Unveiling Research Trends: A Practical Workshop on Bibliometric Analysis with Bibliometrix

Discover, Analyze, and Visualize Scientific Literature

By:

Radhika Gedela

Data Visualization Intern

Indiana Center for Musculoskeletal Health, IUI

Under the Guidance of:

Thomas O'Connell

What is Bibliometric Analysis?

A quantitative approach to analyze scientific literature by evaluating publication patterns, citation relationships, and research trends.

Key Metrics:

- **Citation Analysis:** Measures the influence of papers through citation counts.
- **Co-authorship Analysis:** Identifies collaboration patterns among authors and institutions.
- **Keyword Analysis:** Highlights core research themes and emerging topics.
- **Network Mapping:** Visualizes relationships between authors, keywords, and institutions.

Importance of Bibliometric Analysis

Quantitative Evaluation
Data-driven and objective framework to assess structure & dynamics of scientific research.

Trend Identification
Detects emerging topics, key themes, & evolving research trajectories

Research Impact Assessment
Measures influence of publications, authors, & institutions to gauge contribution to the field.

Collaboration Insights
Maps co-authorship & institutional partnerships, fostering strategic alliances in research.

Strategic Resource Allocation
Informs funding agencies & policymakers on prioritizing impactful research areas.

Field Development
Guides researchers in identifying underexplored areas, enabling innovation and discovery.

Tools for Bibliometric Analysis

Popular Tools in Bibliometric Analysis:

- **Bibliometrix (R Package):** Comprehensive tool for quantitative research, offering diverse features for data analysis and visualization.
- **VOSviewer:** Specialized in network visualizations like co-authorship and citation mapping.
- **CiteSpace:** Focuses on trend analysis and cluster detection in scientific literature.

Why Choose Bibliometrix

Open Source: A free and readily available package, providing powerful functionalities for bibliometric analysis.

Comprehensive Functionality: Includes a wide range of tools for data import, cleaning, analysis, and visualization.

User-Friendly Interface: Equipped with a user-friendly syntax, facilitating the implementation of bibliometric analyses.

Collecting Data from Web of Science

1

Data Collection:

- Open Web of Science and navigate to the advanced search.
- Enter the keywords: **bone AND muscle AND metabolism**

2

Search Criteria:

- Exclude **review** articles.
- Limit the publication range to the **last 5 years.**

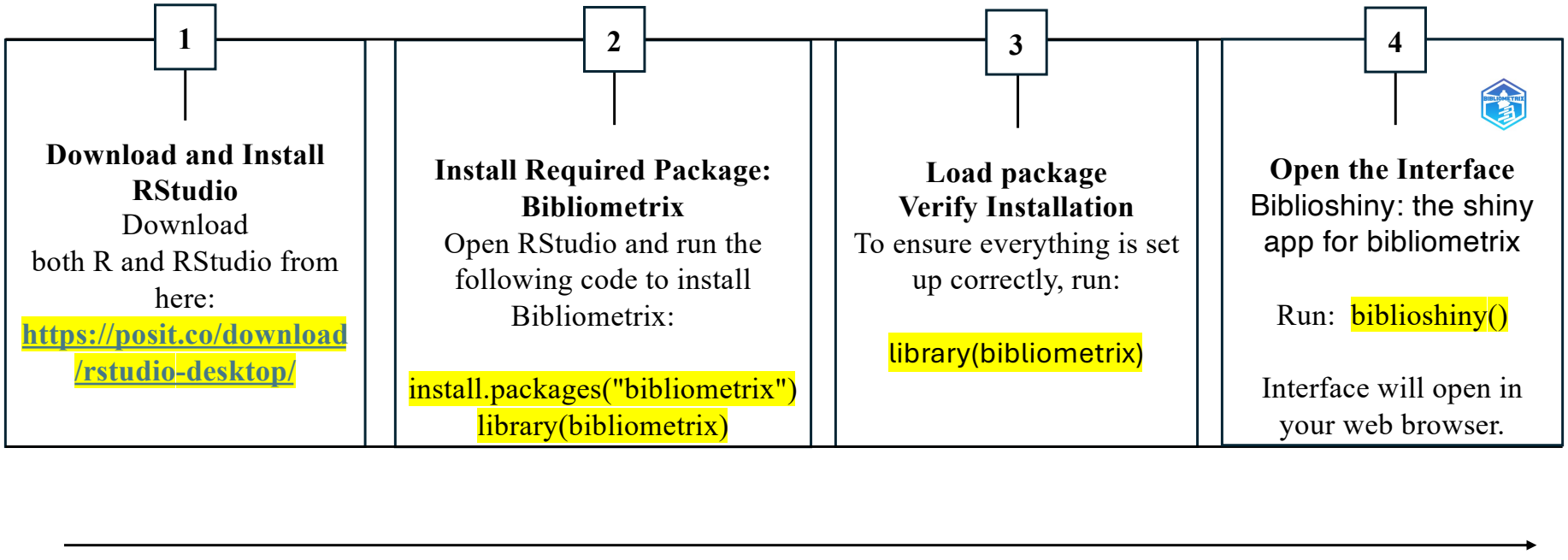
} This results in articles

3

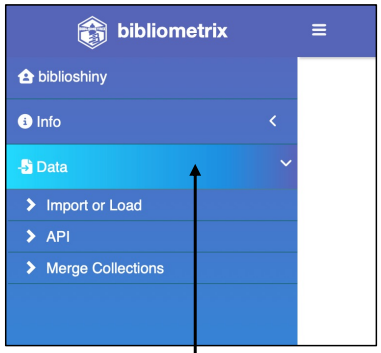
Export Settings:

- Export data in **plain text** format.
- Select **"Full Record and Cited References"** for detailed metadata.
- Download a maximum of 500 articles at a time due to Web of Science limits.

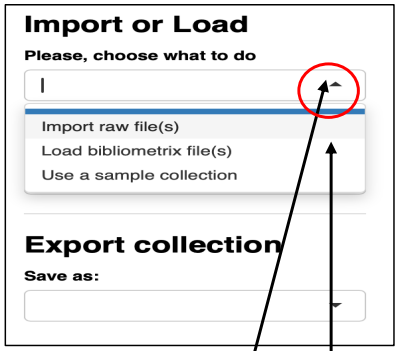
Setting up the R and RStudio Environment



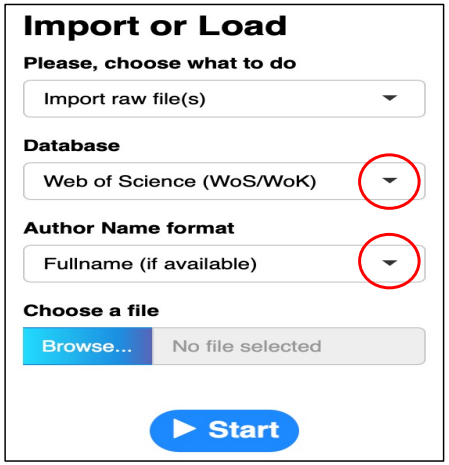
Importing Data into Bibliometrix



Go to Data Tab – click the **Data** tab.



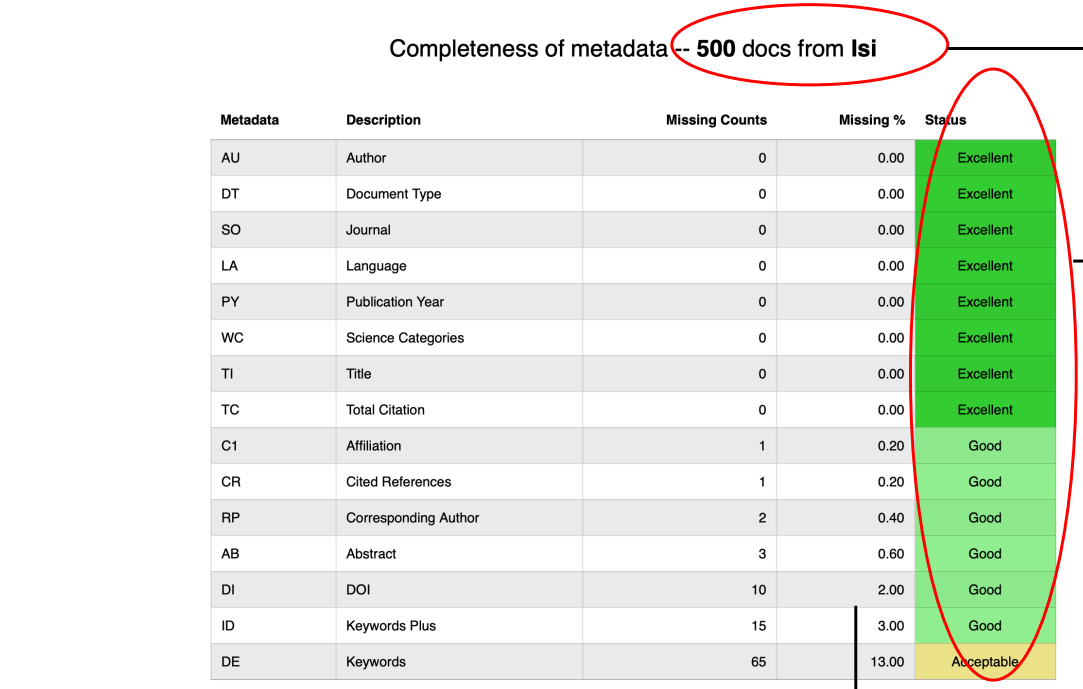
In the Data tab, click on the dropdown menu for **Import** or **Load**.



- Select **Choose Import Raw Files** from the dropdown.
- **Database**: Choose **Web of Science**.
- **Author Name Format**: Select **Full Name** (if available).
- **Choose a File**: Select the file you downloaded earlier from Web of Science.
- Once the file is uploaded, click **Start the Analysis** to begin your bibliometric analysis.

Metadata Completeness Results

After importing the file, a popup will display metadata completeness details



Completeness of metadata -- **500** docs from **Isi**

Metadata	Description	Missing Counts	Missing %	Status
AU	Author	0	0.00	Excellent
DT	Document Type	0	0.00	Excellent
SO	Journal	0	0.00	Excellent
LA	Language	0	0.00	Excellent
PY	Publication Year	0	0.00	Excellent
WC	Science Categories	0	0.00	Excellent
TI	Title	0	0.00	Excellent
TC	Total Citation	0	0.00	Excellent
C1	Affiliation	1	0.20	Good
CR	Cited References	1	0.20	Good
RP	Corresponding Author	2	0.40	Good
AB	Abstract	3	0.60	Good
DI	DOI	10	2.00	Good
ID	Keywords Plus	15	3.00	Good
DE	Keywords	65	13.00	Acceptable

500 documents successfully imported with complete metadata.

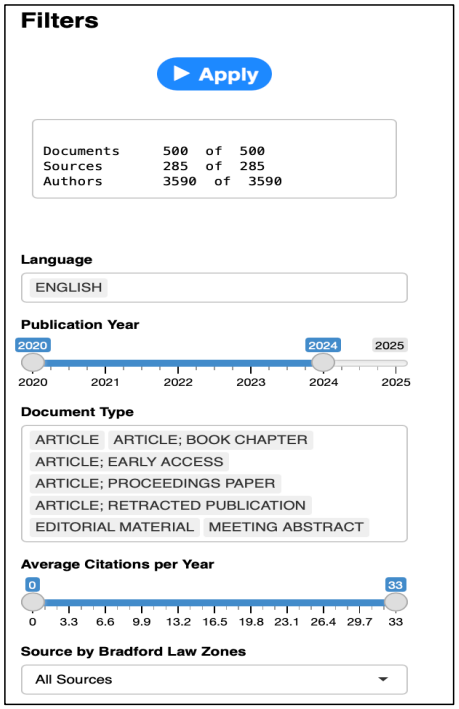
Assesses data quality based on the percentage of missing values

- **Excellent**: Data quality is high, ensuring accurate and reliable analysis.
- **Good:** Data quality is acceptable but may slightly impact results.
- **Acceptable:** Some impact on results; moderate quality.
- **Poor:** Data quality is low, potentially affecting the validity of analysis and insights.

Descriptive fields in the dataset

Percentage of documents

Document Filtering Options in Bibliometrix



- **Language**: Filter documents by publication language.
- **Publication Year**: Select a specific range of publication years to focus the analysis on.
- **Average Citations/Year**: Filter documents by their average citation count over time, helping identify influential papers.
- **Source by Bradford's Law**: Categorize sources into zones based on citation impact.

Features in Bibliometrix

- **Advice**: Get guidance for interpreting visualizations.
- **Report**: Generates a summary report of the analysis and key insights that can be downloaded as an excel file.
- **Save**: Save visualizations as images or files into the PC.
- **Close**: Exit the visualization window.

Comprehensive Set of Functions of Bibliometrix

Bibliometric Analysis

Improvements in Data-analysis and Data-viz through Science mapping

Often, new knowledge emerges at crossroad among structures and time evolution

Focus on relevance

- Main Data
- Annual Scientific Production

Levels of analysis

metrics

Level of Analysis	Metrics
Sources	Bradford's Law H index Source dynamics Most relevant sources
Authors	Most relevant authors Annual production per author Lotka's Law H index Most relevant affiliations Countries
Documents	Most cited doc Cited ref Words: ID, DE, TI, AB

K structures

Structure	Bibliometric Technique	Unit of Analysis	Statistical Techniques
Conceptual	• Co-word	• ID, DE *(keywords)* • TI • AB • Full document	• Network Analysis • Factorial Analysis (CA; MCA; MDS) • Thematic mapping • Thematic evolution • Topic modeling
Intellectual	• Co-citation • Citation	• References • Authors • Journals	• Network analysis • Histograph
Social	• Collaboration	• Authors (co-authorship) • Institution • Journal	• Collaboration network

Overview of the Data

1. Main Information

- Summarizes the dataset with key metrics like total documents, authors, and sources.
- Provides a foundational understanding of the dataset's scope and scale.

2. Annual Scientific Production

- Displays the number of publications produced each year, highlighting trends over time.

3. Average Citations Per Year

- Shows the average number of citations received by publications annually, reflecting research impact.

4. Three-Field Plot

Left Field (Authors): Identifies key researchers in the field.
Middle Field (Keywords): Connects researchers to primary topics studied.
Right Field (Sources/Journals): Shows where research is published, linking authors and topics to specific journals.

- The Three-Field Plot offers a comprehensive view of the relationships among researchers, research topics, and publication venues, enabling better understanding and strategic planning in research.

Analysis of Sources

1. Most Relevant Sources

- Lists the journals with the highest number of publications in the dataset.

2. Most Local Cited Sources

- Highlights journals that are frequently cited within the dataset, indicating influence within the field.

3. Sources' Local Impact

- Measures the impact of journals based on their citation frequency in the dataset.

4. Sources' Production Over Time

- Visualizes trends in the number of publications from key journals over the selected time period.

Analysis of Authors

1. Most Relevant Authors

- Lists authors who contributed the most publications in the dataset.

2. Most Local Cited Authors

- Highlights authors with the highest citation counts within the dataset.

3. Authors' Production Over Time

- Shows trends in authors' publication activity across the years.

4. Author Productivity Through Lotka's Law

- Evaluates the distribution of productivity among authors, revealing a small number of highly prolific contributors.

5. Authors' Local Impact

- Measures the citation impact of authors within the dataset.

Analysis of Affiliations

1. Most Relevant Affiliations

- Identifies institutions with the highest number of contributions in the dataset.

2. Affiliations' Production Over Time

- Displays trends in publications produced by top institutions over the years.

Analysis of Countries

1. Corresponding Author's Countries

- Highlights countries of the corresponding authors, showcasing global participation.

2. Countries' Scientific Production

Lists countries with the most publications in the dataset.

3. Countries' Production Over Time

- Visualizes trends in research output by countries over time.

4. Most Cited Countries

- Indicates countries whose publications have the highest citation counts, reflecting their influence in the field.

Analysis of Documents & References

1. Documents

Most Global Cited Documents

- Highlights the papers with the highest citation counts worldwide, indicating their broad influence.

Most Local Cited Documents

- Focuses on documents most frequently cited within the dataset, showing their relevance to the research area.

2. Cited References

Most Local Cited References

- Lists references that are most frequently cited in the dataset.

Reference Spectroscopy

- Visualizes the distribution and patterns of cited references.

Analysis of Words or Keywords

1. Most Frequent Words

- Identifies the most commonly used words in the dataset, revealing core research themes.

2. WordCloud

- A graphical representation of word frequency, where larger words indicate higher importance.

3. TreeMap

- Visualizes word frequency in a hierarchical structure, showing the prominence of topics.

4. Words' Frequency Over Time

- Tracks how the usage of specific keywords has changed over the years.

5. Trend Topics

- Highlights emerging and declining topics based on keyword analysis.
- Each dot corresponds to a specific term and a year. The size of the dot reflects the **frequency** of the term in the publications for that year.
- **Larger dots** indicate higher usage or importance of the term during that year.
- **Smaller dots** reflect less frequent usage.
- Horizontal lines connected to the terms indicate the **time span** during which a term was actively used in publications

Clustering

- **Clustering by coupling:**

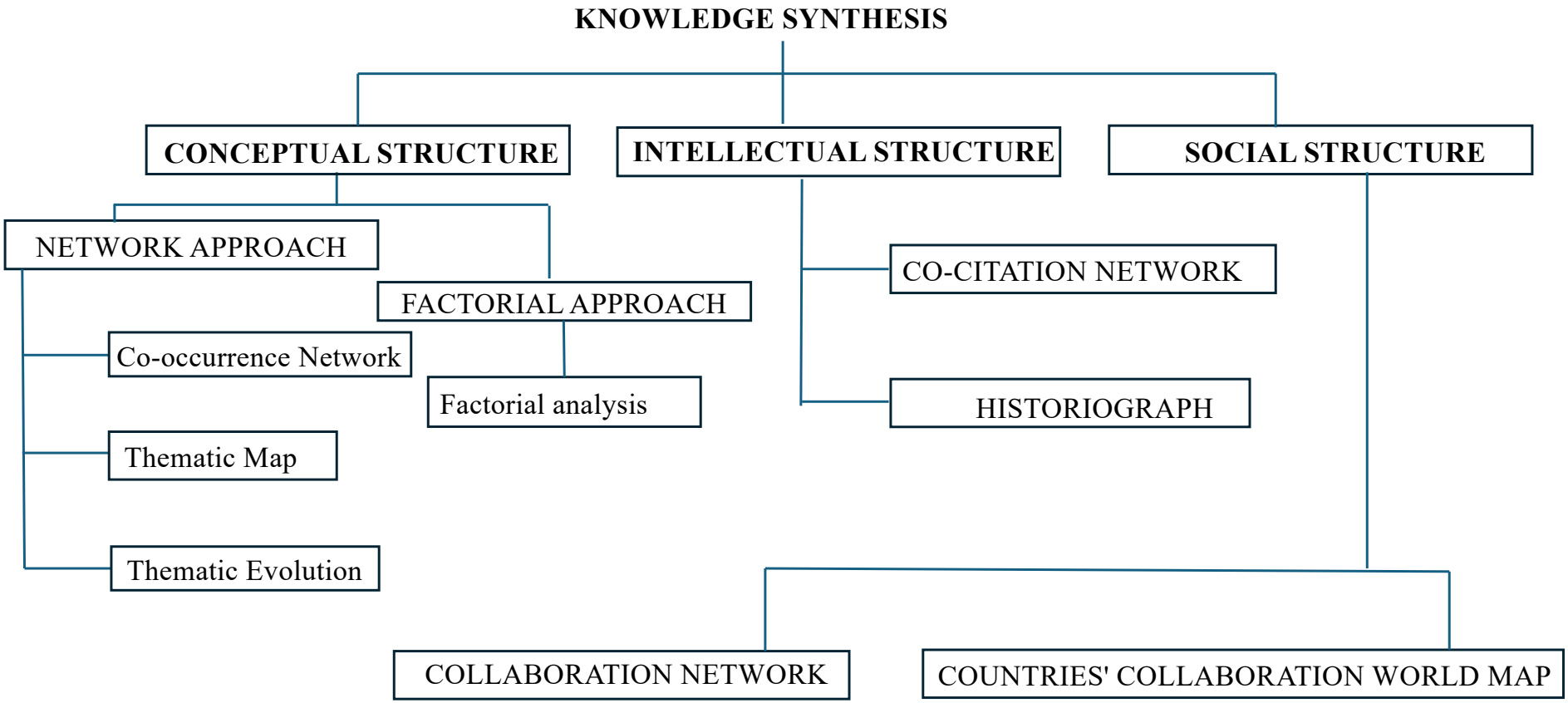
It groups research papers based on shared references. If two documents frequently cite the same sources, they are considered "coupled" and likely belong to the same research theme or cluster.

- **How it Works in Biblioshiny?**

1.Analyzes references: Calculates shared citations between pairs of documents.

2.Groups related papers: Creates clusters based on these shared references.

3.Clusters = Research Themes: Each cluster represents a distinct research area.


KNOWLEDGE SYNTHESIS
CONCEPTUAL STRUCTURE
INTELLECTUAL STRUCTURE
SOCIAL STRUCTURE
NETWORK APPROACH
FACTORIAL APPROACH
CO-CITATION NETWORK
Co-occurrence Network
Factorial analysis
HISTORIOGRAPH
Thematic Map
Thematic Evolution
COLLABORATION NETWORK
COUNTRIES' COLLABORATION WORLD MAP

Co-occurrence Network

- A graph-based visualization showing relationships between items (e.g., keywords, authors, or concepts) based on their co-occurrence in a dataset.

Interpretation

Clusters = Research Topics

Groups of interconnected nodes are color-coded, showing distinct research themes.

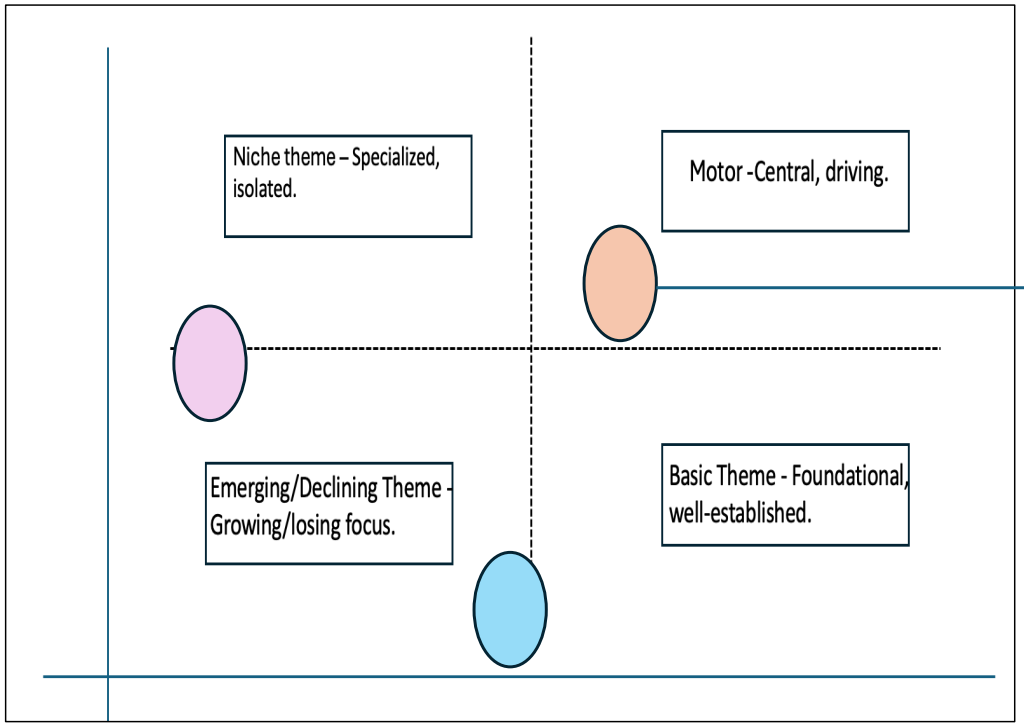
Strength of Connection

Thicker edges indicate stronger relationships or higher co-occurrence frequency.

Central Nodes

Nodes with higher centrality are key topics or concepts within the dataset.

Thematic map



•**Bubble Represents Cluster**: Each bubble is a network cluster.

•**Bubble Names**: Words represent key terms in the cluster.

•**Bubble Size**: Proportional to word occurrence

•.**Bubble Position**: Determined by **Callon Centrality** and **Density**.

•**Centrality**: Importance or connectivity.

•**Density**: Level of development.

Thematic Evolution

- Thematic evolution studies the changes and progression of research topics over time by dividing the time span into distinct **time slices** to analyze topic trajectories.
- **Time Slices:**
- We split the recent research timeline into **three time slices** with **three cutting points**:
- **2021**
- **2022**
- **2023**
- This allows for an analysis of how research themes have emerged, grown, or declined across these periods.

Factorial Analysis

- Factorial analysis uses data reduction techniques to identify subfields and patterns in research. These techniques simplify complex data by reducing dimensions while retaining meaningful relationships between variables (e.g., keywords).

Interpretation

- Words that are close together indicate shared research focus (high co-occurrence in articles).
- Words that are distant suggest they are infrequently treated together in research.

Co-citation Network

- Co-citation analysis identifies relationships between documents by examining how often they are cited together in other research papers.

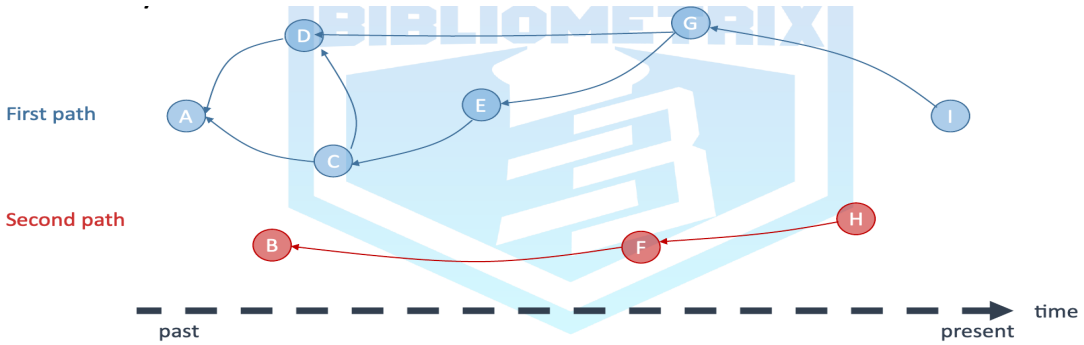
Example:

- Co-citation occurs when **two documents (A & B)** are cited together by other documents (C, D, E).
- Example:
- Document A and Document B are co-cited if Articles C, D, and E reference both A and B.

Historiograph

Historiographic mapping is a visualization method used to analyze the evolution of research topics over time by tracing citation patterns among key documents.

- **Nodes**: Represent documents cited by other papers in the dataset.
- **Edges**: Show direct citations between documents (e.g., A cited by D).
- **Graph Orientation**:
- **Horizontal Axis**: Represents publication years, showing the timeline of research evolution.
- **Node Placement**: Tracks the historical development of topics over time.



Collaboration Network

- A collaboration network visualizes the relationships between authors, institutions, or countries working together on research.

Key Concepts of Collaboration Network

- **Nodes:** Represent entities such as authors, institutions, or countries.
- **Edges:** Show collaboration between entities (e.g., co-authored papers).
- **Network Properties:**
- **Node Size:** Indicates the number of collaborations or contributions by the entity.
- **Edge Thickness:** Reflects the strength or frequency of collaboration.

QUESTIONS?

THANK YOU!