

Contributors, Topics, and Future Trends: A Deep Dive into 20 Years of OpenStreetMap Research

Prerequisite

Install software

Download and install the following software from the official websites, following the official guidelines to ensure correct installation and functionality.

Microsoft Excel: <https://www.microsoft.com/en-us/microsoft-365/excel>

Vosviewer: <https://www.vosviewer.com/>

Biblioshiny: <https://www.bibliometrix.org/home/index.php/layout/biblioshiny>

Python → jupyter notebook: <https://jupyter.org/install>

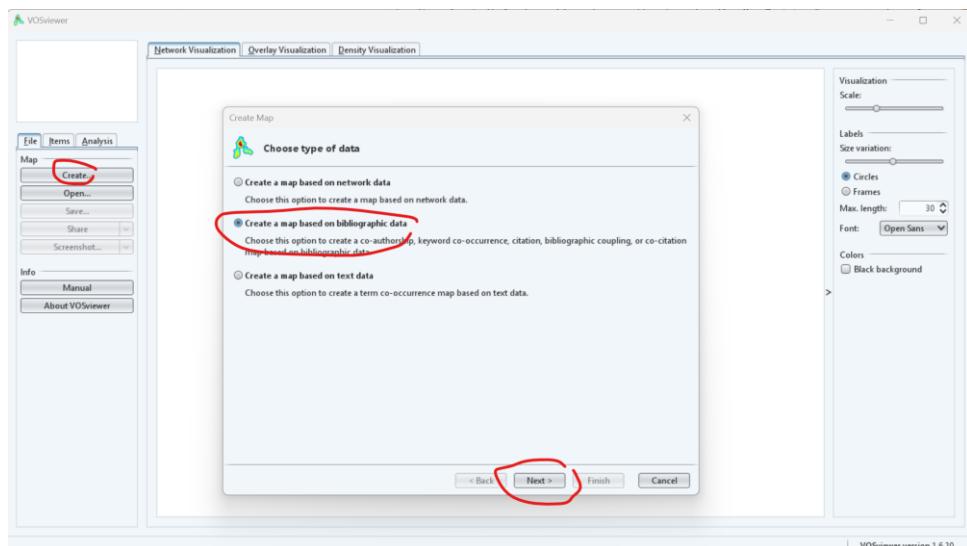
Software UI to start general analysis

Microsoft Excel

- Used for checking data and generating Figure 1.

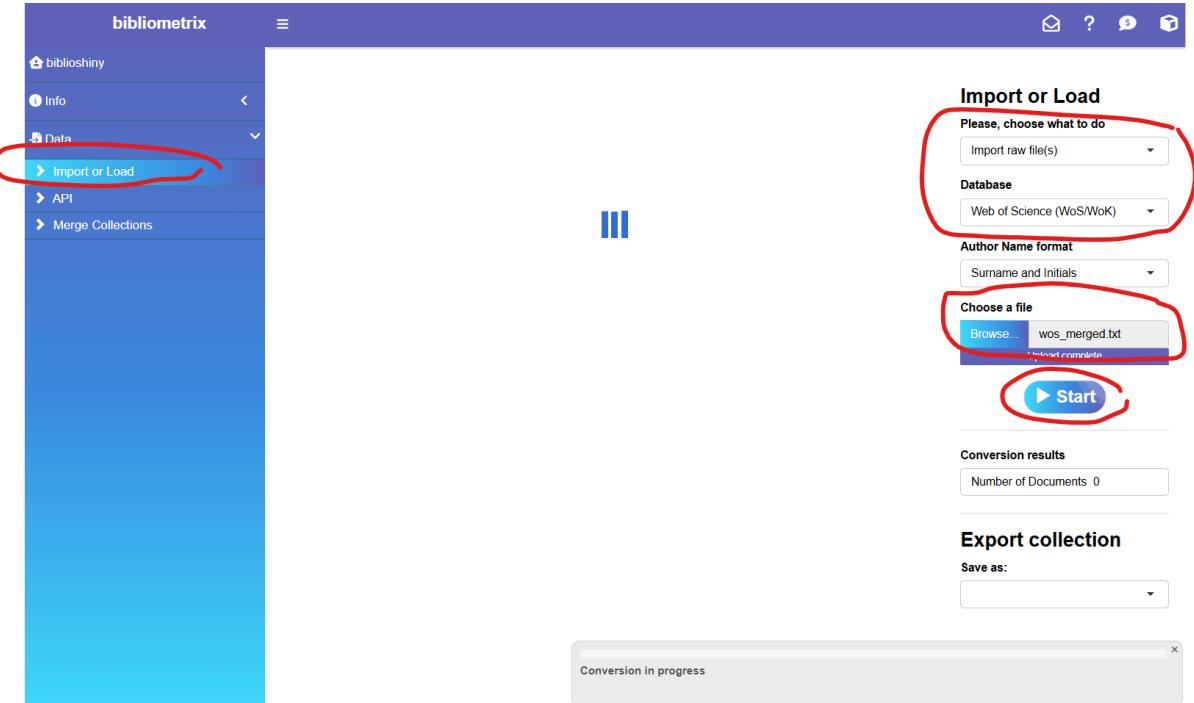
Vosviewer

- Used for generating Figure 5, 6, and 8.
- UI: Load data as shown in the figure below, click “Create”, then choose “Create a map based on bibliographic data”, then click “Next”.



Biblioshiny

- Used for generating Figure 1, 2, 3, 4, 7, and 8.
- UI : as shown below, in "Import or Load," set "Import raw file(s)," set "Database" to "Web of Science (WoS/WoK)," and "Choose a file" to "wos_merged.txt."



Jupyter notebook

- Necessary for generating Figure 2.
- Run figure2.ipynb.

The screenshot shows the Jupyter Notebook interface. At the top, there is a header with the Jupyter logo, "jupyter", "Quit", and "Logout". Below the header, there are tabs for "Files", "Running", and "Clusters", with "Files" being active. A message "Select items to perform actions on them." is displayed above the file list. The file list shows a directory structure under "/Figure2":

- .. (empty folder)
- 256 (empty folder)
- figure2.ipynb (file, status: Running, 299 kB, 2 days ago)
- Annual_Production-2008-2024.xlsx (file, status: Running, 20.2 kB, 2 days ago)
- Countries_Production_Over_Time.xlsx (file, status: Running, 42.1 kB, 2 days ago)
- country-ranking-15.csv (file, status: Running, 2.38 kB, 2 days ago)
- country_rank.png (file, status: Running, 1.34 MB, 2 days ago)

At the top right of the file list, there are buttons for "Upload", "New", and "Edit".

Data

Raw data:

- .\data and code\Raw data\wos
- .\data and code\Raw data\sotm

Ready to analyze data:

- .\data and code\wos_merged.txt
- .\data and code\sotm_merged.xlsx

Figures

- .\data and code\Raw data\

Figure 1

Step 1: load the data

In Biblioshiny software, in "Import or Load," set "Import raw file(s)," set "Database" to "Web of Science (WoS/WoK)," and "Choose a file" to "wos_merged.txt."

Step 2: run analysis

On the left panel, select "Overview" -> "Annual Scientific Production."

On the top panel, select "Table." Download the table: "Annual_Production-2008-2024.xlsx"

Step 3: generate the figure

Generate the line chart in Excel.

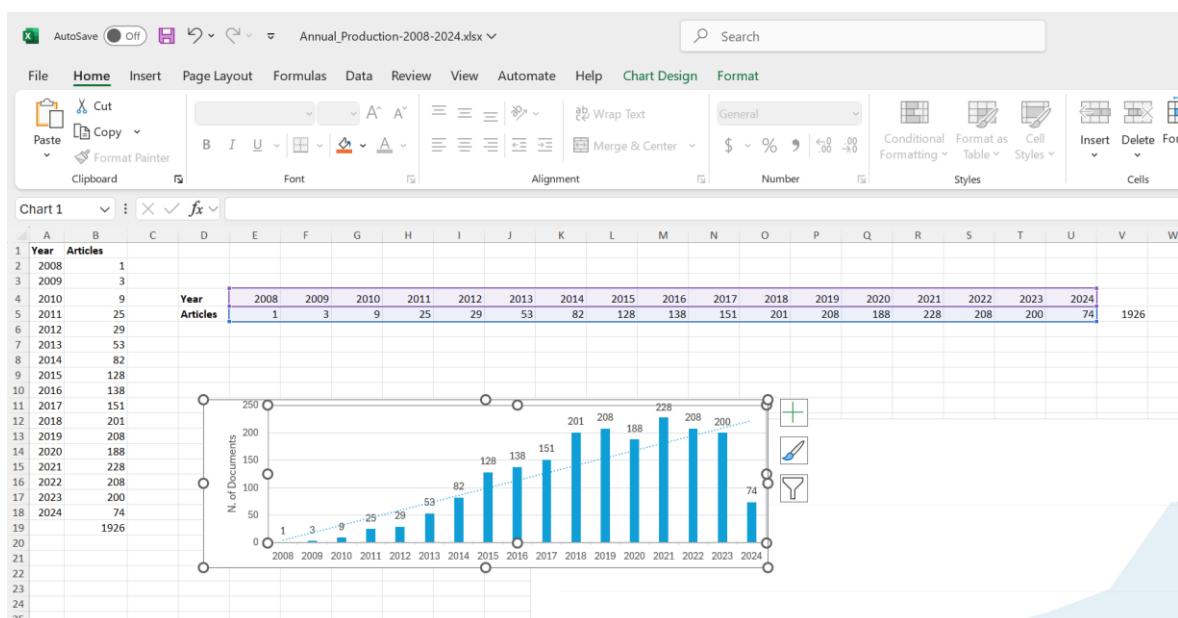


Figure 2

Code directory: .\data and code\Raw data\Figure2

Figure 2 contains two sub-figures, as explained below.

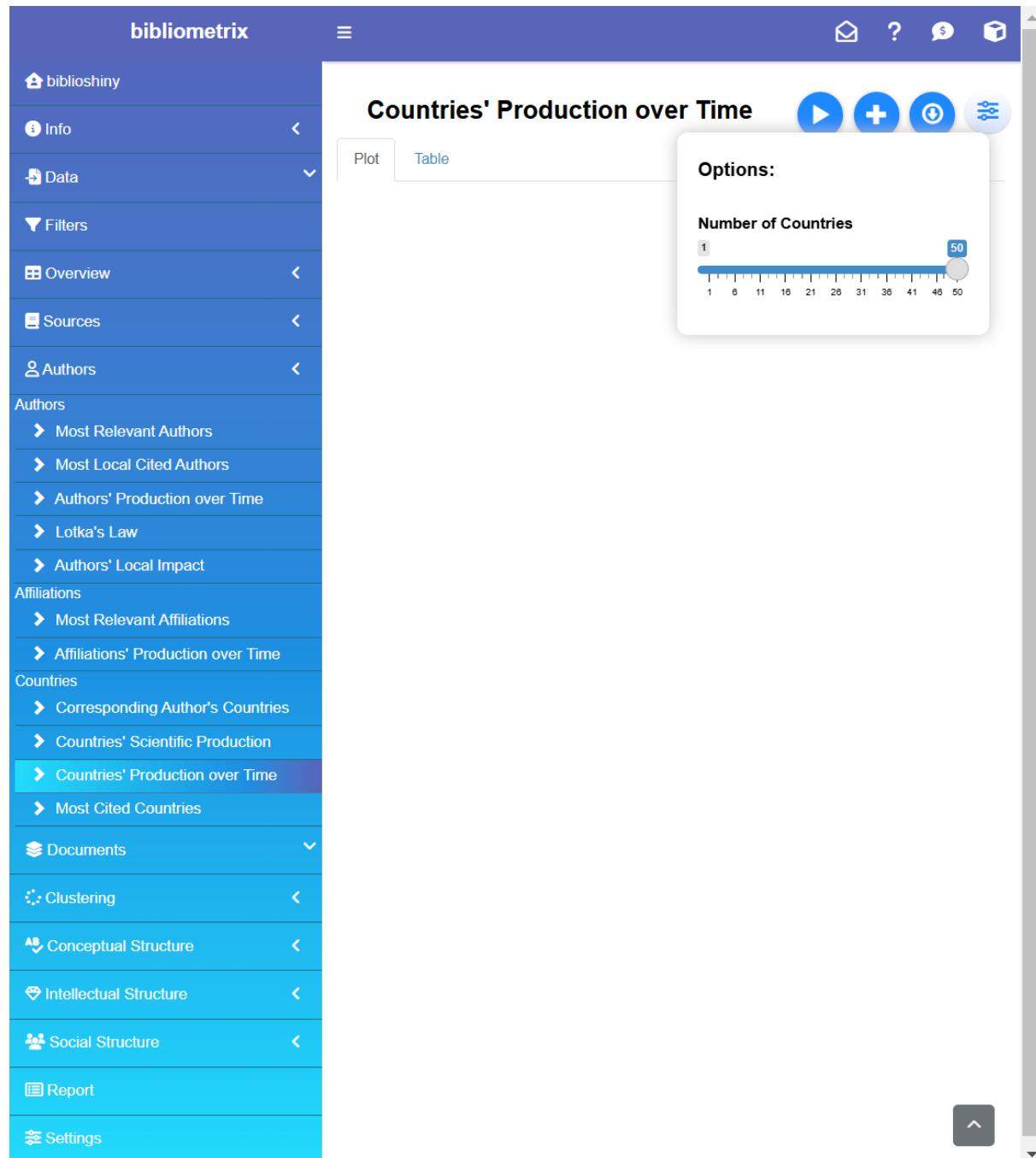
Subfigure 1:

Step 1: load the data.

In Biblioshiny software, in "Import or Load," set "Import raw file(s)," set "Database" to "Web of Science (WoS/WoK)," and "Choose a file" to "wos_merged.txt."

Step 2: run analysis

On the left panel, select "Authors" -> "Countries" -> "Countries' Production over Time," set "Options: Number of Countries" to 50, and click "Run the Analysis."



Step 3: Download the data

On the top panel, select "Table" and download: "Countries_Production_Over_Time.xlsx."

Step 4: rank the countries

In Excel, rearrange the table. For each country per year, compute its production over the total number of papers in that year. Rank them, select the top 15 countries, and make the table "country-ranking-15.csv".

Step 5: Generate the figure. In jupyter notebook, run "figure2.ipynb"

jupyter figure2 Last Checkpoint: Last Wednesday at 1:21 PM (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

In [1]:

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline

from PIL import Image, ImageDraw
from matplotlib.lines import Line2D
from matplotlib.offsetbox import OffsetImage, AnnotationBbox
```

In [2]:

```
def set_seaborn_style(font_family, background_color, grid_color, text_color):
    sns.set_style({
        "axes.facecolor": background_color,
        "figure.facecolor": background_color,
        "axes.grid": True,
        "axes.axisbelow": True,
        "grid.color": grid_color,
        "text.color": text_color,
        "font.family": font_family,
        "xtick.bottom": False,
        "xtick.top": False,
        "ytick.left": False,
        "ytick.right": False,
        "axes.spines.left": False,
        "axes.spines.bottom": False,
        "axes.spines.right": False,
        "axes.spines.top": False,
    })
}
```

In [3]:

```
font_family = "PT Mono"
background_color = "#FFFFFF"
text_color = "#008050"
grid_color = "#FFFFFF"

set_seaborn_style(font_family, background_color, grid_color, text_color)
```

In [4]:

```
def create_rankings(df, columns):
    rank_columns = ["rank_{}".format(i) for i in range(len(columns))]
    for i, column in enumerate(columns):
        df[rank_columns[i]] = df[column].rank(method='min', ascending=False)
    # print(df[rank_columns[1]])
    df[rank_columns[1]] = np.floor(df[rank_columns[1]])

    return df, rank_columns
```

In [5]:

```
def format_ticks(ax, years, padx=0.25, pady=0.25, y_label_size=20, x_label_size=20):
    ax.set(xlim=(-padx, len(years) - 1 + padx), ylim=(-len(df) - pady, -pady))

    xticks = [i for i in range(len(years))]
    ax.set_xticks(ticks=xticks, labels=years)

    yticks = [-i for i in range(1, len(df) + 1)]
    ylabels = ["{}.".format(i) for i in range(1, len(df) + 1)]
    ax.set_yticks(ticks=yticks, labels=ylabels)
    ax.tick_params("y", labelsize=y_label_size, pad=16)
    ax.tick_params("x", labeltop=True, labelsize=x_label_size, pad=8)
```

In [6]:

```
def add_line(ax, row, columns):
    x = [i for i in range(len(columns))]
    y = [row[rc] for rc in columns]

    ax.add_artist(
        Line2D(x, y, linewidth=1., color="#DADBD9")
    )
```

In [7]:

```
def add_pie_and_flag(ax, x, y, name, ratio, size=374, zoom=0.1):
    flag = Image.open("256/{}.png".format(name.lower()))
    image = Image.new('RGB', (size, size))

    draw = ImageDraw.Draw(image)
    pad = int((size - 256) / 2)
```

```

        im, (x, y), frameon=False,
        xycoords="data",
    )))

```

```

In [8]: def add_text(ax, value, max_value, y):
    trillions = round(value / 1e12, 1)
    ratio_to_max = round(100 * value / max_value, 1)

    text = "{}\n${:,}T ({})%".format(
        row.country_name,
        trillions,
        ratio_to_max
    )

    ax.annotate(
        text, (1.03, y),
        fontsize=10,
        linespacing=1.7,
        va="center",
        xycoords="axes fraction", "data")
)

```

```

In [14]: background_color = "#FFFFFF"
text_color = "#000000"
grid_color = "#EAEAEA"

set_seaborn_style(font_family, background_color, grid_color, text_color)

years = ["2008", "2009", "2010", "2011", "2012", "2013", "2014", "2015", "2016", "2017", "2018", "2019"]

df = pd.read_csv("country-ranking-15.csv", index_col=None)
df, rank_columns = create_rankings(df, years)

fig, ax = plt.subplots(nrows=1, ncols=1, figsize=(25, 0.8*len(df)))
format_ticks(ax, years)

for i, row in df.iterrows():

    for j, rc in enumerate(rank_columns):

        add_pie_and_flag(
            ax, j, -row[rc],
            name=row.country_name,
            ratio=row[years[j]])
)

```

```

plt.show()
# plt.savefig('country_rank.png', transparent=True, dpi=300)

```

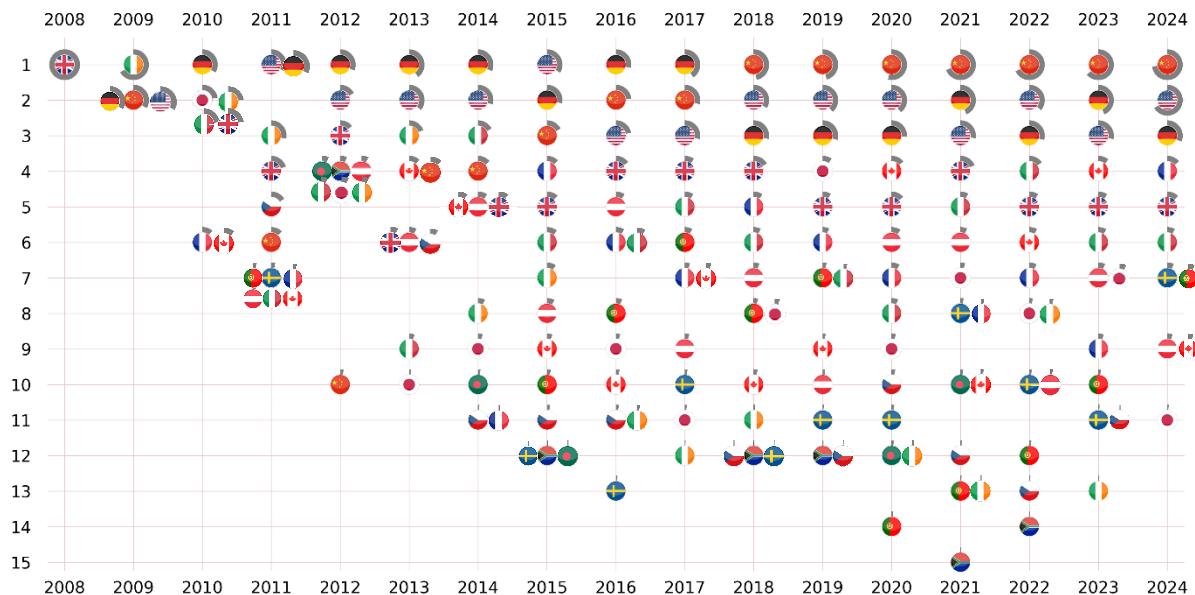
```
In [256]: df[["country_name"] + rank_columns]
```

	country_name	rank_0	rank_1	rank_2	rank_3	rank_4	rank_5	rank_6	rank_7	rank_8	rank_9	rank_10	rank_11	rank_12
0	United Kingdom	1.0	NaN	2.0	4.0	3.0	6.0	5.0	5.0	4.0	4.0	4.0	5.0	
1	Ireland	NaN	1.0	2.0	3.0	4.0	3.0	8.0	7.0	11.0	12.0	11.0	12.0	1
2	Germany	NaN	2.0	1.0	1.0	1.0	1.0	1.0	2.0	1.0	1.0	3.0	3.0	
3	USA	NaN	2.0	NaN	1.0	2.0	2.0	2.0	1.0	3.0	3.0	2.0	2.0	
4	China	NaN	2.0	NaN	6.0	10.0	4.0	4.0	3.0	2.0	2.0	1.0	1.0	
5	Italy	NaN	NaN	2.0	7.0	4.0	9.0	3.0	6.0	6.0	5.0	6.0	7.0	
6	Japan	NaN	NaN	2.0	NaN	4.0	10.0	9.0	15.0	9.0	11.0	8.0	4.0	
7	Canada	NaN	NaN	6.0	7.0	NaN	4.0	5.0	9.0	10.0	7.0	10.0	9.0	
8	France	NaN	NaN	6.0	7.0	NaN	11.0	11.0	4.0	6.0	7.0	5.0	6.0	
9	Czech Republic	NaN	NaN	NaN	5.0	NaN	6.0	11.0	11.0	11.0	13.0	12.0	12.0	1

```
In [ ]:
```

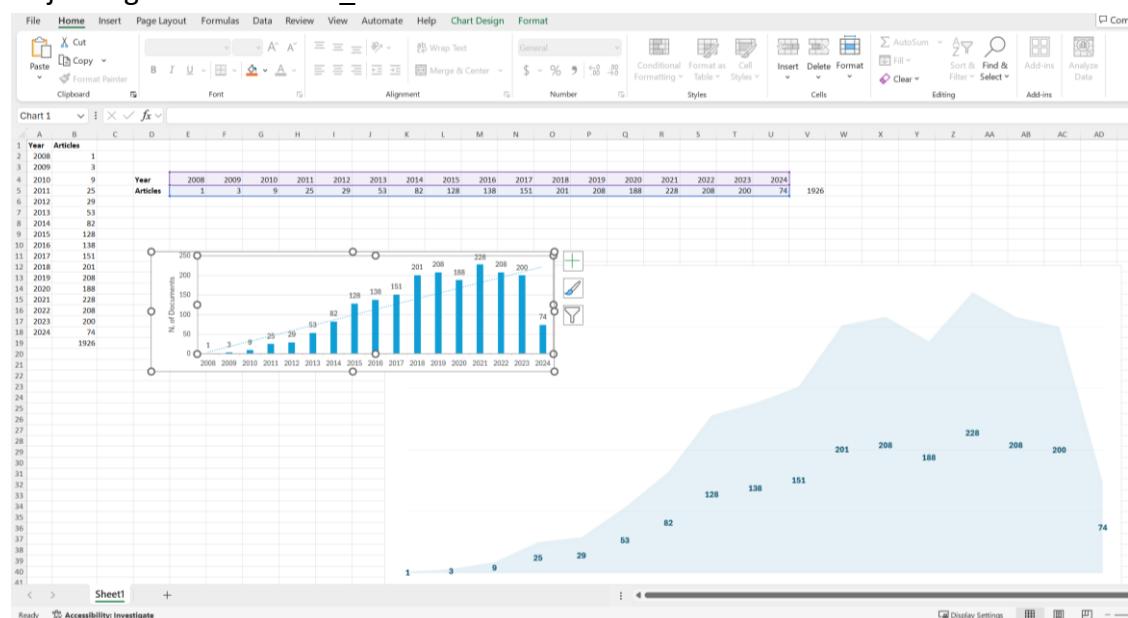
Step 6: Adjust the figure

Countries with the same outputs overlay each other. Adjust the table for all countries and combine the generated figures with the figure below.



Subfigure 2:

Adjust Figure1 in “Annual_Production-2008-2024.xlsx”.



Combine the subfigures:

Finally, add Subfigure 2 to the background of Subfigure 1. Adjust the location of numbers in Subfigure 2. Add labels on the x- and y-axis.

Figure 3

Figure 3 contains three sub-figures, as explained below.

Subfigure 1:

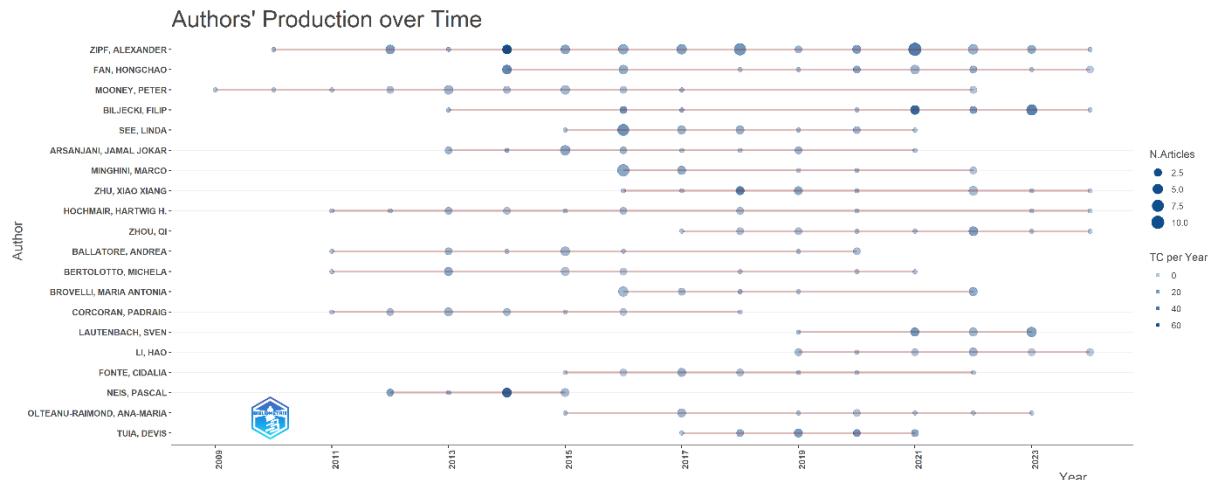
Step 1: load the data.

In Bibioshiny software, in "Import or Load," set "Import raw file(s)," set "Database" to "Web of Science (WoS/WoK)," and set "Author Name format" to "Fullname (if available)," then "Choose a file" to "wos_merged.txt."

Step 2: run analysis

On the left panel, select "Authors" -> "Authors' Productivity over Time," then set "Options: Number of Authors" to 20, and click "Run the Analysis."

The screenshot shows the Bibioshiny software interface. The left sidebar has a dark blue header with the logo 'bibliometrix' and the word 'bibioshiny'. Below the header, there are several sections with expandable arrows: 'Info', 'Data', 'Filters', 'Overview', 'Sources', 'Authors', 'Affiliations', 'Countries', 'Documents', 'Clustering', 'Conceptual Structure', 'Intellectual Structure', 'Social Structure', 'Report', and 'Settings'. The 'Authors' section is expanded, showing sub-options: 'Most Relevant Authors', 'Most Local Cited Authors', 'Authors' Production over Time' (which is highlighted with a blue background), 'Lotka's Law', and 'Authors' Local Impact'. The main content area is titled 'Authors' Production over Time' and contains three tabs: 'Plot' (selected), 'Table - Production per Year', and 'Table - [redacted]'. A large white pop-up window titled 'Options:' is overlaid on the main content. It has a 'Number of Authors' input field containing the value '20'. At the top right of the main content area are several blue circular icons with white symbols: a play button, a plus sign, a magnifying glass, and a gear.



Step 3: Download the tables of top authors.

On the top panel, select "Table - Documents," input the names of the top authors, and download their corresponding tables. Name the table using the format "lastname.xlsx."

Subfigure 2:

Step 1: load the data.

In Biblioshiny software, in "Import or Load," set "Import raw file(s)," set "Database" to "Web of Science (WoS/WoK)," and set "Author Name format" to "Fullname (if available)," then "Choose a file" to "wos_merged.txt."

Step 2: run analysis

On the left panel, select "Authors" -> "Most Relevant Authors," then set "Options: Number of Authors" to 20, and click "Run the Analysis."

bibliometrix

biblioshiny

Info

Data

Filters

Overview

Sources

Authors

Authors

- Most Relevant Authors
- Most Local Cited Authors
- Authors' Production over Time
- Lotka's Law
- Authors' Local Impact

Affiliations

- Most Relevant Affiliations
- Affiliations' Production over Time

Countries

- Corresponding Author's Countries
- Countries' Scientific Production
- Countries' Production over Time
- Most Cited Countries

Documents

Clustering

Conceptual Structure

Intellectual Structure

Social Structure

Report

Settings

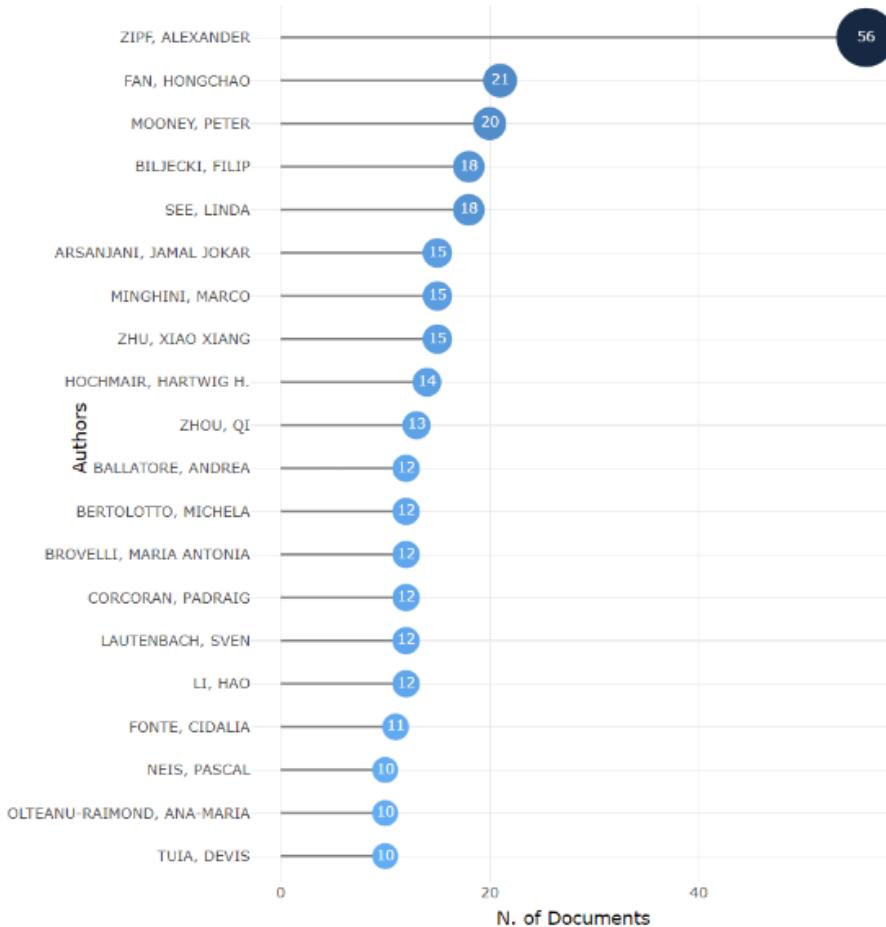
Most Relevant Authors

Plot Table

Options:

Number of Authors: 20

Frequency measure: N. of Documents



Subfigure 3:

Step 1: check affiliations for each author.

For each author, in the table "lastname.xlsx" generated in the previous step, we checked their affiliations corresponding to the publications in each year. Then, add the affiliations to the tables.

Step 2: make a colored table of all affiliations.

Draw a table, as shown below, add all affiliations, and then use different colors to fill the cells in the table.

Ruprecht Karl University of Heidelberg
Wuhan University (WHU)
Norwegian University of Science and Technology (NTNU)
Maynooth University
Environmental Protection Agency, Dublin
Delft University of Technology (TU Delft)
National University of Singapore (NUS)
International Institute for Applied Systems Analysis (IIASA)
Aalborg University Copenhagen (AAU CPH)
Polytechnic University of Milan (PoliMi)
Joint Research Centre (JRC)



Combine the subfigures:

- Overlay a table in Subfigure 1 and color its cells according to Subfigure 3.
- Arrange the three figures in a row and adjust their size. Remove repeated information, e.g., author names in the middle figure.
- Adjust the color of the Subfigure 1 and 2.
- Save as one combined figure, as shown below.

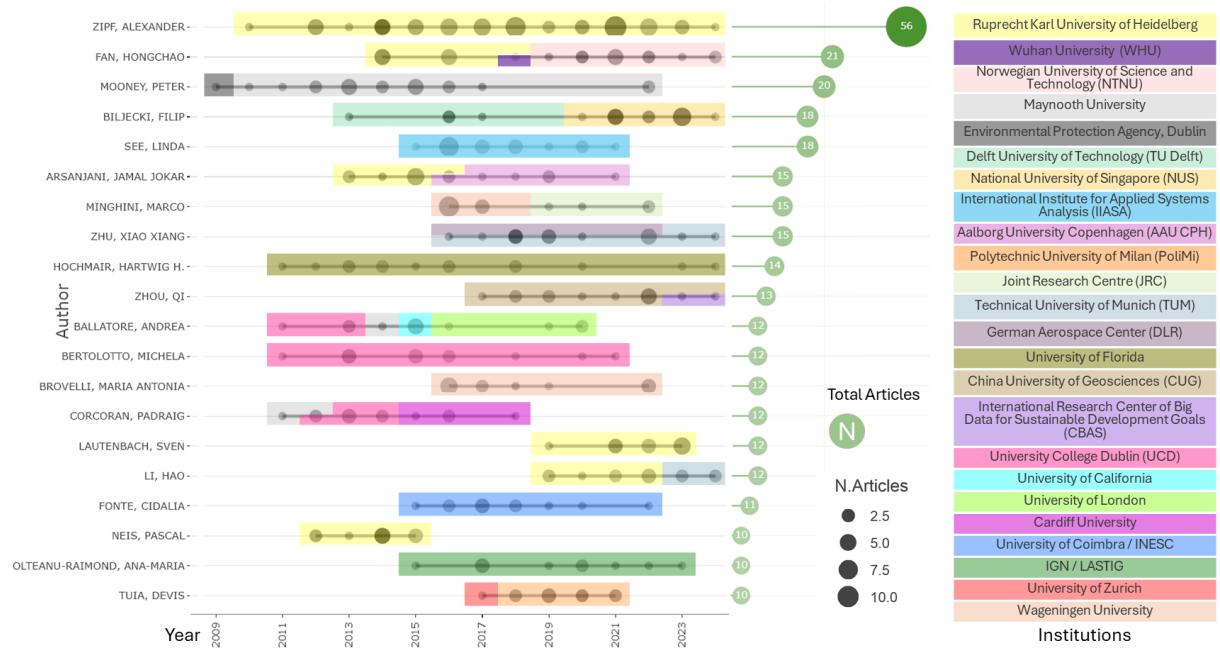


Figure 4

Step 1: load the data.

In Biblioshiny software, in "Import or Load," set "Import raw file(s)," set "Database" to "Web of Science (WoS/WoK)," and "Choose a file" to "wos_merged.txt."

Step 2: run analysis

On the left panel, select "Authors" -> "Countries" -> "Corresponding Author's Countries," set "Options: Number of Countries" to 20, and click "Run the Analysis."

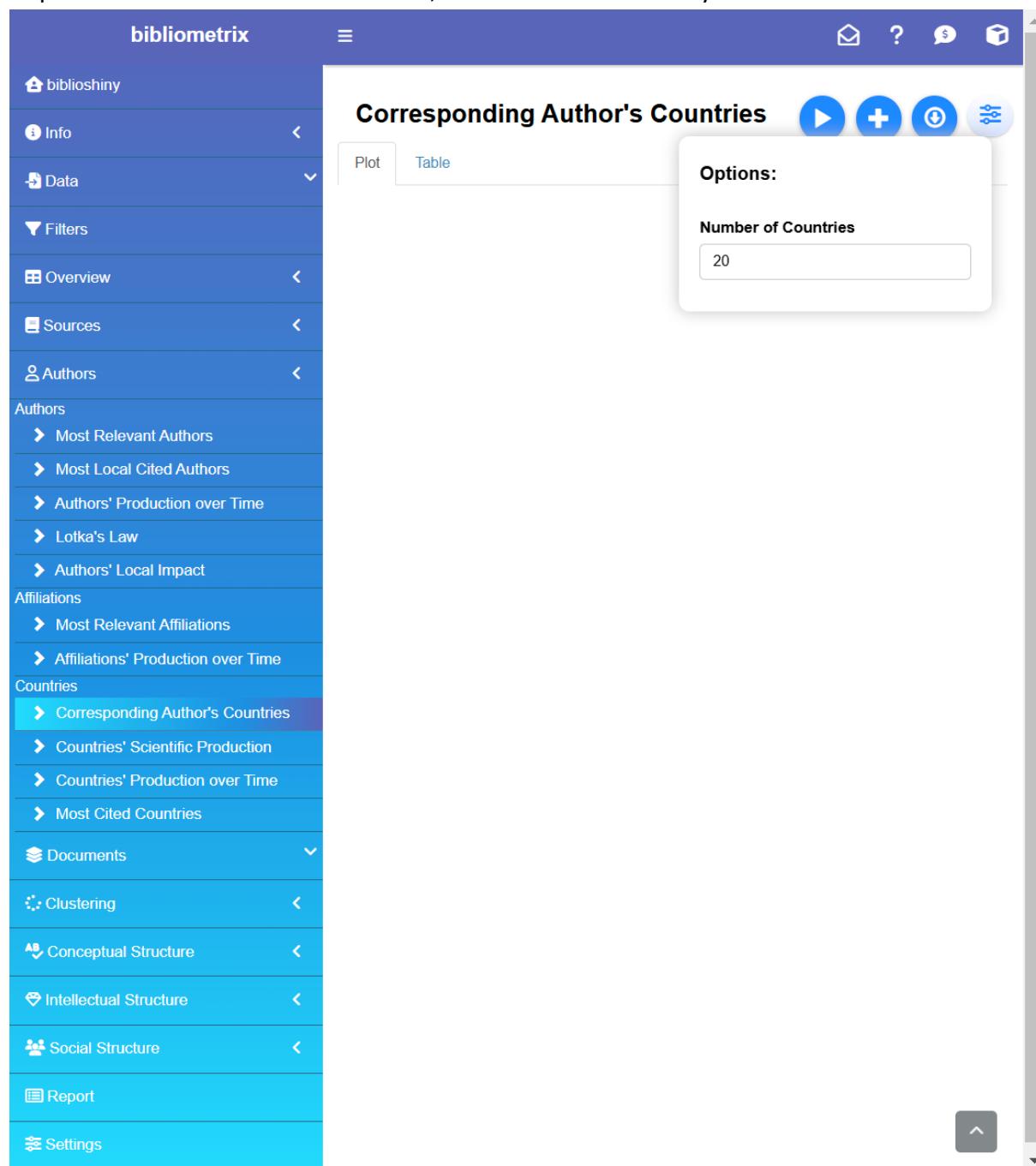


Figure 5

Data directory: .\data and code\Raw data\Figure5

Step 1: load the data.

In Vosviewer software, set "choose the type of data" to "Create a map based on bibliographic data," "Choose data source" to "Read data from bibliographic database files," select "wos_merge_switchC1C3.txt" to load the data.

Step 2: set the methods and parameters:

- "type of analysis" → Co-authorship
- "Unit of analysis" → Organizations
- "counting method" → Full counting
- "choose thresholds" → Minimum number of documents of an organization: 10

Step 3: Create the map (Explained in Section 4.3.2)

Deselect: Arizona State University, Helmholtz Association, Purdue University System, State University System of Florida, Swiss federal institutes of technology domain, University of California System, University of Colorado System, University of Minnesota System, University System of Ohio, and then click "Finish".

- Set parameters according to the screenshot below.

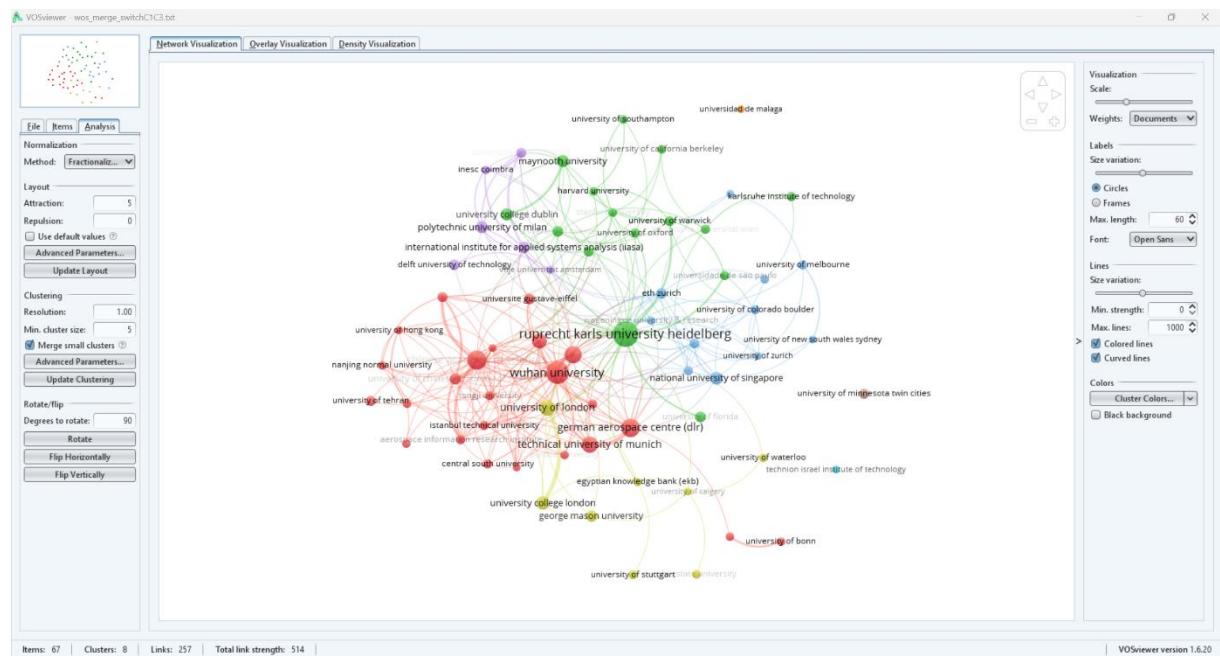


Figure 6

Step 1: load the data.

In Vosviewer software, set "choose type of data" to "Create a map based on bibliographic data", "Choose data source" to "Read data from bibliographic database files", select "wos_merged.txt" to load the data.

Step 2: set the methods and parameters:

- “Type of analysis” → co-authorship
- “Unit of analysis” → Authors
- “Counting method” → Full counting
- “Choose thresholds” → minimum number of documents of an author: 5

Step 3: create map, click “Finish”.

- Set parameters according to the screenshot below.

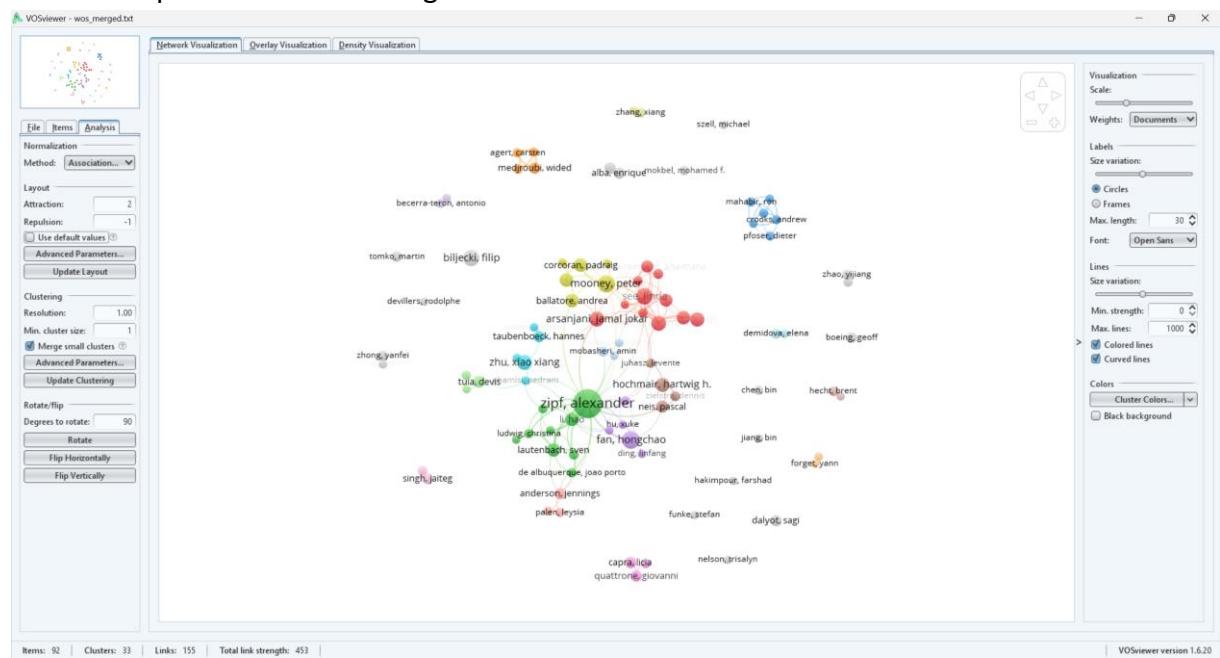


Figure 7

Step 1: load the data.

In Biblioshiny software, in "Import or Load," set "Import raw file(s)," set "Database" to "Web of Science (WoS/WoK)," and "Choose a file" to "wos_merged.txt."

Step 2: run analysis

On the left panel, select "Intellectual Structure" -> "Historiograph," then set "Options: Number of Nodes" to 50, and click "Run the Analysis."

The screenshot shows the Bibliometrix software interface. On the left, a sidebar menu lists various analytical modules: Info, Data, Filters, Overview, Sources, Authors, Documents, Clustering, Conceptual Structure, Intellectual Structure (which is expanded to show Co-citation Network and Historiograph), Social Structure, Report, and Settings. The 'Historiograph' option under 'Intellectual Structure' is currently selected, indicated by a blue background. The main workspace is titled 'Historiograph' and contains two tabs: 'Network' (selected) and 'Table'. To the right of the workspace is a configuration panel titled 'Options:' with the following settings:

- Number of Nodes:** Set to 50.
- Graphical Parameters:** A collapsed section.
- Node label:** Set to "Short id (1st Author, Year)".
- Remove Isolated Nodes:** Set to "Yes".
- Label size:** Set to 2.
- Node size:** Set to 2.

At the top of the main workspace, there are several blue circular icons with white symbols: a play button, a plus sign, a circular arrow, and a gear.

Step 3: Label the key papers

The five key research interests' starting paper(s) are manually highlighted in the figure.

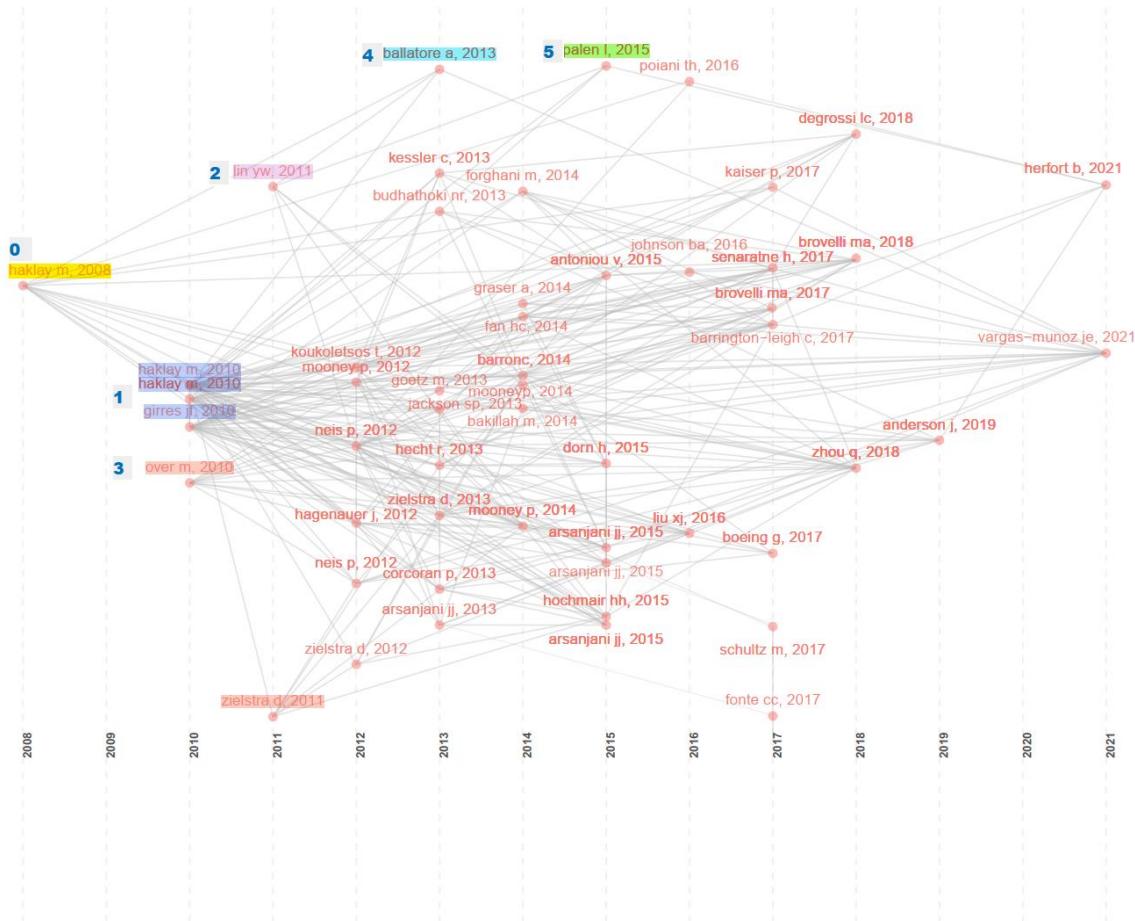


Figure 8

Step 1: load the data.

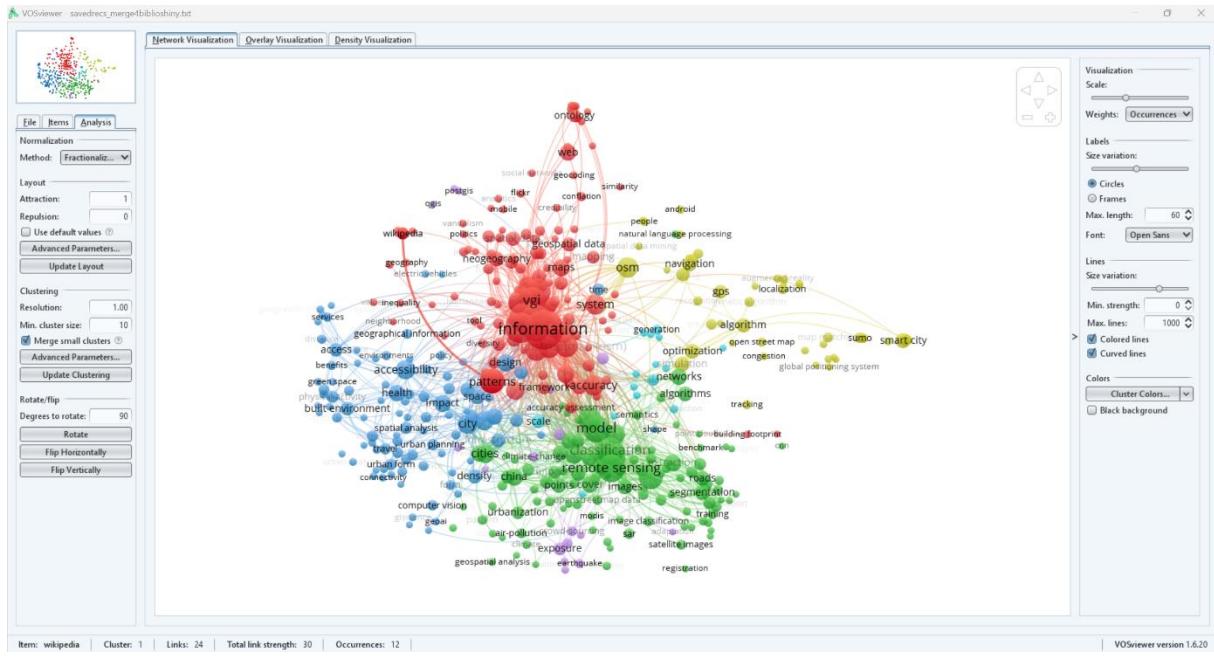
In Vosviewer software, set "choose type of data" to "Create a map based on bibliographic data", "Choose data source" to "Read data from bibliographic database files", select "wos_merged.txt" to load the data.

Step 2: set the methods and parameters:

- “Type of analysis” → Co-occurrence
- “Unit of analysis” → All Keywords
- “Counting method” → Full counting
- “Choose thresholds” → Minimum number of occurrences of a keyword: 5

Step 3: create map, deselect the top 2 keywords, “openstreetmap” and “volunteered geographic information”, then click “Finish”.

- Set parameters according to the screenshot below.



- For Figure 8 (b), change to “Overlay Visualization”.

Figure 9

Figure 9 contains two WoS sub-figures and two SotM subfigures.

WoS figures:

Data directory: .\data and code\Raw data\Figure9\wos

Step 1: load the data.

In Biblioshiny software, in "Import or Load," set "Import raw file(s)," set "Database" to "Web of Science (WoS/WoK)," and "Choose a file" to "wos_merged.txt."

Step 2: set parameters

- For Unigrams:
 - On the left panel, select "Documents" -> "Words" -> "Trend Topics," and set "Options: Filed" to "Titles" and "N-Gram" to "Unigrams."
 - In "Text Editing," at "load a list of terms to remove," load "remove_1.txt"; at "load a list of synonyms," load "synonyms_1.txt".
 - In "Parameters," set "Word Minimum Frequency" to 5 and "Number of Words per Year" to 3.
 - For Bigrams
 - On the left panel, select "Documents" -> "Words" -> "Trend Topics," and set "Options: Filed" to "Titles" and "N-Gram" to "Bigrams."
 - In "Text Editing," at "load a list of terms to remove," load "remove_bi.txt"; at "load a list of synonyms," load "synonyms_bi.txt."
 - In "Parameters," set "Word Minimum Frequency" to 3 and "Number of Words per Year" to 3.

Step 3: run analysis, click "Run the Analysis."

See the screenshot below for getting the Unigram of WoS. Follow the steps to get other figures.

The screenshot shows the bibliometrix application interface. On the left, a sidebar menu lists various analysis options: Info, Data, Filters, Overview, Sources, Authors, Documents, Documents (with sub-options: Most Global Cited Documents, Most Local Cited Documents), Cited References (with sub-options: Most Local Cited References, References Spectroscopy), Words (with sub-options: Most Frequent Words, WordCloud, TreeMap, Words' Frequency over Time, Trend Topics), Clustering, Conceptual Structure, Intellectual Structure, Social Structure, Report, and Settings. The 'Trend Topics' option is selected and highlighted in blue. The main workspace is titled 'Trend Topics' and contains two tabs: 'Plot' (selected) and 'Table'. To the right of the workspace is a large 'Options:' panel. The 'Field' section is set to 'Titles'. The 'N-Grams' section is set to 'Unigrams'. The 'Word Stemming' section is set to 'No'. The 'Timespan' section shows a timeline from 2008 to 2024, with the year 2024 highlighted. Below the timeline, there are sections for 'Text Editing' (with a dropdown set to 'Yes') and 'File Separator' (set to 'Comma ","'). There are also sections for 'Upload complete' (with a file named 'remove_1.txt' listed) and 'Upload synonyms' (with a file named 'synonyms_1.txt' listed). The 'Parameters' section at the bottom includes fields for 'Word Minimum Frequency' (set to 5) and 'Number of Words per Year' (set to 3).

Trend Topics

Options:

- Field: Titles
- N-Grams: Unigrams
- Word Stemming: No
- Timespan: 2008 to 2024

Text Editing

Load a list of terms to remove: Yes

Upload a TXT or CSV file containing a list of terms you want to remove from the analysis.

Terms have to be separated by a standard separator (comma, semicolon or tabulator).

Browse... remove_1.txt Upload complete

File Separator

Comma ","

Load a list of synonyms: Yes

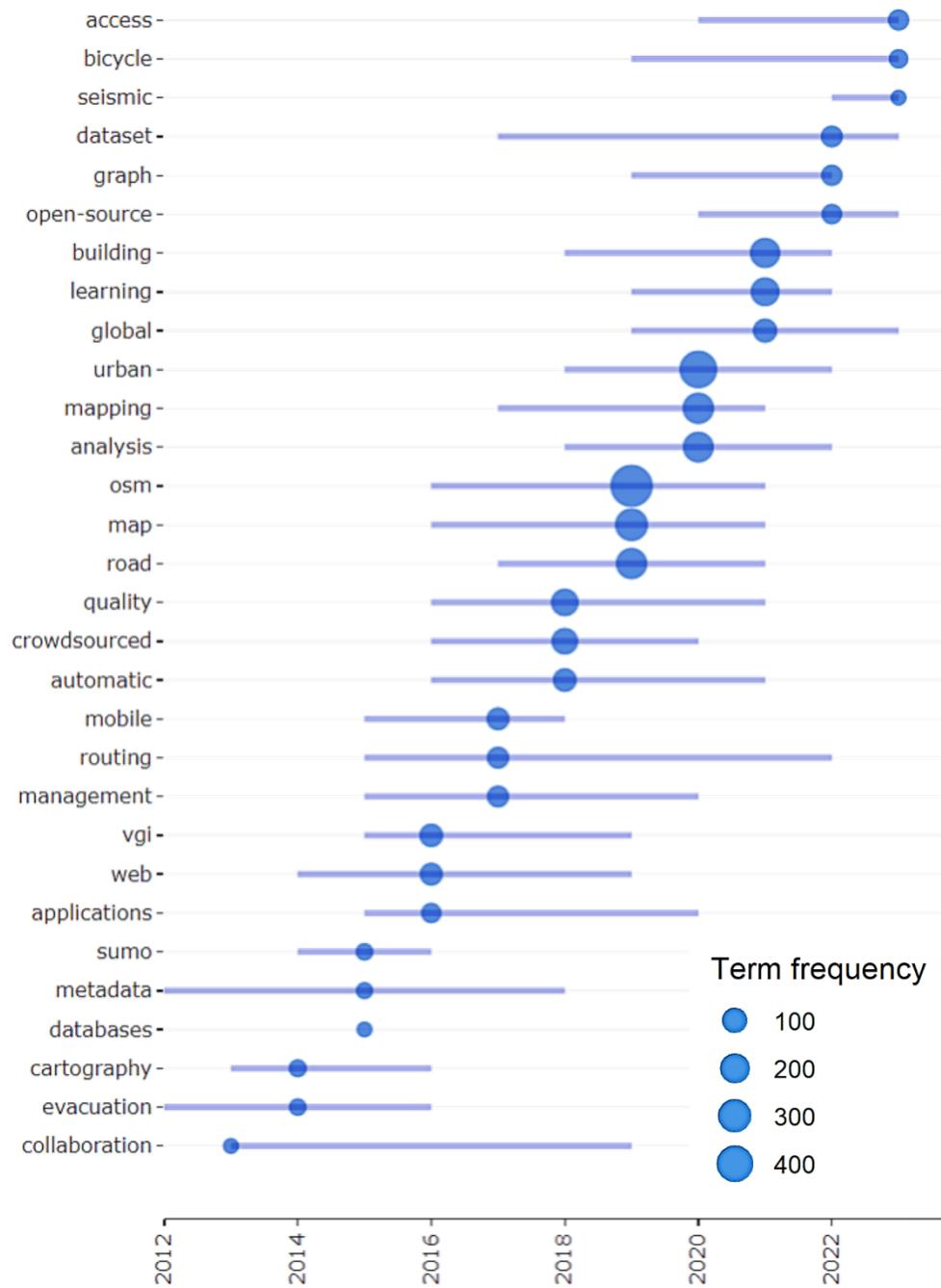
Upload a TXT or CSV file containing, in each row, a list of synonyms, that will be merged into a single term (the first word contained in the row).

Terms have to be separated by a standard separator (comma, semicolon or tabulator). Rows have to be separated by return separator.

Browse... synonyms_1.txt Upload complete

Parameters

Word Minimum Frequency	Number of Words per Year
5	3



SotM figures:

Data directory: .\data and code\Raw data\Figure9\sotm\

Step 0: prepare data for analysis:

- Open "sotm_merged.xlsx," and copy-paste the columns "TI," "AU," and "PY" to the corresponding columns in the template "temp_Bibliometrix.xlsx."
- Name it "SOTM_Bibliometrix.xlsx".

Step 1: load the data.

In Biblioshiny software, in “Import or Load”, set “Load bibliometrix file(s)”, “Choose a file” to “SOTM_Bibliometrix.xlsx”.

Step 2: set parameters

- For Unigrams:
 - On the left panel, select "Documents" -> "Words" -> "Trend Topics," and set "Options: Filed" to "Titles," and "N-Gram" to "Unigrams."
 - In "Text Editing," at "load a list of terms to remove," load "remove_1.txt"; at "load a list of synonyms," load "synonyms_1.txt".
 - In "Parameters", set "Word Minimum Frequency" to 5 and "Number of Words per Year" to 3.
- For Bigrams
 - On the left panel, select "Documents" -> "Words" -> "Trend Topics," and set "Options: Filed" to "Titles," and "N-Gram" to "Bigrams."
 - In "Text Editing," at "load a list of terms to remove," load "remove_bi.txt"; at "load a list of synonyms," load "synonyms_bi.txt."
 - In "Parameters", set "Word Minimum Frequency" to 2 and "Number of Words per Year" to 3.

Step 3: run analysis, click “Run the Analysis”.

Combine the subfigures:

Finally, combine the four subfigures and adjust the colors.