# Chapter 6 / Capítulo 6

*Applied bibliometrics. From data to publication (English Edition)*
**ISBN:** 978-9915-9680-7-0
**DOI:** 10.62486/978-9915-9680-7-0.ch06

# Correlation Graphs / Grafos de Correlación

## 6.1. Theoretical foundations of bibliometric graphs

Correlation graphs are a fundamental tool for analyzing relationships in scientific literature, allowing connections between academic entities to be visualized and quantified. Based on mathematical graph theory, these structures transform bibliographic data into networks in which meaningful relationships interconnect elements of the research system. The analytical power of bibliometric graphs lies in their ability to reveal underlying structural patterns in large volumes of data, facilitating the identification of intellectual communities, emerging trends, and collaborative dynamics that would remain hidden in conventional tabular analyses.

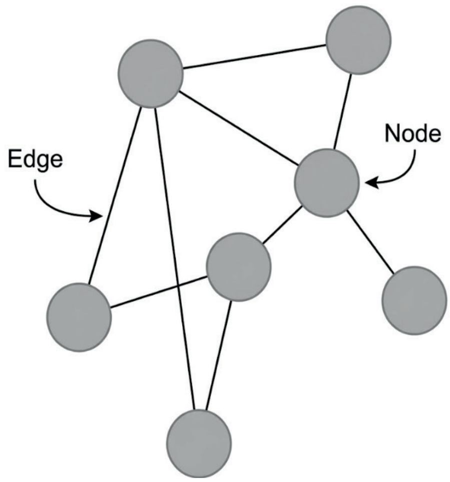


**Figure 6.1.** General structure of a graph

### 6.1.1. Nodes: authors, keywords, and journals as units of analysis

Nodes are the fundamental elements of the graph, representing the academic entities whose study reveals the structure of knowledge. When nodes represent authors, the graph visualizes collaboration communities, where relative position indicates centrality in co-authorship networks, and proximity suggests thematic or institutional affinity. Highly connected researchers function as bridges between different groups, while peripheral nodes may indicate thematic specialization or academic isolation. The density of connections around a specific author reflects their degree of integration into the scientific community, providing insights into collaboration strategies and intellectual leadership.

In keyword-based maps, nodes capture research concepts and themes, transforming intellectual content into interpretable spatial structures. The frequency of term occurrence determines node size, while co-occurrence within the same documents establishes connections between concepts. This approach reveals the conceptual architecture of a scientific field, identifying consolidated thematic nuclei, emerging areas, and research gaps. Natural terminological clusters emerge from community detection algorithms, revealing how knowledge is organized and how different subfields relate to one another within a disciplinary domain.

Scientific journals as nodes allow us to analyze the structure of academic communication and knowledge flows between different specialties. Citation graphs between journals show relationships of intellectual influence, where the direction of citations indicates knowledge flows and the strength of connections reflects disciplinary proximity. This approach identifies central journals that function as dissemination centers, bridge publications that connect different fields, and specialized peripheries. Diachronic analysis of these networks can reveal the evolution of disciplinary boundaries and the emergence of new interdisciplinary research areas.

**6.1.2. Edges: weight and direction as measures of relationship**

Edges are the connecting elements of the graph, quantifying relationships between nodes along two critical dimensions: weight and direction. The weight of an edge reflects the intensity of the connection and is calculated using different metrics depending on the type of analysis. In co-authorship networks, weight can represent the number of joint publications; in co-word networks, the frequency of terminological co-occurrence; in citation networks, the volume of citations between entities. This weighting allows us to distinguish between occasional connections and solid, sustained relationships, which is essential for identifying strategic collaborations and consolidated thematic nuclei.

The direction of the edges introduces a temporal and causal dimension to the analysis, which is particularly crucial in studies of citation and intellectual influence. In citation networks between authors or journals, directed edges indicate flows of knowledge, showing who cites whom and revealing patterns of academic influence. This directionality allows us to identify seminal works that receive many citations but cite few (sink nodes), versus synthesis works that cite extensively and are widely cited (hub nodes). Centrality analysis in directed networks provides more nuanced measures of influence that consider both the impact received and the capacity for knowledge dissemination.

The interaction between weight and direction adds a layer of analysis to the interpretation of bibliometric graphs. A heavy, bidirectional edge between two authors suggests intense, reciprocal collaboration, while a light, unidirectional edge may indicate incidental influence or one-off recognition. In co-citation analysis, the strength and reciprocity of connections reveal intellectual proximity and membership in shared schools of thought. Community detection in weighted and directed graphs identifies groups with high internal cohesion and characteristic patterns of external connections, which are essential for understanding the social and intellectual structure of scientific fields.

Bibliometric visualization programs incorporate distinctive features that determine the graphic significance and interpretive capacity of the graphs generated. VOSviewer, for example, uses layout algorithms based on attraction and repulsion that position nodes according to their similarity, creating natural clusters where visual distance represents thematic or collaborative proximity. CiteSpace, in contrast, uses temporal representations that show the evolution of networks through timelines, where the vertical position indicates publication time and the horizontal position reflects citation relationships. These algorithmic differences produce visualizations with different capacities to reveal specific patterns, making it essential to understand the representation principles of each tool.

The addition of clusters or groupings represents an advanced functionality that substantially enriches the interpretation of graphs. These clusters are generated using community detection algorithms that identify subgroups of nodes with high internal connectivity and lower external connectivity, typically differentiated by distinctive colors. In co-word analysis, clusters reveal consolidated research topics; in co-authorship networks, they show collaboration groups; in citation networks, they identify schools of thought or shared paradigms. Automatic clustering tagging using representative terms extracted from nodes is a key feature that facilitates immediate interpretation of the field of study's underlying structure.

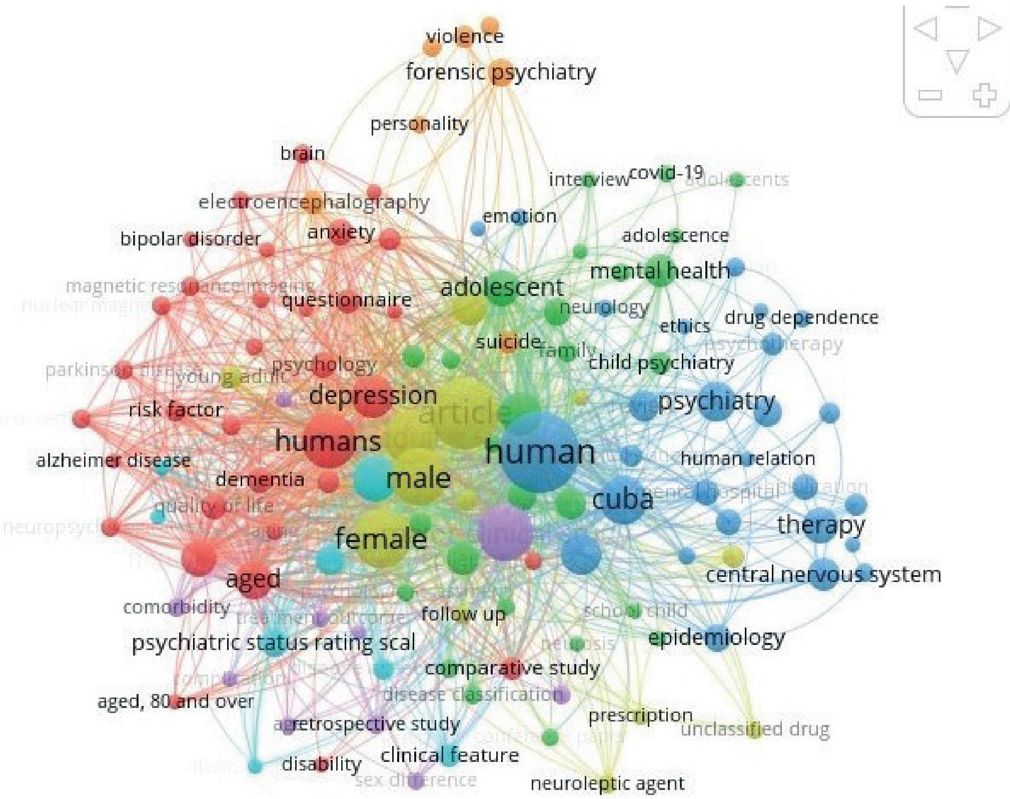


**Figure 6.2.** Example of a co-occurrence graph of terms or words

The most sophisticated tools incorporate additional elements such as concentric rings representing temporal dimensions, node sizes proportional to impact metrics, and edge thicknesses reflecting relationship intensity. Some programs, such as CitNetExplorer, add timelines showing the diachronic evolution of connections, while SciMAT introduces strategic maps that position topics according to their density and centrality. These layers of information transform graphs from simple structural representations into multidimensional analytical tools that simultaneously capture relationships, intensity, temporal evolution, and impact, requiring the user to understand these visual conventions to extract meaningful insights from the generated visualizations.

## 6.2. Key algorithms for constructing bibliometric graphs

Graph construction algorithms represent the methodological core of bibliometric network analysis, transforming bibliographic data into meaningful relational structures. Each algorithm operates under different conceptual principles and answers different research questions, requiring the analyst to have a deep understanding of its theoretical foundations and practical limitations. The appropriate selection of the algorithm determines not only the validity of the results but also the ability to extract meaningful insights into the structure and dynamics of scientific knowledge. This chapter examines in depth three fundamental algorithms—co-occurrence, bibliographic coupling, and co-citation—unraveling their operating mechanisms, characteristic applications, and optimal contexts of use.

### 6.2.1. Co-occurrence analysis: mapping conceptual structure

Co-occurrence analysis is based on the principle that the joint appearance of terms in academic documents reveals conceptual and thematic proximity. This algorithm constructs networks in which nodes represent concepts, typically extracted from keywords, titles, or abstracts, and edges reflect their co-occurrence within the same document. The technical implementation involves multiple processing stages: initially, the text is normalized through lemmatization or stemming to unify morphological variants; subsequently, non-significant terms are filtered using domain-specific stopword lists; finally, a document co-occurrence matrix is constructed where each cell records the frequency of joint occurrence.

Measuring the associative strength between terms requires the application of advanced normalization coefficients. The Jaccard index, calculated as the ratio of observed co-occurrences to the union of individual frequencies, is particularly effective at correcting for bias toward widespread terms. Alternatively, the cosine similarity coefficient measures the angle between frequency vectors in multidimensional spaces, providing robustness for analyzing large volumes of data. For contexts where terminological rarity contains valuable information, measures such as weighted specific association preserve connections between specialized terms. The selection of the appropriate coefficient depends critically on the terminological distribution of the analyzed corpus and the particular research objectives.

The applications of co-occurrence analysis range from identifying emerging research niches to mapping the cognitive structure of established disciplines. In technology watch, it enables the detection of convergences between previously disjoint fields, indicating potential areas of innovation. In interdisciplinary studies, it reveals conceptual bridges between different epistemological domains. The temporal evolution of co-occurrence networks, obtained through diachronic analysis segmented by periods, shows the dynamics of the formation, consolidation, and dissolution of research topics, providing unique perspectives on the processes of conceptual change in science.

### 6.2.2. Bibliographic coupling: connections through shared references

Bibliographic coupling establishes relationships between documents based on their shared citation profile, operating on the principle that works that cite familiar sources are likely to share theoretical, methodological, or conceptual frameworks. Unlike other algorithms, bibliographic coupling generates static networks whose connections are fixed at the time of publication and do not evolve. This feature makes it particularly valuable for analyzing recent literature that has not yet accumulated enough citations for other types of relational analysis.

The technical implementation requires the construction of a document-reference matrix where each row represents a document in the corpus and each column a cited reference. The matrix is then transformed into a similarity network through matrix multiplication and the application of association measures. The simple coupling index, based on the raw count of shared references, tends to favor documents with extensive reference lists, so in practice, normalized measures such as the Salton index are preferred, which divides the number of shared references by the square root of the product of the total references in each document. For specialized analysis, the bibliographic proximity index weights references by age, giving greater weight to recent citations.

The applications of bibliographic coupling are particularly relevant in the context of contemporary literature evaluation and the mapping of active scientific frontiers. In systematic reviews of recent advances, it allows the identification of groups of publications with similar approaches without waiting for citation traditions to be established. In interdisciplinary studies, documents that function as bridges between fields are revealed by their mixed citation patterns. The main methodological limitation lies in its sensitivity to individual citation styles and its inability to capture indirect influences or subsequent developments in intellectual relationships.

### 6.2.3. Co-citation analysis: the collective perception of the scientific community

Co-citation analysis is based on the principle that the joint citation of two documents by subsequent works reflects a relationship perceived by the scientific community. Unlike bibliographic coupling, which examines citations from source documents, co-citation analyzes citations from citing documents, thereby capturing a collective, dynamic assessment of intellectual relationships. This algorithm generates evolving networks in which the strength of connections between documents can increase, decrease, or reconfigure over time, reflecting changes in scientific consensus about the structure of knowledge.

The construction of co-citation networks involves complex technical phases, including the identification of co-cited pairs, the calculation of co-citation frequencies, and the application of minimum thresholds to include meaningful connections. Normalization of co-citation strength typically employs the standard co-citation index, which is simply the count of common citers. However, advanced implementations use Pearson's correlation coefficient to capture similarities in co-citation patterns over time. For diachronic analysis, the data is segmented into successive time windows, allowing the evolution of intellectual groupings and the emergence of new lines of research to be tracked.

The applications of co-citation analysis are compelling in historical studies of science and analyses of the evolution of intellectual paradigms. It allows for the identification of seminal works that have remained relevant over time, the detection of mergers between previously separate intellectual traditions, and the analysis of processes of specialization and disciplinary fragmentation. In scientific evaluation, it provides robust indicators of lasting intellectual impact beyond conventional citation metrics. Among its limitations are the inherent time lag before documents accumulate enough citations for meaningful analysis and the possible reinforcement of established canons to the detriment of marginal but potentially transformative contributions.

#### Author co-citation graphs

Author co-citation graphs represent the connections between researchers who are cited together in the references of other works. When two authors are frequently mentioned together in the same publications, it can be inferred that their contributions belong to a common theoretical or thematic framework. The intensity of co-citation reflects their shared influence in a field of study, allowing us to identify schools of thought, intellectual leaders, and the invisible structure of scientific trends. A temporal analysis of these graphs can reveal the evolution of paradigms and the emergence of new approaches.

#### Document co-citation graphs

These graphs map the relationships between publications that are simultaneously cited by other articles. Each node represents a document, and the edges symbolize its co-occurrence in the bibliographic sections of subsequent works. This structure allows us to identify the literary foundations of a field: the seminal articles that form the theoretical core, as well as the peripheral works that connect different areas. The density of connections around a document indicates its centrality in the construction of disciplinary knowledge.

**Co-citation graphs of countries and institutions**

At the macro level, co-citation graphs can be applied to countries or institutions, with nodes representing these entities and edges reflecting the frequency with which their research is cited together. This indicates a thematic or methodological proximity between their scientific systems. Two countries with high co-citation rates tend to specialize in similar research niches or collaborate intensively. These graphs help analyze international competitiveness, identify clusters of excellence, and design global positioning strategies in science and technology.

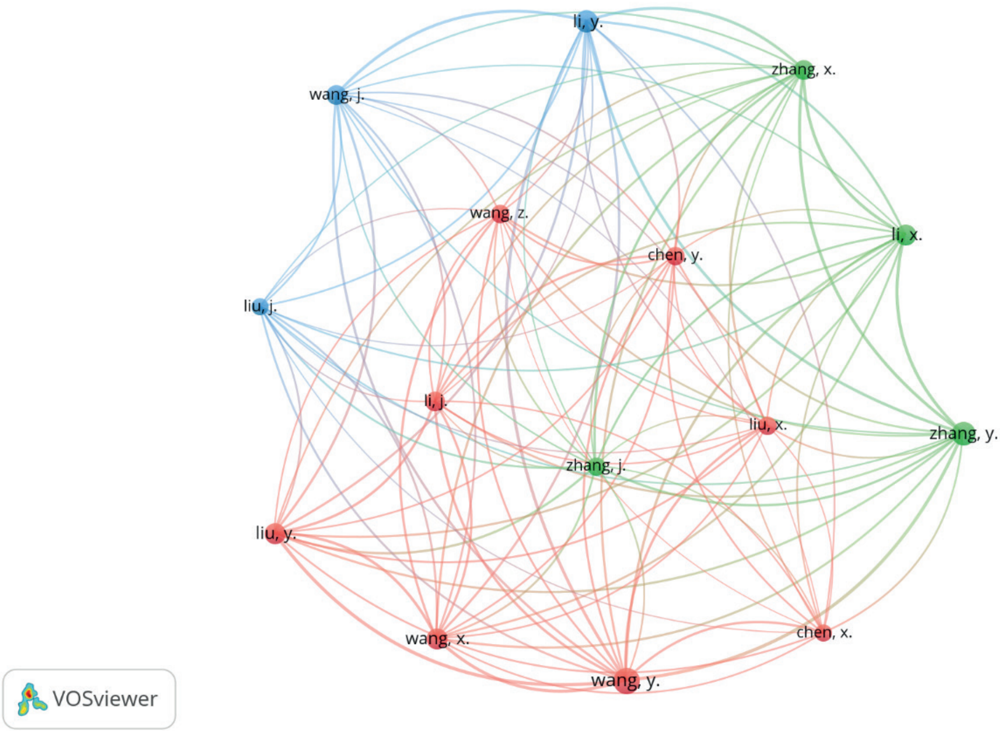


**Figure 6.3.** Example of a co-citation graph

### 6.2.4. Methodological integration and comparative perspectives

Specific epistemological, temporal, and pragmatic considerations should guide the selection between these algorithms. Co-occurrence analysis is optimal for exploring the immediate conceptual structure of a field, particularly in disciplines where specialized terminology efficiently encodes intellectual content. Bibliographic coupling offers decisive advantages for analyzing recent literature and identifying contemporary intellectual alignments. Co-citation analysis, on the other hand, provides historical depth and captures consolidated collective assessments of intellectual relationships.

The most sophisticated methodological approaches combine multiple algorithms in triangulated designs that mitigate the individual limitations of each method. A typical design might employ co-occurrence to identify emerging themes, bibliographic coupling to analyze their current structuring, and co-citation to contextualize their historical development. This multi-method integration allows us to distinguish between immediate conceptual connections (co-occurrence), contemporary intellectual alignments (bibliographic coupling), and consolidated perceptions of relationships (co-citation), thereby producing richer, methodologically robust bibliometric analyses.

A deep understanding of these algorithms, their theoretical foundations, practical implementations, and characteristic biases is an essential skill for any researcher who aspires to produce rigorous and meaningful bibliometric analyses. Far from interchangeable tools, each algorithm illuminates distinct dimensions of the complex topography of scientific knowledge, requiring conscious selection and contextualized application to realize its analytical potential fully.

### 6.3. Visual tools

The specialized tools used to construct graphs in the field of bibliometrics include VOSviewer, CiteNetExplorer, CiteSpace, R Bibliometrix, and Pybibx. However, the latter two are not specialized in graph construction. The procedure for each is as follows:

#### 6.3.1. VOSviewer: creation of thematic maps

1. Export complete bibliographic data from Scopus, WoS, or PubMed in compatible formats (RIS, CSV, EndNote).
2. Start VOSviewer and select “Create” → “Create a map based on bibliographic data.”
3. Choose the type of analysis according to the research objective:
   - Co-occurrence for analysis of terms and concepts.
   - Co-authorship for scientific collaboration networks.
   - Co-citation for citation-based intellectual structures.
   - Bibliographic coupling for relationships through shared references.

4. Configure threshold parameters according to data volume:
   - Minimum number of occurrences of a term (typically 5-10).
   - Minimum number of documents per author (2-5 for collaboration analysis).
   - Minimum number of citations per reference (10-20 for co-citation).

5. Apply the VOS clustering and visualization algorithm.
6. Customize the final visualization:
   - Adjust node size according to frequency or impact metrics.
   - Modify the color scheme to differentiate clusters.
   - Configure labels and zoom levels to optimize readability.
   - Apply network smoothing to reduce visual overlap.

The fundamental feature of VOSviewer lies in its visualization algorithm, which is based on multidimensional stress function minimization techniques, where the distance between nodes directly represents their similarity, calculated using normalized association measures. This tool generates maps where clusters emerge naturally as dense spatial groupings, differentiated chromatically to facilitate immediate identification. The visual representation prioritizes thematic interpretability over absolute metric precision, making VOSviewer the preferred choice for exploratory analysis and communication of results to non-specialist audiences.

#### 6.3.2. CitNetExplorer: citation map creation

1. Prepare complete citation data from Web of Science in standard export format
2. Load the data file via “File” → “Open Citation Network.”
3. Configure the time and impact filtering parameters:
   - Define the range of years to be analyzed.
   - Set the minimum number of citations for document inclusion.
   - Specify selection criteria for seminal documents.

4. Generate the complete citation network with all interconnections.
5. Apply specific exploration tools:
   - Use the zoom function to examine particular time periods.
   - Apply dynamic filters by number of citations or year of publication.
   - Use the connected components selection tool.

6. Perform citation trajectory analysis:
   - Identify foundational documents and their lines of descent.
   - Trace paths of intellectual development between publications.
   - Analyze patterns of thematic convergence and divergence.

CitNetExplorer stands out for its explicit temporal representation, in which the vertical axis unambiguously encodes publication years, creating a visual timeline that reveals the diachronic evolution of citation relationships. This specialized tool allows you to trace the historical development of scientific ideas through citation connections between publications, showing how concepts are transmitted, transformed, and branched over time. The ability to animate the evolution of the network year by year provides unique insights into patterns of intellectual inheritance and moments of paradigmatic change in scientific development.

**6.3.3. CiteSpace: Detection of Research "Bursts."**

1. Download and install CiteSpace with the updated Java Runtime Environment.
2. Create a new project and configure the directory structure for data and results.
3. Import and convert bibliographic data from WoS or Scopus to CiteSpace's native format.
4. Configure the temporal analysis parameters exhaustively:
   - Divide the study period into time segments (1-3 years recommended).
   - Establish selection criteria by percentile (Top 10 %, 20 %, etc.) or by top N per segment.
   - Define network pruning strategies (Pathfinder, Minimum Spanning Tree, or none).

5. Run the burst detection algorithm using:
   - Select entities for analysis (terms, references, authors).
   - Configuring Kleinberg algorithm parameters for detection sensitivity.
   - Specifying minimum thresholds for burst duration and intensity.

6. Generate and analyze integrated visualizations:
   - Interpret concentric rings of annual citations in nodes.
   - Identify thematic clusters through automatic tag analysis.
   - Analyze centrality metrics and connections for bridge nodes.

The uniqueness of CiteSpace lies in its ability to integrate multiple analytical dimensions into a single visualization: spatial position indicates similarity relationships, concentric rings show temporal citation patterns, colors differentiate thematic clusters, and red rings highlight burst periods of citation activity. This tool uses specialized algorithms to detect turning points in scientific literature, identifying not only emerging topics but also specific moments of acceleration in research attention. The resulting visual representation provides a dynamic map of scientific evolution that simultaneously captures spatial structure, temporal development, and intellectual turning points.

### 6.3.4. Graphs with PyBibx

Both R Bibliometrix and Pybibx offer functionality for constructing co-authorship graphs, each with its own algorithmic details. Pybibx is characterized by implementing a more straightforward construction procedure with flexible variants. One of these representations generates a graph in which nodes representing individual articles, referenced by their unique IDs, interact without explicit edges. This alternative visualization enables analysis of the proximity or coexistence of publications within a conceptual or temporal space defined by the researcher, offering a different perspective from that of a traditional network of direct links.

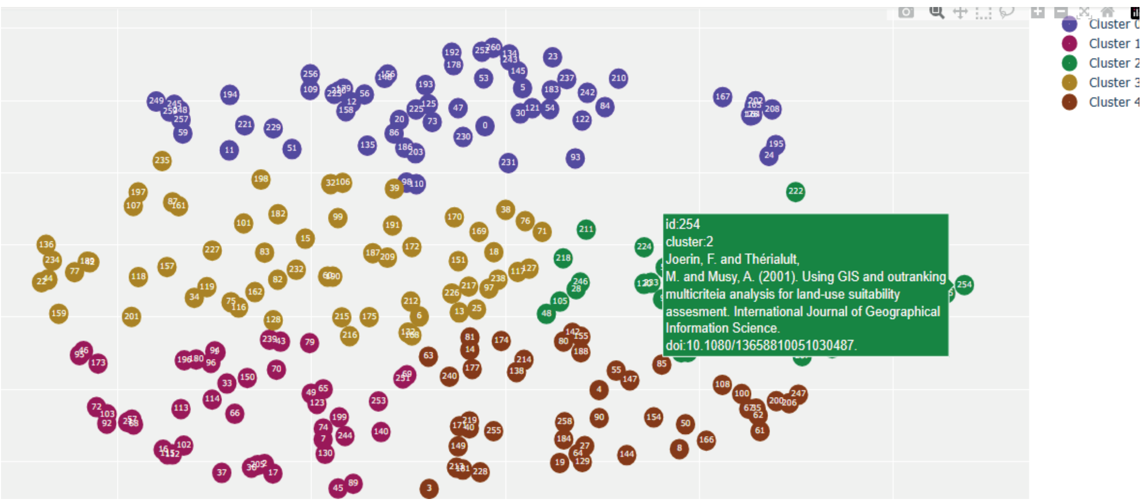


**Figure 6.4.** Example of a co-occurrence graph of terms or words

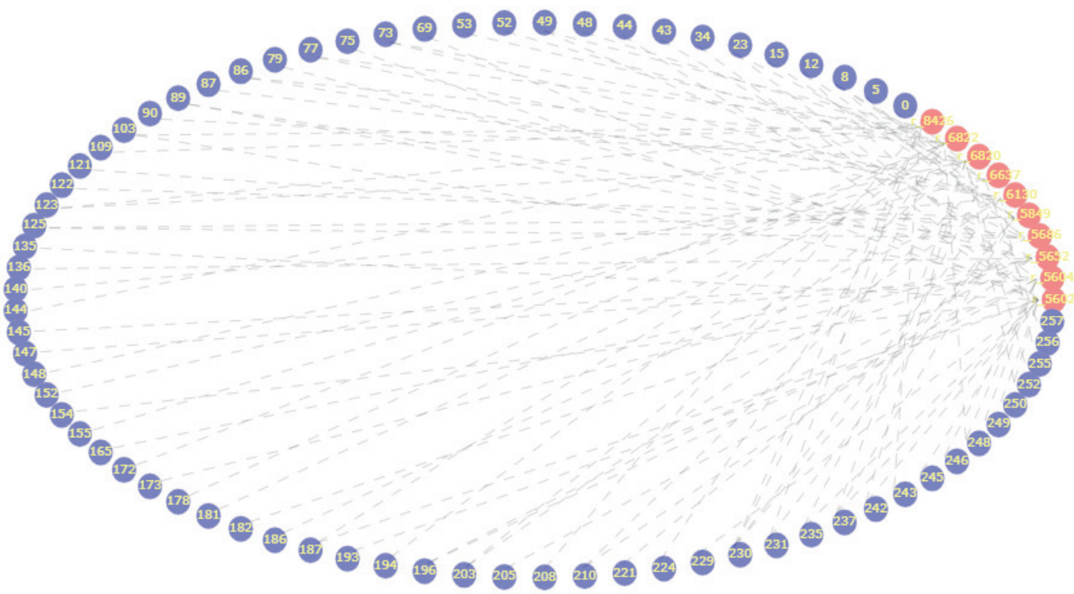

**Figure 6.5.** Example of a co-occurrence graph of terms or words

Beyond edgeless representations, these tools also allow the generation of classic relational graphs. In these graphs, the connections or edges between nodes are explicitly defined, representing citation or co-citation links between individual articles. Unlike maps that group elements into thematic clusters, this visualization focuses on showing the network of direct connections and its structure in its purest form, without applying clustering algorithms. This allows the analyst to identify connection patterns, bridge articles, or network density without the influence of prior automatic clustering.

One of the most eloquent visualizations of the spatial dimension of research is a graph superimposed on a geopolitical map. In this representation, nodes are positioned at the geographic locations of the authors' affiliations, illustrating the global distribution of scientific production. The edges connecting these institutions represent co-authorship links, allowing for immediate visualization of the flows and intensity of international collaboration. This graph is indispensable for identifying centers of scientific gravity, patterns of regional cooperation, and the geographical projection of research networks, offering an extremely valuable layer of contextual analysis.

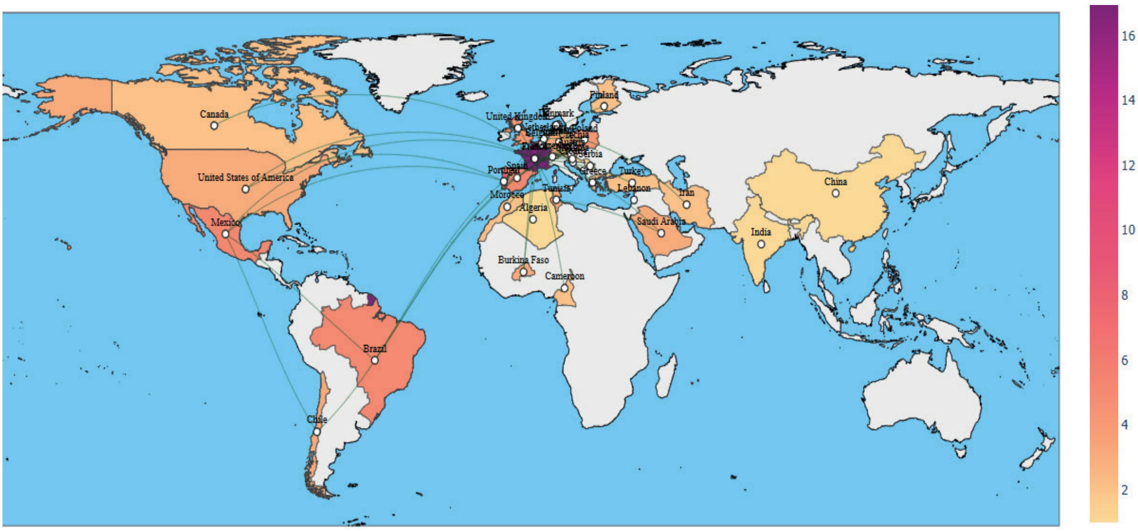


**Figure 6.5.** Example of a co-citation graph with the geographical distribution of authors

PyBibx offers other graph types that can be consulted in its official documentation and examples.

### 6.3.5. Graphs with R Bibliometrix

R Bibliometrix can create graphs of author co-citations, bibliographic links between documents, and collaboration between countries and institutions based on co-authorship relationships. To create graphs in R Bibliometrix, proceed as follows:

1. Install packages:

*install.packages("bibliometrix")*
*install.packages("tidyverse")*

2. Load libraries

*library(bibliometrix)*
*library(tidyverse)*

3. Load and process data:

```
File <- "file.bib"
M <- convert2df(file, dbsource = "wos", format = "bibtex")
```

4. Create a co-occurrence network:

```
NetMatrix <- biblioNetwork(M, analysis = "co-occurrence",
                   network = "author_keywords",
                   sep = ";")
```

5. Create graph:

```
net <- networkPlot(NetMatrix,
             n = 50,  # Number of terms to display
             type = «fruchterman",  # Layout
            Title = «Term Co-occurrence - Author Keywords»,
             labelsize = 0,8,
             size = 5,
             remove.isolates = TRUE,
             cluster = «walktrap")  # Clustering method
```

All of these are displayed in the plot panel or can be exported to an image using:

```
png("term_co-occurrence.png", width = 1200, height = 900, res = 150)
print(net)
```

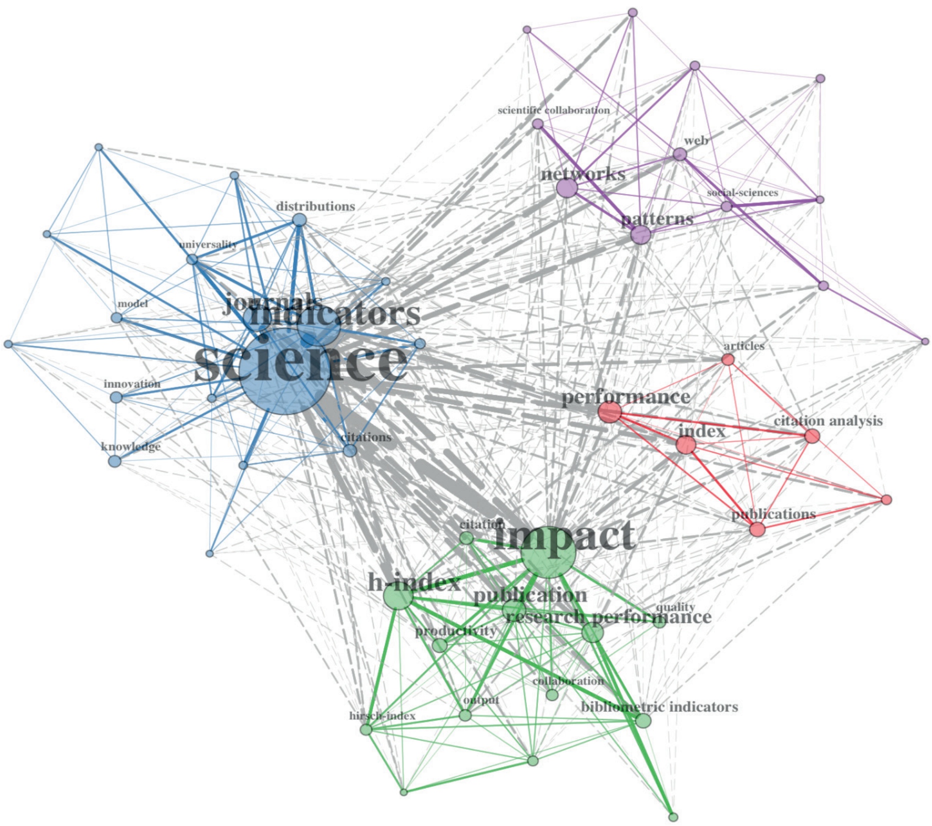


**Figure 6.6.** Example of a co-occurrence graph

## 6.4. Network interpretation

### 6.4.1. Identification of thematic clusters

The identification of thematic clusters represents the analytical process by which conceptual communities within a bibliometric network are discovered and delimited. These clusters emerge naturally from community detection algorithms that identify subgroups of nodes with high internal connectivity and relatively low external connectivity. In practice, each cluster encapsulates a coherent thematic domain, where nodes, whether terms, authors, or publications, share significant semantic, methodological, or theoretical characteristics. The interpretation of these clusters requires a multidimensional analysis that considers not only the internal composition of the group but also its relationships with other clusters and its position within the overall network structure.

The interpretation process begins with an examination of the representative labels that algorithms automatically assign to each cluster, typically derived from the most frequent or central terms within each group. However, this automatic approach must be complemented by a qualitative assessment that examines representative publications from each cluster to understand their substantive content.

The validation of internal thematic coherence is carried out by analyzing the foundational documents, those with the most extraordinary centrality within the cluster, and identifying the core concepts that define the thematic identity of the group. This dual quantitative and qualitative approach allows us to transcend mere structural description to achieve a deep understanding of the intellectual meanings encapsulated in each grouping.

The relative position of clusters within the global map provides crucial information about the structure of the field of study.

Spatially proximate clusters typically share conceptual frameworks, methodologies, or applications, whereas distant clusters represent distinct intellectual traditions or specialized fields. The connections between clusters, manifested as bridge links, point to potentially fertile thematic interfaces for interdisciplinary research. Diachronic analysis of the evolution of these clusters reveals dynamics of disciplinary fragmentation, paradigm fusion, or the emergence of new hybrid fields, providing valuable insights into the trajectory of knowledge development in the domain under study.

### 6.4.2. Centrality vs. density

Centrality and density are two fundamental analytical dimensions that capture complementary aspects of the structure and dynamics of bibliometric networks. Centrality measures the strategic position of a node within the global network, identifying elements that function as connectors between different regions of the graph. In contrast, density quantifies the degree of internal interconnection within a specific cluster or subnetwork, reflecting the cohesion and mature development of a thematic community. The joint interpretation of these metrics allows for a sophisticated characterization of the field of study's intellectual architecture and the specific roles that different elements play within that cognitive ecosystem.

In analytical practice, centrality manifests itself in multiple forms, each revealing different aspects of structural influence.

Degree centrality identifies nodes with many direct connections, typically fundamental concepts or highly collaborative authors. Intermediation points to nodes that connect different clusters, functioning as conceptual bridges between separate thematic communities. Closeness detects nodes that can quickly reach the rest of the network, indicating concepts or authors with broad diffuse influence. For example, in a co-word map of artificial intelligence, terms such as “machine learning” would show high centrality.

At the same time, “explainable AI” could exhibit high intermediation by connecting clusters of computational ethics and learning algorithms.

Density, on the other hand, characterizes the internal development of thematic clusters. High-density clusters, with numerous internal connections, represent mature, highly structured fields in which the constituent concepts have well-defined, stable relationships. Low-density clusters suggest emerging or developing areas, where conceptual relationships are still incipient, and the internal structure is still forming. In an authorship analysis, a dense cluster would indicate a consolidated collaboration group with multiple joint projects.

In contrast, a sparse cluster would indicate an emerging collaboration network with more sporadic or bilateral interactions.

The combination of these dimensions in strategic matrices, such as centrality versus density analysis, provides a robust framework for the typological characterization of clusters and nodes. For example, driving themes appear as clusters with high density and centrality, representing consolidated areas that drive the field’s development. Niche topics show high density but low centrality, indicating mature but isolated specializations.

Emerging topics exhibit low density but high centrality, signaling bridge concepts with potential for future development. Peripheral topics have low density and low centrality, representing incipient specializations or areas in decline. This typology enables strategic prioritization of research areas based on specific scientific development or research policy objectives.

In this interpretation phase, technical mastery of bibliometrics is necessary but not sufficient to extract substantive meaning from the networks generated. The researcher must have in-depth knowledge of the specific domain of study to establish meaningful connections between seemingly disparate terms and recognize conceptual relationships that transcend mere statistical coincidences. This specialized disciplinary understanding allows one to discern between superficial terminological associations and deep intellectual links, between passing terminological fads and foundational concepts with actual structuring capacity. Thematic expertise thus becomes the indispensable lens through which quantitative patterns acquire qualitative meaning and intellectual relevance.

The capacity for abstraction emerges as a critical competence for transcending immediate visualization and constructing mental models that explain the underlying architecture of the knowledge represented.

This ability allows fragmentary findings to be integrated into coherent narratives about the structure and dynamics of the field of study, identifying not only what the networks explicitly show but also what they implicitly suggest through their gaps, asymmetries, and relational patterns. The researcher must constantly move between the micro level of individual terms and specific connections, the meso level of thematic clusters and their interrelationships, and the macro level of the field’s overall structure, synthesizing perspectives across multiple scales of analysis into an integrated, hierarchically organized understanding.

The final interpretation, therefore, represents a creative synthesis where metric rigor is combined with disciplinary sensitivity and the researcher's inferential capacity. This process transforms relational data into actionable knowledge, identifying research opportunities, revealing fertile interfaces between specialties, and proposing explanatory narratives about the field's evolution and current state. The quality of this interpretation depends fundamentally on the researcher's ability to exercise informed judgment, contextualize findings within broader intellectual traditions, and communicate complex perspectives in an accessible manner without sacrificing analytical depth or conceptual precision.

An example of a graph and its interpretation is as follows:

Context: search result in SCOPUS for: *( TITLE-ABS-KEY ( cardiolog* OR cardiac OR heart OR coronary OR "myocardial infarction" OR arrhythmia OR echocardiography OR hypertension ) ) AND ( AFFILCOUNTRY ( Cuba )*

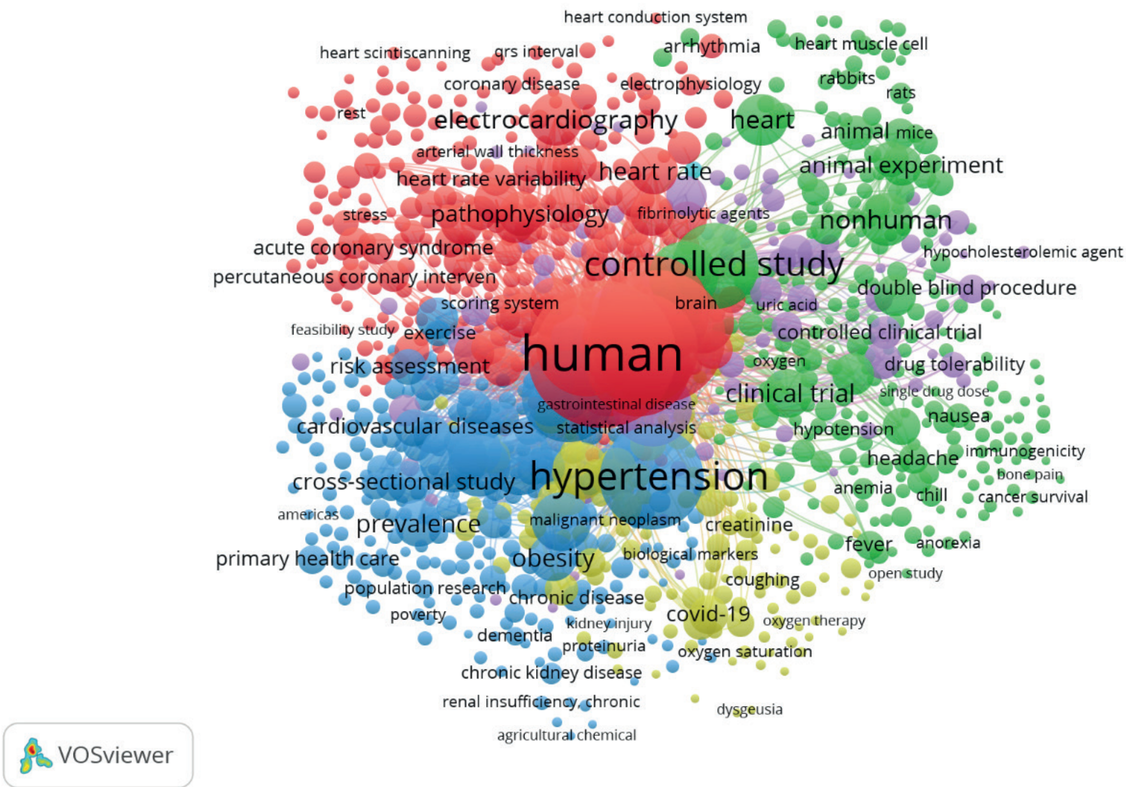


**Figure 6.6.** Example of a co-occurrence graph

**Possible interpretation**

Red cluster (top left): this cluster focuses on clinical cardiology and electrophysiology. It includes terms such as "heart," "electrocardiography," "arrhythmia," "heart rate," "coronary disease," and "acute coronary syndrome." This shows a focus on research into coronary heart disease and arrhythmias, and on the use of diagnostic techniques such as electrocardiograms. The connections to "human" and "controlled study" demonstrate that these studies are conducted on patients.

Blue cluster (bottom left): this cluster focuses on the epidemiology and risk factors of cardiovascular diseases. Key terms are “hypertension,” “cardiovascular diseases,” “prevalence,” “cross-sectional study,” “primary health care,” and “obesity.” This indicates a strong interest in public health, disease prevalence, and their relationships with other risk factors in the population.

Green cluster (upper right): this cluster is related to experimental and pharmacological research. The terms “nonhuman,” “animal experiment,” “rats,” “rabbits,” “double blind procedure,” “drug tolerability,” and “drug efficacy” are prominent. This suggests a line of research that uses animal models and controlled clinical trials to test the efficacy and safety of new treatments.

Yellow cluster (center): this is the core of the research. Although not as large, it contains the most central terms such as “human,” “controlled study,” “clinical trial,” and “statistical analysis.” This cluster serves as the connector among the others, confirming that cardiology research in Cuba is primarily based on clinical and controlled human studies.

In summary, the graphs reveal that cardiology research in Cuba is multifaceted, with a strong emphasis on clinical research into cardiovascular disease and hypertension. Topics such as epidemiology, electrophysiology, and treatment evaluation are addressed, along with an experimental research line in animal models.

**Recap**

- Correlation graphs are visual representations of relationships between variables or bibliometric elements (e.g., authors, keywords, institutions, journals, etc.).
- They are based on network theory and enable the analysis of the structure and dynamics of scientific knowledge.
- A graph is composed of nodes (entities) and edges (links) that reflect relationships or associations between these elements.
- In bibliometrics, nodes can represent authors, articles, keywords, or countries, and edges can represent their degree of correlation, co-occurrence, or co-citation.
- Graph analysis allows us to identify patterns of collaboration, thematic affinity, and cognitive structures in scientific production.
- The strength of correlation between two nodes is quantified using statistical measures such as Pearson’s or Spearman’s correlation coefficient.
- There are different types of networks:
    - Co-authorship networks (collaboration between researchers).
    - Co-citation networks (articles cited together).
    - Co-word or co-occurrence networks (terms that appear together).

- Graphs can be constructed from correlation or similarity matrices derived from bibliographic data.
- The density of a network indicates the overall degree of connection between nodes.
- Centrality (degree, betweenness, closeness) measures the relative importance of a node within the network.
- Clusters or communities are groups of highly interconnected nodes that usually correspond to scientific topics or subfields.
- Modularity analysis allows these communities to be identified using clustering algorithms.
- The most commonly used tools for generating correlation graphs are VOSviewer,

Gephi, CiteSpace, BibExcel, and Pajek.

- VOSviewer constructs similarity maps and calculates distances between elements based on their degree of co-occurrence.
- Gephi offers advanced interactive visualization and structural analysis functions for large networks.
- The color and size of the nodes usually represent the intensity of correlation and the weight of the connections.
- Graphs allow us to observe the temporal evolution of topics and the emergence of new research areas.
- Proper interpretation requires combining quantitative (statistical) and qualitative (semantic) analysis of the results.
- Correlation graphs contribute to scientific monitoring, the identification of opinion leaders, and the detection of thematic gaps.
- Their correct application requires methodological rigor, careful data selection, and ethical and comprehensible visualizations.

**Self-assessment questions**

1. What do the nodes and edges in a correlation graph represent?
2. What type of information can be analyzed using graphs in bibliometric studies?
3. What is the difference between a co-authorship network and a co-citation network?
4. What does the correlation coefficient measure in the construction of a graph?
5. What does the density of a bibliometric network indicate?
6. What types of centrality exist, and what does each one represent?
7. What does the existence of a cluster in a correlation graph mean?
8. What tools allow correlation graphs to be constructed and scientific networks to be visualized?
9. What information is conveyed by the color and size of the nodes in a graph?
10. Why is it essential to combine quantitative and qualitative analysis when interpreting bibliometric networks?

## BIBLIOGRAPHY

1. Newman MEJ. Networks: An introduction. Oxford: Oxford University Press; 2010. ISBN: 9780199206650.

2. Barabási AL. Network science. Cambridge: Cambridge University Press; 2016. ISBN: 9781107076266.

3. Börner K, Chen C, Boyack KW. Visualizing knowledge domains. In: Cronin B, Sugimoto CR, editors. Beyond bibliometrics: Harnessing multidimensional indicators of scholarly impact. Cambridge (MA): MIT Press; 2014. p. 197–228. doi: 10.7551/mitpress/9780262026792.003.0010.

4. Chen C. Mapping scientific frontiers: The quest for knowledge visualization. 2nd ed. London: Springer; 2017. doi: 10.1007/978-3-319-32043-4.

5. Leydesdorff L, Welbers K. The semantic mapping of science: Co-words, co-authors, and co-citations. J Informetrics. 2011;5(1):1–14. doi: 10.1016/j.joi.2010.10.002.

6. Hanneman RA, Riddle M. Introduction to social network methods. Riverside (CA): University of California; 2005. https://faculty.ucr.edu/~hanneman/nettext/.