gloria Gheno. BIBLIOBICLUSTER: a bicluster algorithm for Bibliometrics. 2021. hal-03176527

HAL Id: hal-03176527

<https://inria.hal.science/hal-03176527>

Preprint submitted on 22 Mar 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

BIBLIOBICLUSTER: a bicluster algorithm for Bibliometrics

Gloria Gheno¹

¹ gloria.gheno@ronininstitute.org

Ronin Institute, 127 Haddon Place, Montclair, NJ, 07043-2314 (USA)

Abstract

Bibliographic coupling and co-citation analysis methodologies were proposed in the early 60's and 70's to study the structure and production of scientific communities. Bibliographic coupling is fundamental to understand the current state of a particular research area and its possible and potential future direction. Co-citation analysis, instead, is used to map the roots of academic works, fundamental to the development of a specific research field. With the first method, papers which have a common reference are grouped and the strength of their link results from the number of references in common. The second, instead, groups together the papers co-cited by one or more documents. Both methodologies assume which the papers, citing the same articles or cited from the same article, have similar aspects. Since until now these two methodologies have been considered separately, I propose a new algorithm, based on bicluster analysis, which applies them together. The importance of this new methodology is therefore to group together the paired bibliographic papers and the co-cited references, but keeping them divided. In the resulting bicluster, the references grouped together represent the roots from which the trend to which the citing papers, grouped together, was born.

Introduction

Given the significant increase in academic and scientific production, its analysis becomes essential to understand the state of the art and the evolution of new currents in different scientific fields. To carry out this analysis, Bibliometrics mainly uses two procedures: performance analysis and science mapping (Cobo et al., 2011). Performance analysis analyzes the performance of groups of scientific actors, for example universities, departments, researchers, etc. (Cobo et al., 2011). Science mapping, instead, analyzes the structural and dynamic aspects of scientific research (Cobo et al. 2011). It is mainly based on the principle of information co-occurrence, i.e., if two or more elements are present together in one or more documents, they are closely linked (Noyons, 2004). Both performance analysis and science mapping are based on the analysis of the citation (Van Raan, 2014).

The analysis of the citation to study performance, in its basic form, counts the number of citations of specific papers (Van Raan, 2014). There are two measures to calculate performance: popularity and prestige, also called Page rank. Popularity is measured by considering the number of citations (Ding and Cronin, 2011). For example, in a network analysis, it is represented by the in-degree measure, which counts the number of head endpoints adjacent to a node (Cherven, 2013). The prestige, instead, is measured considering the number of citations obtained from the most cited papers (Ding et al. 2009, Fahimnia et al., 2015).

To study science mapping, the scholar proceeds to the analysis of the citation dividing the main network into two secondary networks, defined as co-citation and bibliographic coupling, and grouping their elements in clusters using the strength of the link among the elements of the secondary networks (Chen and Morris, 2003). Co-citation analysis groups the papers co-cited by one or more documents and the strength of their link is given by the number of citing papers in common (Small, 1973). Bibliographic coupling, instead, groups the papers which have a common reference, and the strength of their link is given by the number of references in common (Kessler, 1963). To illustrate this distinction, I use a network composed of four papers (Figure 1). Paper A cites paper C, while paper B cites papers C and D. Then papers A and B are the citing papers, while papers C and D are the cited papers or references. For bibliographic coupling, papers A and B are linked, having both cited C. For co-citation analysis, papers C and D are linked because both are cited by paper B. Vogel and Guttel (2013) consider the differences between these two methodologies. Bibliographic coupling is useful for understanding the state

of the art and the possible and eventual developments of a particular research area, while co-citation analysis is used to map its roots. Bibliographic coupling is a static approach, while co-citation analysis is dynamic. Bibliographic coupling is an indicator of publication activity, while co-citation analysis is an indicator of the impact of the publications on research.

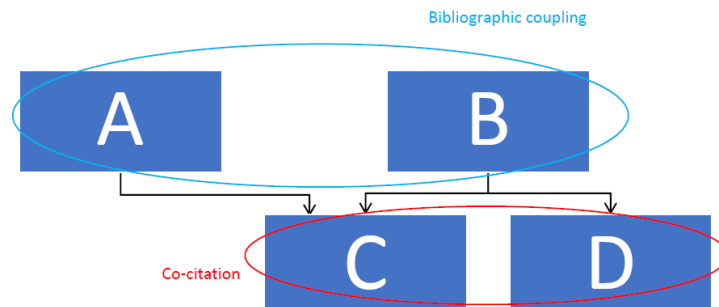


Figure 1: Subdivision of a network into elements connected according to bibliographic coupling and those connected according to co-citation analysis

Until now, these two methodologies have never been considered together, but it would be interesting to be able to do it so to understand the roots from which the state of the art and its future developments are developed. For this reason, I propose a new bicluster method which considers together bibliographic analysis, co-citation analysis and text analysis, which is generally applied to the results of the first two. Liu and Hsu (2019) tried to link bibliographic coupling, co-citation analysis and text mining, proposing a method of bibliographic coupling with a grouping rule based on co-citation analysis and text analysis. Using bicluster methodology to link the three methods together, my analysis becomes more complete because in my results I have both the citing papers and the cited papers unlike Liu and Hsu which have only the citing papers as result of their methodology. My method can be considered as a co-citation analysis and a bibliographic coupling done together, while the method proposed by Liu and Hsu (2019) is a bibliographic coupling with only the addition of the information obtained from co-citation analysis. The advantage of my method is given by the use of the bicluster technique.

Bicluster analysis was developed to study biological data (Chen and Church, 2000) and has subsequently been applied to many fields, but until now never to Bibliometrics. A bicluster algorithm groups both rows and columns simultaneously. There are many types of bicluster depending on the similarity which is considered: bicluster with constant values, with constant values in the rows or columns, bicluster with consistent values on the rows and columns and bicluster with consistent correlation on the rows and columns (Madeira and Oliveira, 2004). In this paper I propose a bicluster method which produces biclusters with constant values, and which is based on the bicluster methodology for binary data proposed by Prelić et al. (2006) and its modified version (Dolničar et al., 2012; Wang et al., 2016). The difference between the method proposed by Prelić et al. (2006) and the modified version (Dolničar et al., 2012; Wang et al., 2016) consists in the possibility of the biclusters to be overlapping. In Prelić's method the biclusters can have both columns and rows in common, while in the modified version only the columns can be in common. The bicluster algorithm, which I propose, finds biclusters with only columns in common. The difference between the method proposed in this paper and that proposed by Dolničar et al. (2012) and by Wang et al. (2016) consists in the training and selection procedure of the biclusters.

Methodology

Linking co-citation analysis, which determines the roots, to bibliographic coupling, which determines the current state of the art and its future direction, it is possible to better study the

evolution of a particular research area. To link the two analyses, I propose a modified version of the bicluster methodology for binary data (Prelić et al., 2006). I call this new algorithm BIBLIOBICLUSTER, which simultaneously groups the references and the citing papers. The references grouped together represent the roots from which the trend of the citing papers, grouped together, is developed. The simplified example in Figure 2 explains the result of the algorithm. To study a specific research area, the scholar considers only papers A and B, which cite papers C, D and E. Co-citation analysis finds the link only among the three references and therefore finds only the roots of this specific field of research. Applying to the titles a textual analysis, based on the frequency of the words, it is found that this field of research is mainly based on the study of dogs and cats. Bibliographic coupling links the two citing papers and, applying again textual analysis based on the frequencies, finds three main trends: food, health and environment. By combining the two analyses, the algorithm obtains that, from a generic study of dogs and cats, scholars have moved on to focus on their food, their health and the environment in which they live.

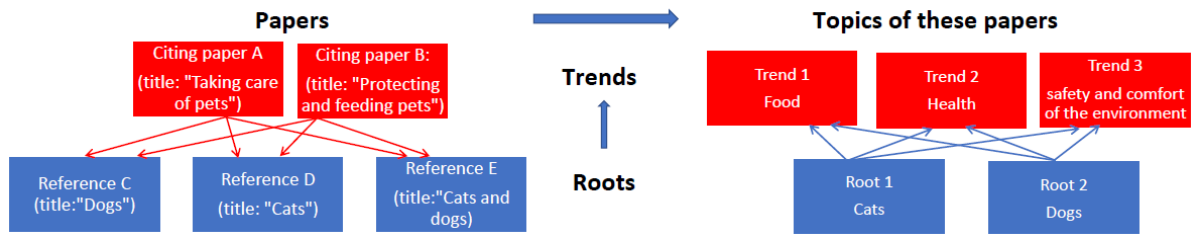


Figure 2: Example of data interpretation

The algorithm analyses a matrix (Figure 3) in which the columns represent the references and the rows the citing papers. If paper 1 cites reference 1, a_{11} is equal to 1, if it does not cite it, a_{11} is equal to 0, so the input of the algorithm is a matrix of 0 and 1.

	Reference 1	Reference 2	...	Reference k	...	Reference K
Paper 1	a_{11}	a_{12}	...	a_{1k}	...	a_{1K}
Paper 2	a_{21}	a_{22}	...	a_{2k}	...	a_{2K}
\vdots	\vdots	\vdots	\ddots	\vdots	\ddots	\vdots
Paper j	a_{j1}	a_{j2}	...	a_{jk}	...	a_{jK}
\vdots	\vdots	\vdots	\ddots	\vdots	\ddots	\vdots
Paper J	a_{J1}	a_{J2}	...	a_{Jk}	...	a_{JK}

Figure 3: Example of analysed dataset

From this matrix, the algorithm derives all possible biclusters using the method proposed by Prelić et al. (2006). Among these found biclusters, the algorithm selects the one that maximizes the function

$$C_m * R_m + \lambda \max \text{freq of } B_m \quad (1)$$

where C_m and R_m respectively represent the number of the columns and that of the rows of the generic bicluster B_m , “ $\max \text{freq of } B_m$ ” represents the frequency with which the word is mostly present in the titles of the papers in B_m and λ is a parameter whose value is included in the interval [0,1], measuring the importance of text analysis. If the parameter λ is equal to 0, text analysis is not considered. The introduction of λ can eliminate one of the weaknesses of bibliographic coupling. Bibliographic coupling, indeed, can group two unrelated articles just because they cited the same reference (Liu and Hsu, 2019). The introduction of the study of the

titles of the references seeks to reduce this problem by trying to favor biclusters in which the references have titles with similar words and therefore more connected. There are cases in which two or more biclusters maximize formula (1) and then among them the algorithm chooses the one that maximizes the following function

$$Index = \sum_k \left(\frac{(\sum_j a_{jk} - 1)_+}{R_m} + I \left(\sum_j a_{jk} \right)_{\sum_j a_{jk}=0} \right) \quad (2)$$

where j is the row j of the bicluster B_m and k is the column k excluded from the bicluster B_m . This index analyzes the goodness of the bicluster considering the ungrouped columns. They represent the references and therefore the more a reference is cited by the members of a bicluster, the more similar are the elements of the bicluster. Indeed, it is sufficient that only one of the members of the bicluster does not mention this reference and this column is excluded. The same consideration is valid if no member of the bicluster mentions a reference. The more the number of references which are not cited by any member of a bicluster increases, the more the goodness of the bicluster increases because there are no other arguments. To explain how this mechanism works, I use the matrix of Figure 4 and consider λ equal to 0. The blue and pink biclusters have the same size ($C_{blue} * R_{blue} = C_{pink} * R_{pink} = 2 * 4 = 8$) and then the algorithm applies formula (2) to choose the best.

	Reference 1	Reference 3	Reference 4	Reference 2
Paper A	1	1	0	0
Paper B	1	1	0	1
Paper C	1	1	0	1
Paper D	1	1	1	0
Paper E	0	1	1	0
Paper F	0	1	1	0
Paper G	0	1	1	0

Figure 4: Choice between two biclusters with the same size

The analysed columns are those in blue for the blue bicluster and the pink ones for the pink bicluster. For simplicity we denote the rows as A, B, C, D, E, F, G and the columns as 1,3,4,2.

$$\begin{aligned}
 Index_{blue} &= \sum_{k=4,2} \left(\frac{(\sum_{j=A,B,C,D} a_{j,k} - 1)_+}{R_{blue}} + I \left(\sum_{j=A,B,C,D} a_{jk} \right)_{\sum_{j=A,B,C,D} a_{jk}=0} \right) \\
 &= \frac{(1-1)_+}{4} + I(1)_{\sum_{A,B,C,D} a_{j1}=0} + \frac{(2-1)_+}{4} + I(2)_{\sum_{A,B,C,D} a_{j2}=0} \\
 &= 0 + 0 + 0.25 + 0 = 0.25
 \end{aligned}$$

$$\begin{aligned}
 Index_{pink} &= \sum_{k=1,2} \left(\frac{(\sum_{j=D,E,F,G} a_{j,k} - 1)_+}{R_{pink}} + I \left(\sum_{j=D,E,F,G} a_{jk} \right)_{\sum_{j=D,E,F,G} a_{jk}=0} \right) \\
 &= \frac{(1-1)_+}{4} + I(1)_{\sum_{D,E,F,G} a_{j4}=0} + \frac{(0-1)_+}{4} + I(0)_{\sum_{D,E,F,G} a_{j2}=0} \\
 &= 0 + 0 + 0 + 1 = 1
 \end{aligned}$$

Using formula (2), the algorithm selects the pink bicluster. Reference 2 represents a sub-topic of the blue bicluster and therefore confirms an even closer link among some papers of the bicluster. Reference 2 is never cited by the pink bicluster papers, thus confirming a close link among them. The concept expressed by this index is summarized in Figure 5. Therefore, formula (2) tries to eliminate the second and final weakness noted by Liu and Hsu (2019) of bibliographic coupling. Liu and Hsu (2019) noted that two related articles may cite different references. The study of non-grouped references, therefore, determines the goodness of the bicluster, preferring biclusters whose non-grouped references are similar or those which group all the references.

Number of excluded References	Concept	Similarity
Number of references >1	Subgroup topic	similarity (\uparrow number $\Rightarrow \uparrow$ similarity)
Number of references =1	Different topic	dissimilarity
Number of references =0	No excluded topics	maximum similarity

Figure 5: Concepts expressed by the goodness index

Once the bicluster has been chosen, the algorithm eliminates the rows which belong to the selected bicluster from the data matrix. Then the algorithm keeps repeating as long as it continues to find biclusters with a size greater than or equal to two. The elimination of the rows and not of the columns is to emphasize that the same roots can lead to more trends. This last step determines that the obtained biclusters have exclusive rows, i.e. that only the columns can belong to more biclusters. Formula (1) and formula (2) highlight the main difference between BIBLIOBICLUSTER and the bicluster method proposed by Dolničar et al. (2012) and by Wang et al. (2016).

Comparison with other methods

To better explain the functioning of BIBLIOBICLUSTER algorithm and compare it with the methods already present in the literature, I use some simplified datasets. The methods of comparison are direct citation, also called inter citations (Klavans and Boyack, 2006; Shibata et al., 2008; Boyack and Klavans 2010), co-citation analysis (Small, 1973) and bibliographic coupling (Kessler, 1963). Small (1997, 1999) grouped the three methods together, considering all the papers, cited and citing, and, using the three methodologies, he calculates the strength of the relationships. The difference between my method and Small's method is substantial because I distinguish between the citing papers and the cited ones, while he considers them together without distinguish them. To implement these analyses, I use Gephi software (Bastian et al., 2009) and the modularity algorithm (Blondell et al., 2008) implemented in the software itself. In order to highlight the main difference between BIBLIOBICLUSTER and the three methodologies, the substantial difference between a bicluster and a cluster must be well understood. BIBLIOBICLUSTER produces biclusters, while the three methodologies present in the literature produce clusters. Bicluster method groups a set of objects considering two variables, the cluster method, instead, considers only one variable. For this reason, direct citation has the limitation of not being able to distinguish between the citing papers and the cited papers, co-citation analysis that of considering only the cited papers and the bibliographic coupling that of considering only the citing papers. BIBLIOBICLUSTER, instead, manages to group both the cited papers and the citing papers considering them together, but distinguishing them. In the first example I show how the solution of the algorithm varies using $\lambda = 0$ and $\lambda = 2$. The dataset used is shown in Figure 6 and it is created with titles without any underlying scientific theory, just to show the functioning of the algorithm. In the dataset the first row is represented by the titles of the references and therefore, for example, reference 1 has the title “Animals: white cat”. Since I assume that the citing papers are selected using the word Animals,

I delete this from text analysis, assuming that even the references often have the word animals in the title.

	Animals: white cat	Animals: European cat	Animals: dog	Animals: pets
Paper A	1	0	1	0
Paper B	1	1	1	0
Paper C	1	1	1	0
Paper D	1	0	1	1

Figure 6: Dataset of the example 1

When $\lambda = 0$, the algorithm finds only a bicluster (rows: Paper A, Paper B, Paper C, Paper D; columns: “Animals: white cat”, “Animals: dog”). When $\lambda = 2$ is set, the algorithm finds 2 biclusters (Figure 7). Without text analysis the algorithm selects only the bicluster with the maximum size (green bicluster in Figure 7), with text analysis the algorithm, when it applies formula (1), finds 2 biclusters (green bicluster and blue bicluster in Figure 7) which have the same value. To choose, therefore, the best of them, the algorithm applies formula (2). From the algorithm without text analysis, the papers on the white cat and on the dog are the roots of all of the four citing papers, while from the algorithm with text analysis, the papers on the white cat and on the dog are the roots of the trend of papers A and D, while the same papers, together with the one on the European cat, represent the basis of the trend of papers B and C.

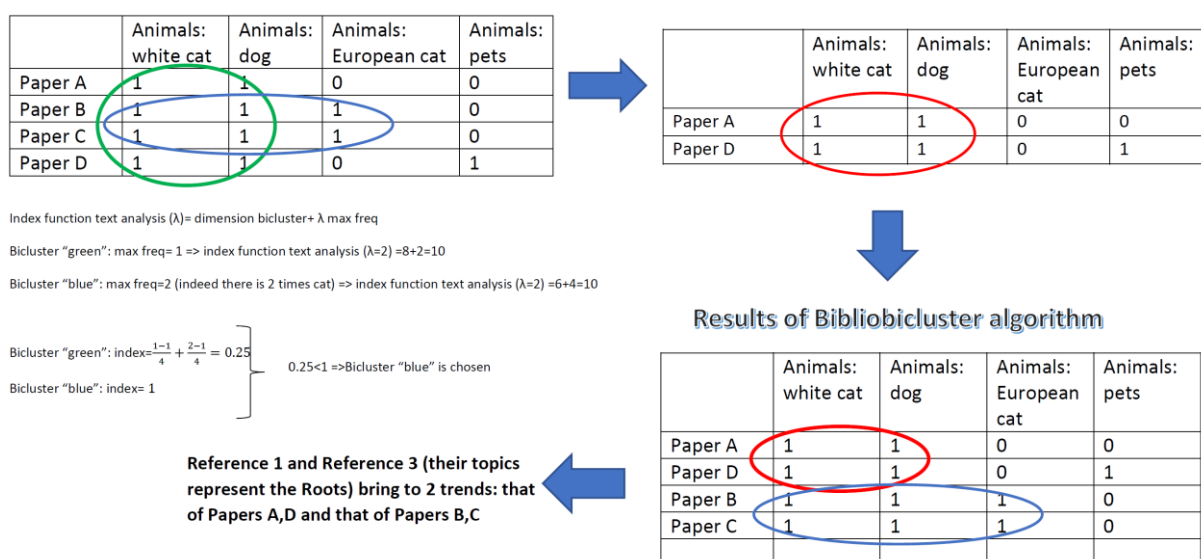


Figure 7: BIBLIOBICLUSTER applied to the first example with $\lambda = 2$

In citation analysis, co-citation analysis and bibliographic coupling, text analysis is carried out subsequently and then the two results obtained by BIBLIOBICLUSTER are compared with a single result for each method. Direct citation finds three clusters, while the other methodologies find only one cluster (Figure 8). The thickness of the lines, which join the various elements, represents the strength of the link between the two documents. BIBLIOBICLUSTER, without text analysis, groups the two references, which are more closely linked by co-citation analysis, and the four citing papers grouped together by bibliographic coupling. BIBLIOBICLUSTER, therefore, considers the groups of bibliographic coupling and the strong relationships of co-citation analysis. The results of BIBLIOBICLUSTER with text analysis, adding information which other methods do not consider, find new and unprecedented links. BIBLIOBICLUSTER is also useful as a method of data reduction, indeed, both with text analysis and without text

analysis, the algorithm eliminates the references which are less cited, while the three methods proposed in the literature continue to consider all the data.

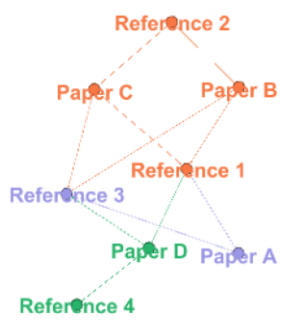
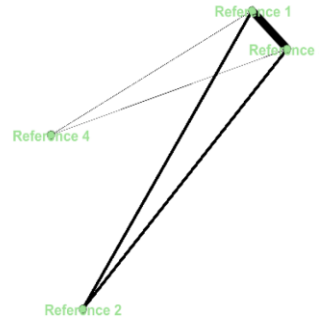
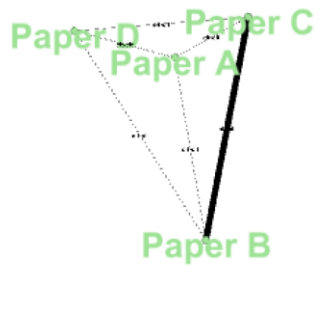
Direct citation analysis	Co-citation analysis	Bibliographic coupling analysis
		
3 Clusters (orange, purple, green)	1 Cluster (green)	1 Cluster (green)

Figure 8: Results of the methodologies present in the literature relating to example 1

To further show the functioning of BIBLIOBICLUSTER and compare it with the 3 methods proposed in the literature, I use a second example, whose dataset is shown in Figure 4. With this example it is possible to better understand the problem of the different references of related articles. To analyse this problem, I preferred not to consider text analysis and therefore set λ equal to 0. In the first step, the algorithm finds two biclusters with the same size (Figure 9), the pink one and the blue one and, applying formula (2), the pink bicluster is chosen. The blue bicluster has, indeed, as residual columns reference 4 and reference 2, which are cited respectively 1 and 2 times. The pink bicluster, instead, has reference 1 and reference 2 as residual columns, which are cited 1 and 0 times respectively. Since it is more likely that the linked papers of the pink bicluster did not mention different references, having a column equal to 0 and one with only 1, the pink bicluster is preferred to the blue bicluster. Now the rows relative to the citing papers of the pink bicluster are removed and the biclusters are looked for again. As in the previous step, also in this case there are two biclusters with the same size (the green one and the red one) and, always applying formula (2), the green one is chosen. The green bicluster has, indeed, as residual columns reference 2 and reference 4, which are cited 2 and 0 times respectively. The red bicluster, instead, has reference 4 as its residual column, which is cited 0 times. Since reference 2 is cited 2 times out of 3 by the green bicluster papers, it is very likely that paper A is linked to B and C even if they have cited different articles, being reference 4 not cited by anyone. The red bicluster cannot be chosen because it excludes only one column of all zeros. The exclusion of a single column of all zeros is insufficient to determine the goodness of a bicluster.

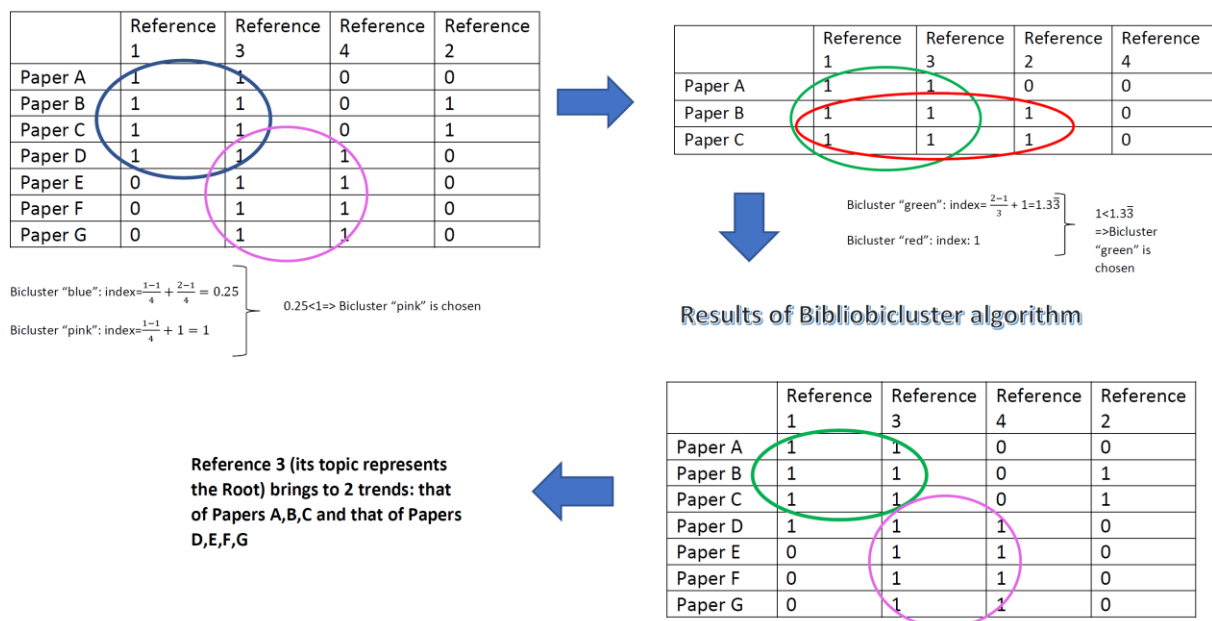


Figure 9: BIBLIOBICLUSTER applied to the second example with $\lambda = 0$

In conclusion, BIBLIOBICLUSTER finds two biclusters and excludes reference 2 from the analysis. If, instead, I apply direct citation, co-citation analysis and bibliographic coupling to the same dataset, I get the results shown in Figure 10. Therefore, the three methods do not delete the reference 2, as BIBLIOBICLUSTER does. Direct citation produces two clusters, co-citation analysis one cluster and bibliographic coupling two clusters. As in example 1, also in this case BIBLIOBICLUSTER considers the strong relationships of co-citation analysis and the clusters of bibliographic coupling.

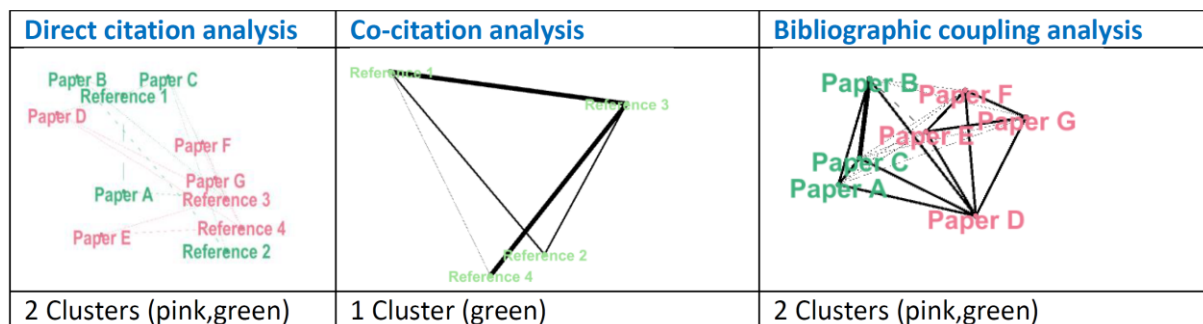
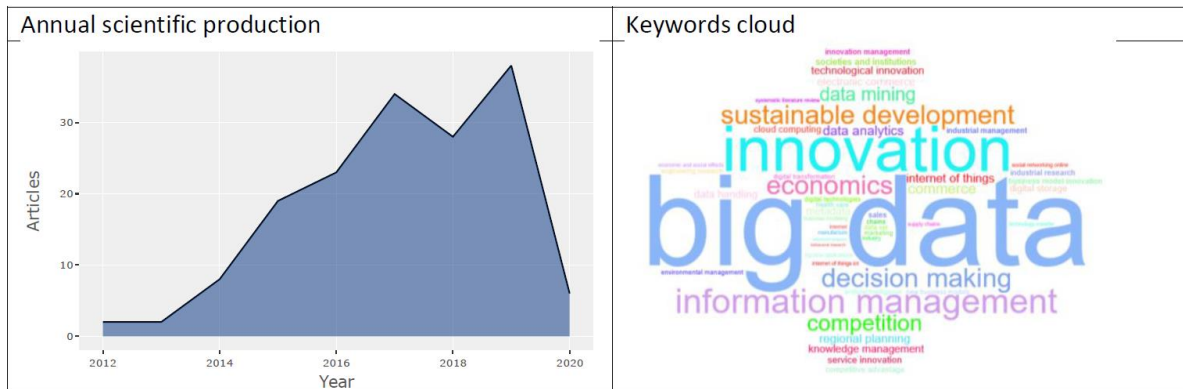


Figure 10: Results of the three methodologies present in the literature relating to example 2

Application to a real case

I apply my new algorithm to the analysis of economic papers relative to accounting published between 2011 and the first quarter of 2020, which have big data among the keywords. As it can be seen from Figure 11, the papers on big data tend to increase in number, demonstrating the increase in importance of this topic. From the keywords cloud graph, in which the increase in font size of the word demonstrates the highest frequency, it can be seen that big data is the word with the greatest frequency actually. Precisely this term is the discriminating factor in order that a paper is included in the dataset. The initial dataset is made up of 160 documents extracted from Scopus. These documents are articles, books, book chapters, conference articles, editorials and short questionnaires. I consider only 91 articles, eliminating, however, three of them because they have no references.



I apply the algorithm without text analysis ($\lambda = 0$) to the dataset, to show how this can be done later, as in the methods already present in the literature. BIBLIOBICLUSTER finds sixteen biclusters, but I apply text analysis only to the biclusters that are larger than 8, believing that the smallest ones represent niche topics. Of course, all biclusters could also be analyzed. In this case I examine both the titles of the references and those of the citing papers. I also consider both the frequency with which a word is found in titles and the one with which it is found next to another word (Hornik and Grün, 2011). Of course, this analysis, or other similar ones, can also be applied to the results of the algorithm with λ greater than 0. The results of the text analysis are shown in Figure 12. For example, if I consider bicluster 1, I find that the study of the ability to learning about the business and its performance led to three main analyses: the innovation, the role of intuition and the impact of data on businesses.

	trends	roots
Bicluster 1	1)Innovation and enhance of the firm 2)Role of insight in the firm 3)The impact of the data on the firm	The capacities of learning of firm and its performance
Bicluster 2	Resilient enterprise and sustainability	The sustainability: innovation, the keys and the drivers, ecologic and social aspects (for example psychology and work)
Bicluster 3	Understanding the determinants of the use of the big data	Adoption of different technologies (cloud and data, e-commerce and its process, communication and electronic interchange)
Bicluster 4	Innovation and big data analysis	Competition and management
Bicluster 5	Green innovation management	New data and environmental
Bicluster 6	Business model (initiative and process, impacts and success)	Business model (innovation, strategy, design)
Bicluster 7	Innovation service and incubators	Services (innovation, research, learning)
Bicluster 8	Empirical analysis	Business and technological performance
Bicluster 9	Technology systems and emergent technology	Emergent technologies and innovation: information and competition
Bicluster 10	Relation among supply chain, performance and innovation	Supply chain prediction

Conclusions

Until now, co-citation analysis, bibliographic coupling and text analysis have been applied separately with consequent limitations. I propose a new algorithm, which I call BIBLIOBIBLUSTER to solve these limitations. This algorithm produces biclusters, in which

the references, grouped together, represent the roots from which is born the trend to which the citing papers, grouped together, belong. BIBLIOBICLUSTER can be used both considering only the citations and considering together the citations and text analysis. Obtaining a mixed method is fundamental and, indeed, some authors (Liu and Hsu, 2019) are trying to link bibliographic coupling, co-citation analysis and text mining using the theory of clusters. Liu and Hsu (2019), indeed, propose a method of bibliographic coupling with a selection rule based on co-citation analysis and text analysis. Indeed, linking them together using the biclusters, I obtain a more complete analysis and with cited papers and citing papers present in the results, unlike them which group only the citing papers

References

- Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. *lcwsm*, 8(2009), 361-362.
- Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10), P10008.
- Boyack, K. W., & Klavans, R. (2010). Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately?. *Journal of the American Society for information Science and Technology*, 61(12), 2389-2404.
- Chen, C., & Morris, S. (2003, October). Visualizing evolving networks: Minimum spanning trees versus pathfinder networks. In *IEEE Symposium on Information Visualization 2003 (IEEE Cat. No. 03TH8714)* (pp. 67-74). IEEE.
- Cheng, Y., & Church, G. M. (2000, August). Biclustering of expression data. In *Ismb* (Vol. 8, No. 2000, pp. 93-103).
- Cherven, K. (2013). *Network graph analysis and visualization with Gephi*. Packt Publishing Ltd.
- Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the fuzzy sets theory field. *Journal of informetrics*, 5(1), 146-166.
- Ding, Y., Yan, E., Frazho, A., & Caverlee, J. (2009). PageRank for ranking authors in co-citation networks. *Journal of the American Society for Information Science and Technology*, 60(11), 2229-2243.
- Ding, Y., & Cronin, B. (2011). Popular and/or prestigious? Measures of scholarly esteem. *Information processing & management*, 47(1), 80-96.
- Dolnicar, S., Kaiser, S., Lazarevski, K., & Leisch, F. (2012). Biclustering: Overcoming data dimensionality problems in market segmentation. *Journal of Travel Research*, 51(1), 41-49.
- Fahimnia, B., Bell GH, M., Hensher A, D., & Sarkis, J. (2015). *Green Logistics & Transportation: A Sustainable Supply Chain Perspective*.
- Liu, R. L., & Hsu, C. K. (2019). Improving Bibliographic Coupling with Category-Based Cocitation. *Applied Sciences*, 9(23), 5176.
- Hornik, K., & Grün, B. (2011). topicmodels: An R package for fitting topic models. *Journal of statistical software*, 40(13), 1-30.
- Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *American documentation*, 14(1), 10-25.
- Klavans, R., & Boyack, K. W. (2006). Identifying a better measure of relatedness for mapping science. *Journal of the American Society for Information Science and Technology*, 57(2), 251-263.
- Madeira, S. C., & Oliveira, A. L. (2004). Biclustering algorithms for biological data analysis: a survey. *IEEE/ACM transactions on computational biology and bioinformatics*, 1(1), 24-45.
- Noyons, C. M. (2004). Science maps within a science policy context. In *Handbook of quantitative science and technology research* (pp. 237-255). Springer, Dordrecht.
- Prelić, A., Bleuler, S., Zimmermann, P., Wille, A., Bühlmann, P., Gruissem, W., ... & Zitzler, E. (2006). A systematic comparison and evaluation of biclustering methods for gene expression data. *Bioinformatics*, 22(9), 1122-1129.
- Shibata, N., Kajikawa, Y., Takeda, Y., & Matsushima, K. (2008). Detecting emerging research fronts based on topological measures in citation networks of scientific publications. *Technovation*, 28(11), 758-775.
- Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for information Science*, 24(4), 265-269.
- Small, H. (1997). Update on science mapping: Creating large document spaces. *Scientometrics*, 38(2), 275-293.

- Small, H. (1999). Visualizing science by citation mapping. *Journal of the American society for Information Science*, 50(9), 799-813.
- Van Raan, A. F. (2014). Advances in bibliometric analysis: research performance assessment and science mapping. *Bibliometrics Use and Abuse in the Review of Research Performance*, 17-28.
- Vogel, R., & Güttel, W. H. (2013). The dynamic capability view in strategic management: A bibliometric review. *International Journal of Management Reviews*, 15(4), 426-446.
- Wang, B., Miao, Y., Zhao, H., Jin, J., & Chen, Y. (2016). A biclustering-based method for market segmentation using customer pain points. *Engineering Applications of Artificial Intelligence*, 47, 101-109.