



ShinyMasters

A discovery tool for patterns of research activity
in Master's programmes

User Guide

Version 1.0

Sergio A. Rojas, PhD.

Universidad Distrital Francisco José de Caldas

Bogotá, Colombia, 2021

ShinyMasters Version 1.0 - User guide

Copyright © 2021 Sergio A. Rojas

This document is distributed under the CC BY-NC-ND license (*Creative Commons Attribution-Noncommercial-NoDerivatives 3.0*). Any other unauthorised form of distribution, copying, duplication, reproduction, or sale (total or partial) of the content of this document, both for personal and commercial use, will constitute an infringement of copyright. This guide is an original work of its author, and therefore it is protected by the laws that regulate copyright and intellectual property. The opinions and points of view expressed in this document are personal to the author and do not compromise the policies, intentions, strategies, or official position of any other organism, company, organisation, service or person mentioned in it.

The author has made every effort to ensure that this guide is free from errors or omissions. However, the author accept no responsibility for offence, damage or loss caused to any person acting or endorsing actions using the material contained in this document.

First Edition, March 2021
Bogotá, Colombia

Overview

ShinyMasters is a software tool designed to for exploring hidden patterns that underlie the research activity of a master’s program. Such activity can be seen as a self-organizing system that evolves complex structures and dynamics mainly as a result of the interaction between its faculty and students. The tool leverages the advantages of bibliometric analyses on the metadata from completed dissertations during a given observation window to help revealing such patterns.

ShinyMasters implements a bibliometric workflow intended to obtain insights about production growth, dominant and emerging topics of research, keyword usage trends and collaboration networks, among other facets of the research landscape developed by the Master’s degree. These insights can be useful as a guidance to better assess the strengths of the programme and to prioritise strategic goals.



ShinyMasters v1.0 is available online at:

<https://srojas.shinyapps.io/shinymasters/>

Contents

Overview	iii
1 A discovery tool	1
1.1 What is ShinyMasters ?	1
1.2 How it works	1
1.3 How to run it	2
1.4 Setup	3
1.5 References	4
2 Production dynamics	5
2.1 Production growth	5
2.2 Production distribution	6
2.3 Citation count	6
2.4 Citation distribution	7
2.5 Author's timelines	7
2.6 Word trends	8
3 Conceptual structures	9
3.1 Frequent words	9
3.2 Word cloud	10
3.3 Topic map	10
3.4 Word dendrogram	11
3.5 Co-occurrence network	11
3.6 Thematic map	12
4 Social and intellectual structures	13
4.1 Collaboration network	14
4.2 Authors coupling network	15
4.3 Co-citation network	15
4.4 Manuscript coupling network	16
4.5 Energy flow diagrams	16
5 Using your own dataset	19
5.1 Loading the dataset	19
5.2 Extending your preprocessing pipeline	20
6 Software license	23

Chapter 1

A discovery tool

1.1 What is **ShinyMasters**?

ShinyMasters is a software tool designed to exploring hidden patterns that underlie the research activity of a Master's program. Such activity can be seen as a self-organizing system that evolves complex structures and dynamics mainly as a result of the interaction between its faculty and students. The tool leverages the advantages of bibliometric analyses on the metadata from completed dissertations during a given observation window to help revealing such patterns.

The tool implements a bibliometric workflow intended to obtain insights about production growth, dominant and emerging topics of research, keyword usage trends and collaboration networks, among other facets of the research landscape developed by the Master's degree. These insights can be useful as a guidance to better assess the strengths of the programme and to prioritise strategic goals.

When applying the workflow to a collection of metadata from dissertations accomplished by students and supervisors of the Master's programme and completed during a given observation window, the analysis may provide a critical appraisal of academic production in terms of its bibliometric performance, as well as of the development of conceptual, intellectual and social structures of its associated research activity. We refer to the overall collection of the insights obtained with such analyses as the *research landscape* of the Master's programme.

1.2 How it works



Figure 1.1: The workflow of the tool

The initial stage focuses on setting up the tool, which includes choosing the dataset of dissertation metadata on which the analyses will be performed. The following stages correspond to the execution of the performance bibliometrics and scientific mapping techniques aiming at the discovery of the dynamics and structural facets of the the research landscape of the Master's programme. These are explained in detail in the following sections. For an in-depth discussion of the workflow and its associated bibliometric techniques we refer the reader to the [original paper](#) (see reference [1]).

1.3 How to run it

ShinyMasters is a dashboard application that is hosted on the Web and is available for public usage free of charge. To use this application simply open your favourite Internet browser and point it to the following URL:

<https://srojas.shinyapps.io/shinymasters/>

The visual design of the dashboard is optimised for desktop or laptop browser navigation. However the application is also completely functional for mobile-device browsers.

The application runs in most popular operating systems (Linux, Windows, MacOS, Android, iOS). Upon connection with the hosting server, the following welcome page will show up:

ShinyMasters Welcome! Production dynamics Conceptual structures Social and intellectual structures Settings

Revealing patterns of research activity in Master's programmes

Welcome to ShinyMasters, a tool for exploring hidden patterns that underlie the research activity of a master's program. Such activity can be seen as a self-organizing system that evolves complex structures and dynamics mainly as a result of the interaction between its faculty and students. ShinyMasters leverages the advantages of bibliometric analyses on the metadata from completed dissertations during a given observation window to help revealing such patterns.

ShinyMasters follows a bibliometric workflow intended to obtain insights about production growth, dominant and emerging topics of research, keyword usage trends and collaboration networks, among other facets of the research landscape developed by the Master's degree ([see paper for further details](#)). These insights can be useful as a guidance to better assess the strengths of the programme and to prioritise strategic goals. The bibliometric techniques included in ShinyMasters are implemented using the [bibliometrix R package](#).

Instructions: Use the selector input to choose the dataset with the collection of dissertations to analyse. Choices include preloaded datasets for two engineering Masters' programmes (in spanish) and an option to upload a custom BIB file ([see paper for further details](#)). Then press 'Setup' and navigate to the analyses tabs in the top of the page to run the bibliometric techniques.

Choose dataset:
MSc. Industrial Engineering, UDFJC (2010-2020)▼

Setup ShinyMasters

Connected to ShinyMasters at: 2021-03-18 10:26:31

Select a dataset and press 'Setup' to start analyses with ShinyMasters

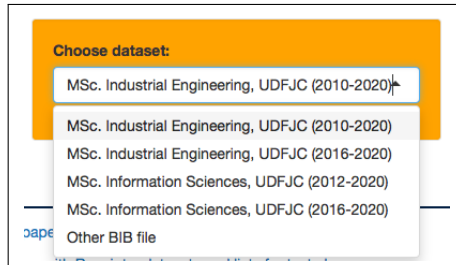
ShinyMasters v1.0
Copyright (c) 2021 Sergio Rojas-Galeano
Universidad Distrital Francisco José de Caldas
srojas@udistrital.edu.co

See pre-print paper on arXiv for details about the tool
See Github repo with R scripts, datasets and list of publications

Now you are ready to use the tool. Follow the instructions or the pop-up notifications in order to perform the analyses.

1.4 Setup

First, use the selector input to choose the dataset with the collection of dissertations to analyse. Choices include preloaded datasets for two engineering Masters' programmes (in Spanish) and an option to upload a custom BIB file:

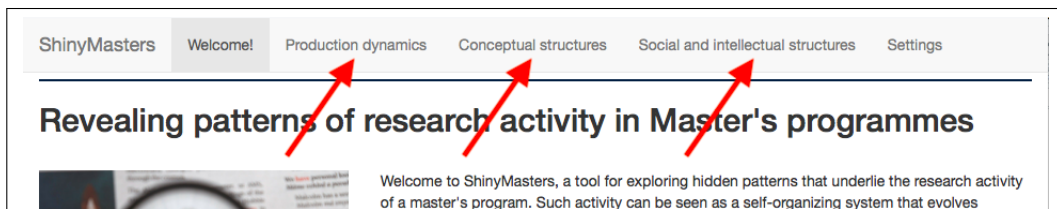


Then press "Setup" to start loading and processing the raw data. The application will indicate when this process is done with a green message and a notification reminder. The green messages also gives you information about the number of documents in the dataset, in the example ($n = 170$).



Notice that the original language of the preloaded datasets is Spanish. Therefore in some specific analyses the results will contain terms in Spanish. Besides, during the collection of dissertations metadata, both supervisors and students names were included as the list of authors.

Now that the setup is finished, you may continue to any of the analyses tabs shown in the menu bar at the top of the page:



The following chapters explain the results that can be obtained in the panels corresponding to each of these analyses tabs.

1.5 References

- [1] Chaparro, N., and Rojas-Galeano, S. (2021). Revealing the research landscape of Master's degrees via bibliometric analyses. In: arXiv preprint, *arXiv:2103.09431*. (<https://arxiv.org/pdf/2103.09431.pdf>).

Chapter 2

Production dynamics

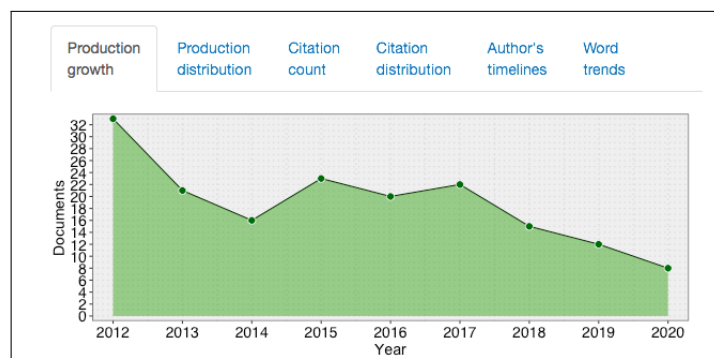
The tool carries out the discovery of production dynamics using performance bibliometrics. These techniques are related to growth, distribution and descriptive statistics of research production, citations, timelines of author's activity as well as trends of word usage. In order to compute these analyses, you need to press the "Run analysis" button first:



Then you can navigate on the tabs menu of this panel, in no particular order. Below we illustrate the results obtained with these analyses, when applied to the dataset called *MSc. Information Sciences UDFJC (2012-2020)* (file name: MIS-dissertations-2012-2020.bib).

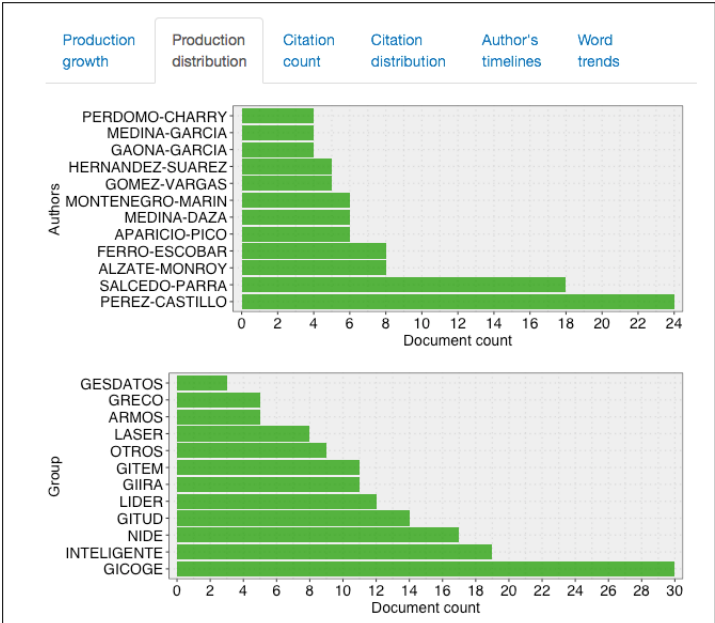
2.1 Production growth

This panel shows a plot with a curve that illustrates the evolution of production counts arranged by year:



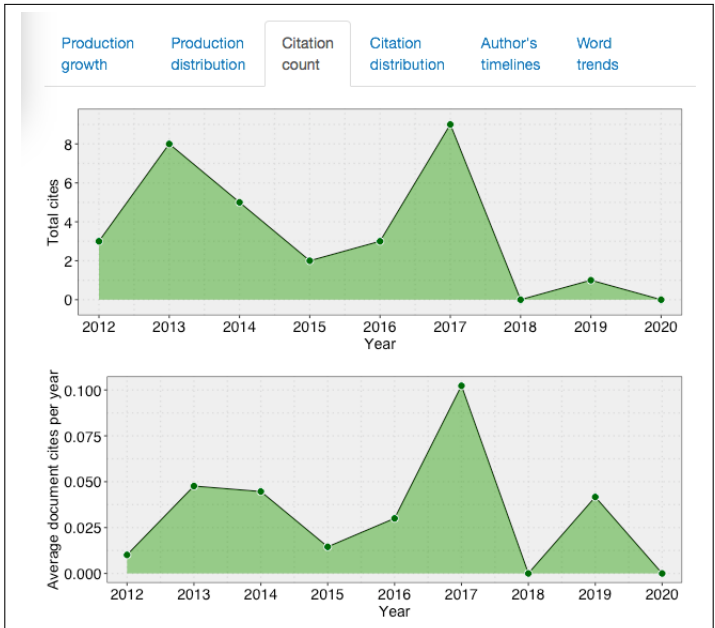
2.2 Production distribution

This panel shows frequency histograms of total number of dissertations per authors or groups:



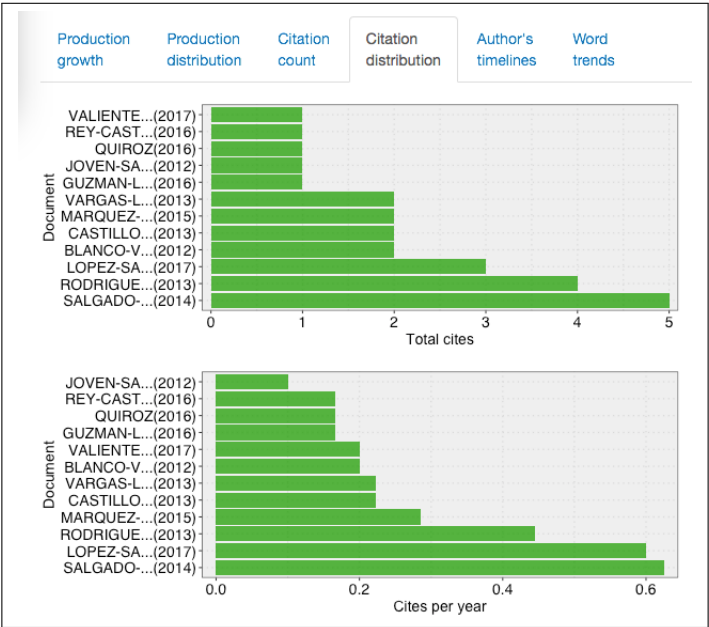
2.3 Citation count

This panel shows a plot with curves representing citation counts (total or averaged) arranged by year:



2.4 Citation distribution

This panel shows frequency histograms of citations for dissertations (total or yearly):



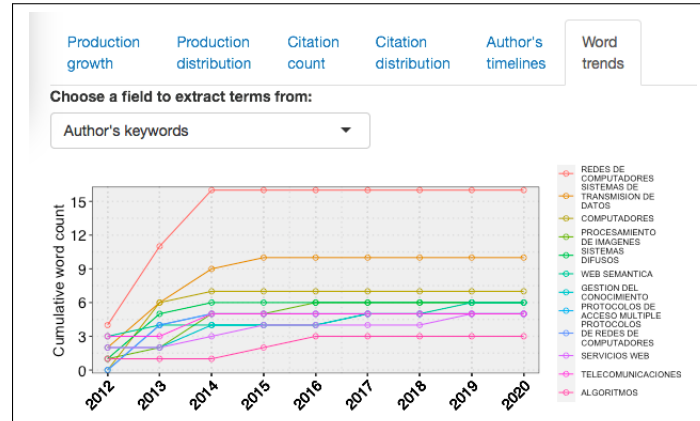
2.5 Author's timelines

This panel shows a stack of 1D bubble diagrams representing dynamics and frequency of author's production (or also groups) over individual timelines:

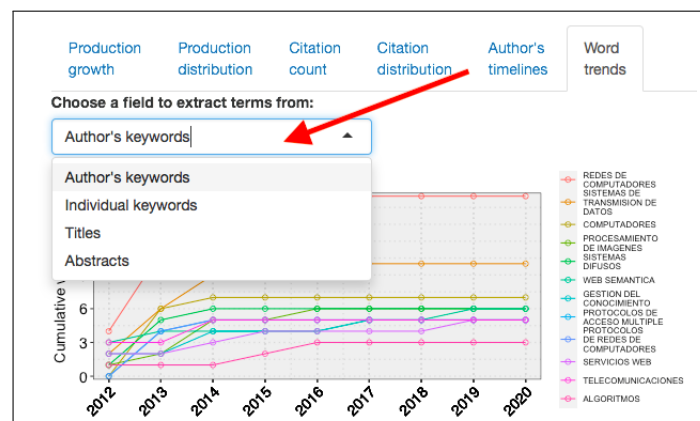


2.6 Word trends

This panel shows a plot of word usage trends over the years, as obtained by title, abstract or keywords:



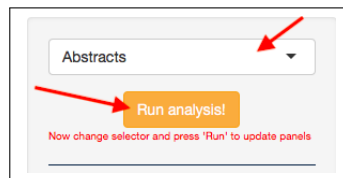
Use the field selector on the left-top corner of the panel to define what terms to extract. The panel would be updated instantly at every selection (no need to press "Run analysis" button again):



Chapter 3

Conceptual structures

The tool carries out the discovery of conceptual structures with science mapping bibliometrics, including techniques such as topic maps, word dendrograms, co-occurrence networks, thematic maps, frequent words and word clouds. In order to compute these analyses, you first need to choose the field to extract the terms from, and then press the "Run analysis" button:

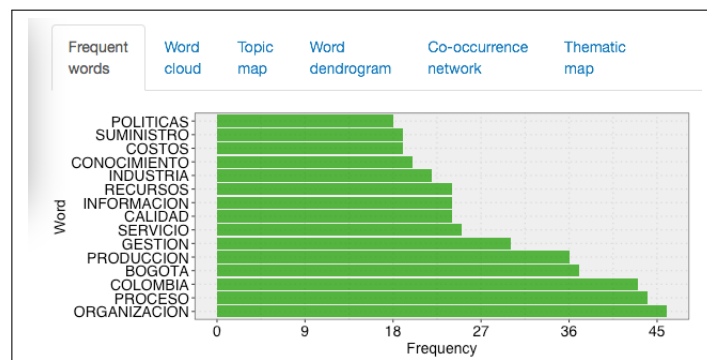


Then you can navigate on the tabs menu of this panel, in no particular order. Below we illustrate the results obtained with these analyses, when applied to the dataset called *MSc. Information Sciences UDFJC (2012-2020)* (file name: MIS-dissertations-2012-2020.bib).

Important: Recall to press the "Run" button every time you select a different field so as to update the panels.

3.1 Frequent words

This panel shows a plot with a frequency histogram of words appearance, obtained by title, abstract or keywords:



3.2 Word cloud

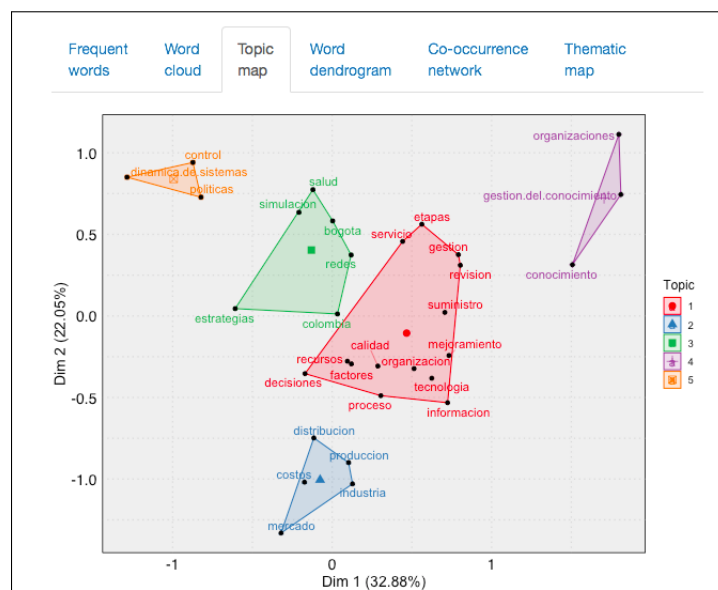
This panel shows a cloud-shaped visual design of most frequent words, obtained by title, abstract, keywords:



Notice that you can hover over any of the words in the cloud with the mouse cursor, to obtain the count of its appearances in the dataset.

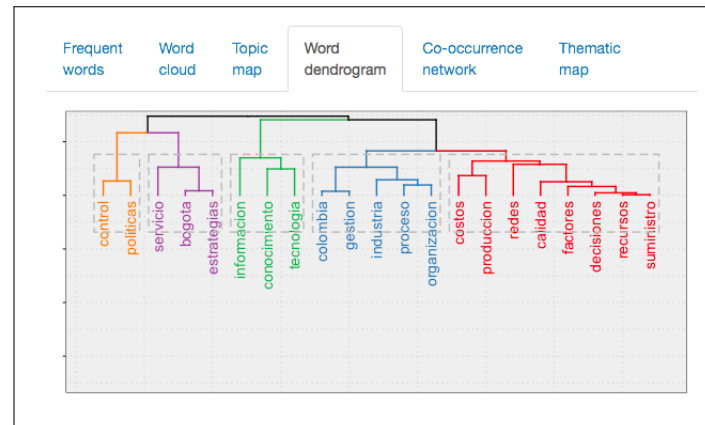
3.3 Topic map

This panel shows a plot where the proximity of words co-occurring in multiple documents are depicted as clusters defining topics or concepts, in a 2D map:



3.4 Word dendrogram

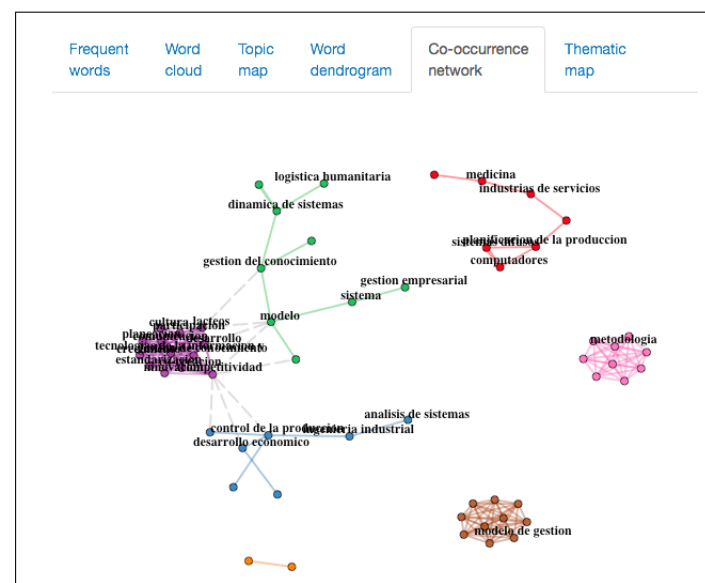
This panel shows an alternative visual format to depict proximity of word co-occurrence, using a hierarchical tree displaying level-dependant partitions:



Notice how the clusters defining topics in the previous topic map, are related to those found in the dendrogram when cutting the branches at the level showed within dashed lines.

3.5 Co-occurrence network

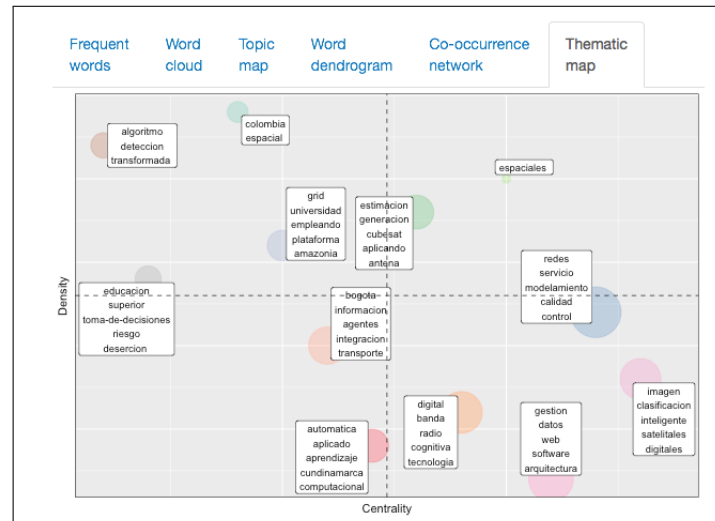
This panel shows a network plot representing different features of words and relationships among them: co-occurrence, dominance and clusters of related terms:



Notice that the layout of the network changes every time you press the "Run Analysis" button. In this way you may obtain different depictions of the network.

3.6 Thematic map

In this panel, by clustering words according to centrality (importance in the field) and density (development in the field) a 2D map is shown depicting motor, emerging, declining and fundamental themes:

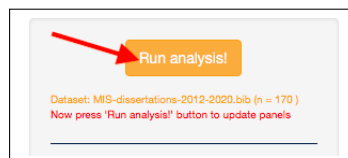


Here the topics of the network are projected onto a 2D map whose dimensions are centrality (relevance of a theme in the research field) and density (maturity on the development of a theme). Therefore, the resulting four quadrants of the map (counterclockwise) would represent motor themes (first quadrant), isolated but highly specialised themes (second quadrant), emerging themes (third quadrant) and fundamental themes (fourth quadrant).

Chapter 4

Social and intellectual structures

The tool carries out the discovery of social and intellectual structures also with science mapping bibliometrics, including techniques such as collaboration and co-citation networks, authors and manuscripts couplings networks and energy flow diagrams (also known as alluvial diagrams). In order to compute these analyses, you need to press the "Run analysis" button first:



Then you can navigate on the tabs menu of this panel, in no particular order. Below we illustrate the results obtained with these analyses, when applied to the dataset called *MSc. Industrial Engineering UDFJC (2010-2020)* (file name: `MIE-dissertations-2010-2020.bib`).

Important: By pressing the "Run" button many times in the same panel, the tool displays a different layout or rendering of the networks. Thus, you are able to choose the one that better suits your visualisation needs.

4.1 Collaboration network

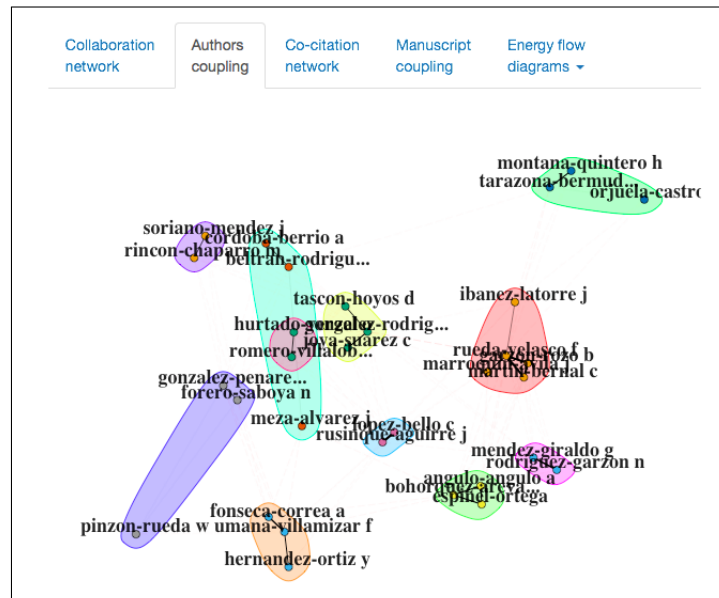
This panel shows a network of co-authorship patterns revealing collaboration links between authors, supervisors and groups:



Given that the list of authors in the dataset comprises the students and their supervisors, three different types of structures emerge from in this network. First, there are star-shaped clusters, in which a single supervisor (central node in the cluster) collaborates with many students to produce several dissertations (the number of dissertations being proportional to the size of the central label). Secondly, there are some triangle-shape clusters, which indicate that two supervisors collaborated with a student in his/her dissertation. And thirdly, there are larger clusters that combine the previous two types, representing extended links of collaboration between several supervisors and students, suggesting the formation of communities.

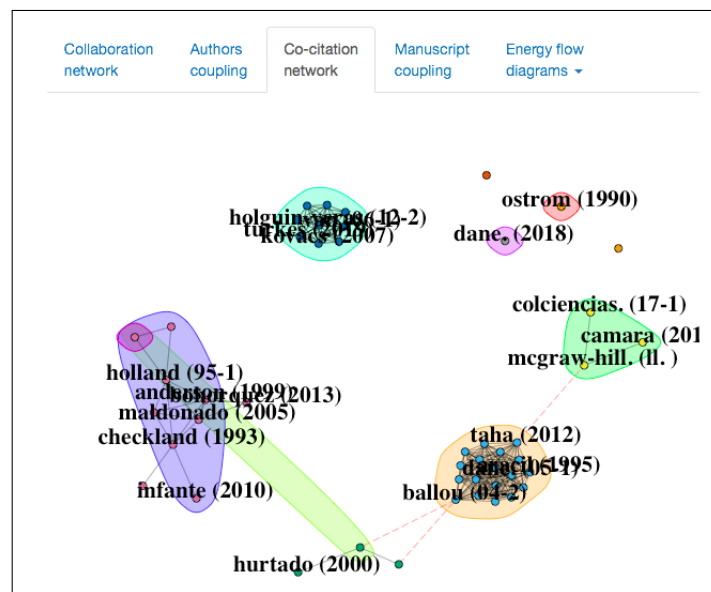
4.2 Authors coupling network

This panel shows a network of authors connected if they share references cited in their entire oeuvres' bibliography. Note that in the case of authors corresponding to supervisors, their oeuvres would be the entire list of dissertations they have supervised:



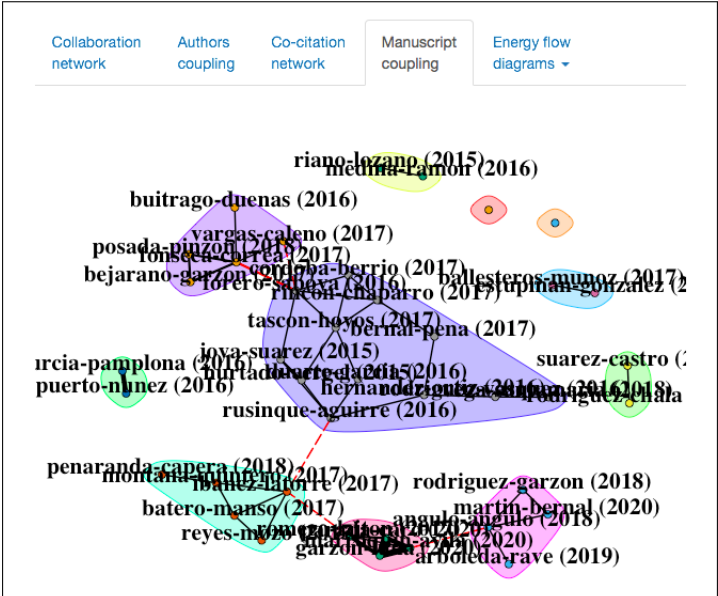
4.3 Co-citation network

This panel shows a network plot of co-occurrence of citations, revealing structures of literature, dominant theories and authorship relevance:



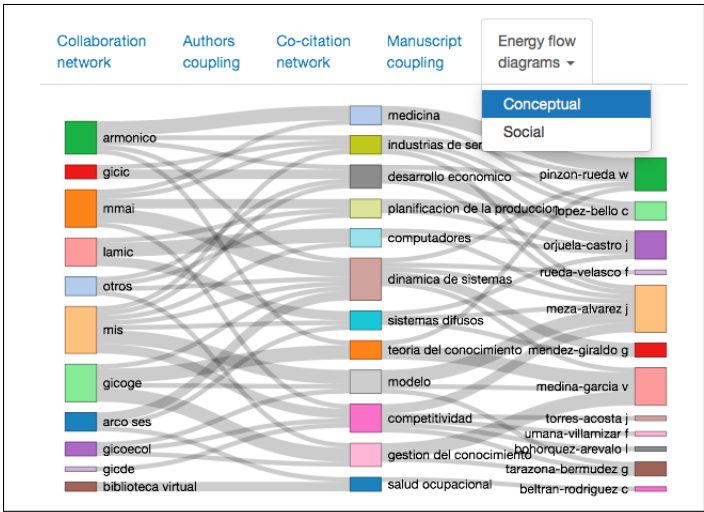
4.4 Manuscript coupling network

This panel shows a network plot of dissertations that are linked when they refer to shared works in their bibliographies:

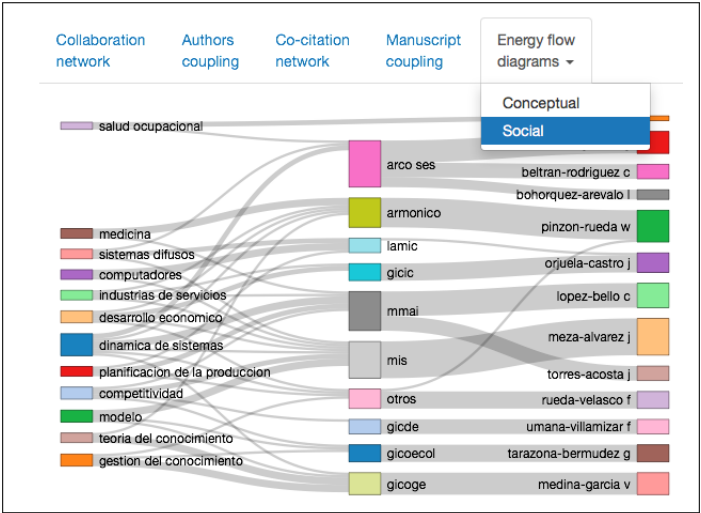


4.5 Energy flow diagrams

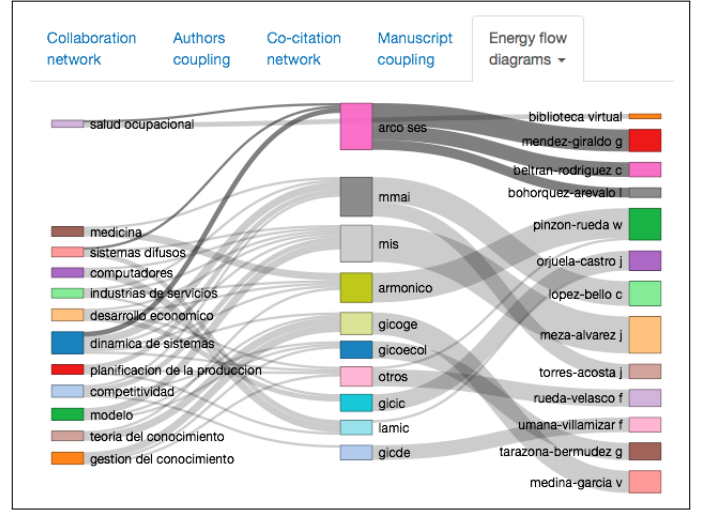
This panel shows a visual representation of energy exchange (i.e. the outflow and inflow of contributions) between bibliographic units. Useful to visualise collaboration structures (social) among groups and authors, or the extent to which they contribute to topic development (conceptual). For example, the conceptual energy flow diagram shows in the middle column the most representative keywords, along with the contribution they receive from groups (left column) and from individual authors (right column):



In contrast, the social energy flow diagram shows in the middle column the most representative groups, along with the contribution they receive from authors (right column) and the topics they feed (left column).



Lastly, notice that this panel allows for the analyst to re-arrange the position of the blocks in each column for particular visualisation needs. For example, below we have moved the groups blocks of the previous diagram such that they now appear sorted by size from top to bottom:



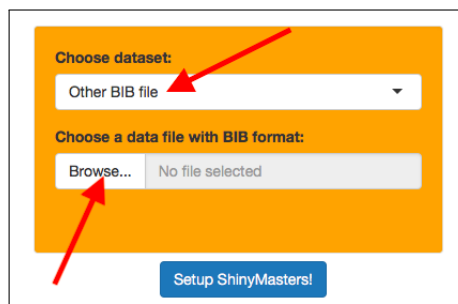
Chapter 5

Using your own dataset

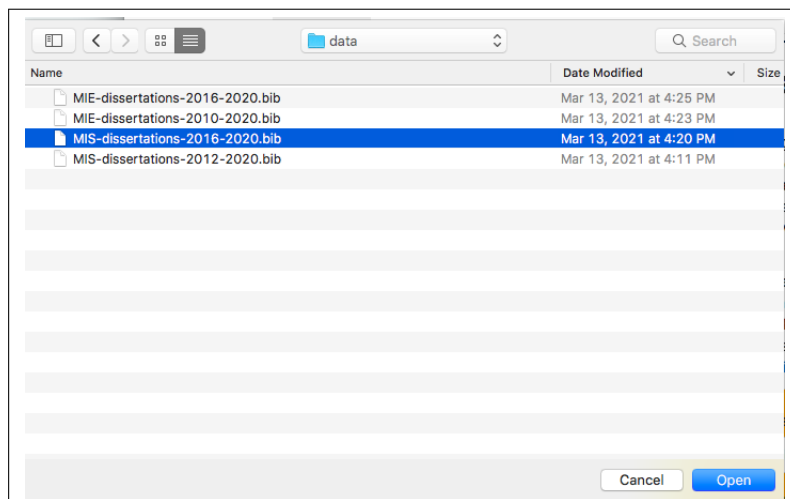
5.1 Loading the dataset

ShinyMasters allows you to load your very own dataset to perform customised analyses. The only requirement is that the dataset has to be formatted as a valid BIB file. A description of the BIB scheme used by the tool can be found in the [original paper](#). The actual BIB files of the preloaded dataset can be seen [in this repo](#).

Once you collected your data in a BIB file, go to the "Welcome!" tab and choose the "Other BIB file" option in the dataset selector:



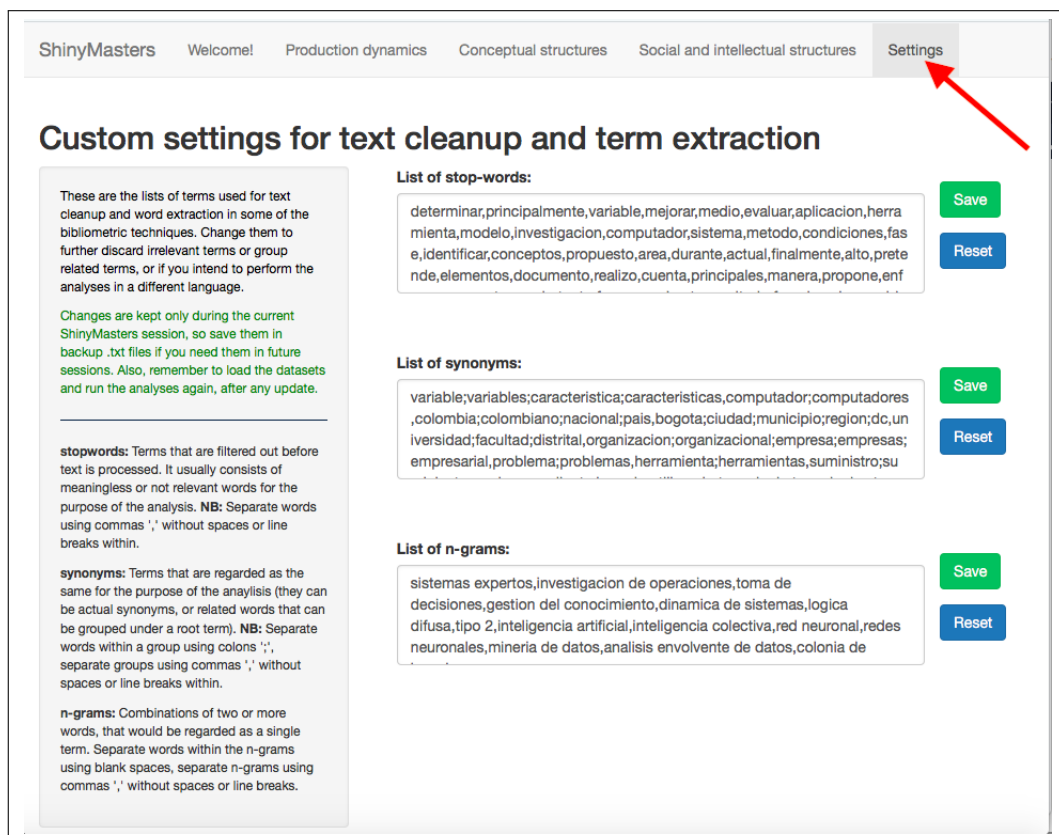
Press the "Browse" and locate your BIB file using the pop-up window (make sure you choose a `.bib` otherwise the application may crash):



Once uploading is finished, press the "Setup" button and you will be ready to go and run the dynamics and structure analyses.

5.2 Extending your preprocessing pipeline

You can customise the preprocessing pipeline that **ShinyMasters** applies to your dataset before computing the results of the bibliometric analyses. The pipeline consists of three steps: removing **stopwords** (terms that are useless for the purposes of the analyses), grouping **synonyms** (terms that are actually synonyms or hyperonyms, or that can be considered equivalent in the context of the analyses), and maintaining **n-grams** specifically relevant for the particular dataset. These lists are found in the "Settings" tab panel:



The screenshot shows the 'Settings' tab in the ShinyMasters application. The title is 'Custom settings for text cleanup and term extraction'. On the left, there is a text box explaining that these are lists of terms used for text cleanup and word extraction, and that changes are kept only during the current session. Below this, there are three sections: 'stopwords', 'synonyms', and 'n-grams', each with a brief explanation of what they are and how to format them. On the right, there are three input fields, each with a 'Save' button and a 'Reset' button. The first field is 'List of stop-words:' with a text area containing a list of terms like 'determinar', 'principalmente', etc. The second field is 'List of synonyms:' with a text area containing terms like 'variable', 'variables', 'caracteristica', etc. The third field is 'List of n-grams:' with a text area containing terms like 'sistemas expertos', 'investigacion de operaciones', etc. A red arrow points to the 'Settings' tab in the top navigation bar.

ShinyMasters Welcome! Production dynamics Conceptual structures Social and intellectual structures **Settings**

Custom settings for text cleanup and term extraction

These are the lists of terms used for text cleanup and word extraction in some of the bibliometric techniques. Change them to further discard irrelevant terms or group related terms, or if you intend to perform the analyses in a different language.

Changes are kept only during the current ShinyMasters session, so save them in backup .txt files if you need them in future sessions. Also, remember to load the datasets and run the analyses again, after any update.

stopwords: Terms that are filtered out before text is processed. It usually consists of meaningless or not relevant words for the purpose of the analysis. **NB:** Separate words using commas ',' without spaces or line breaks within.

synonyms: Terms that are regarded as the same for the purpose of the analysis (they can be actual synonyms, or related words that can be grouped under a root term). **NB:** Separate words within a group using colons ':'; separate groups using commas ',' without spaces or line breaks within.

n-grams: Combinations of two or more words, that would be regarded as a single term. Separate words within the n-grams using blank spaces, separate n-grams using commas ',' without spaces or line breaks.

List of stop-words:

determinar,principalmente,variable,mejorar,medio,evaluar,aplicacion,herramienta,modelo,investigacion,computador,sistema,metodo,condiciones,fase,identificar,conceptos,propuesto,area,durante,actual,finalmente,alto,preliminar,elementos,documento,realizo,cuenta,principales,manera,propone,enf

Save Reset

List of synonyms:

variable;variables;caracteristica;caracteristicas,computador;computadores,colombiano;colombiano;nacional;pais,bogota;ciudad;municipio;region;dc,universidad;facultad;distrital,organizacion;organizacional;empresa;empresas;empresarial,problema;problemas,herramienta;herramientas,suministro;su

Save Reset

List of n-grams:

sistemas expertos,investigacion de operaciones,toma de decisiones,gestion del conocimiento,dinamica de sistemas,logica difusa,tipo 2,inteligencia artificial,inteligencia colectiva,red neuronal,redes neuronales,mineria de datos,analisis envolvente de datos,colonia de

Save Reset

The default lists are shown when you open the "Settings" panel. The lists are editable by typing in new or deleting existing words in the corresponding text area. The format of each list is explained in the left-hand sidebar panel. To update changes in any list, press the corresponding "Save" button. To restore the default values for any of the lists, press the respective "Reset" button.

Below we illustrate an example of how to customise the list of stopwords. For this aim, recall the thematic map we pictured before in section 3.6, when the analysis was being performed for the `MIS-dissertations- 2012-2020.bib` dataset:


It can be seen that the highlighted Spanish term *empleando* (translation: employing, utilising or using) appeared as part of the *grid* theme node. Clearly this is an irrelevant term in this context. Therefore we can add it to the list of stopwords and save the updated list:

A notification warning shows up indicating that we need to load again a dataset in order to begin the analyses:

So we choose the above dataset and press "Setup":

ShinyMasters
Welcome!
Production dynamics
Conceptual structures
Social and intellectual structures
Settings

Revealing patterns of research activity in Master's programmes



Welcome to ShinyMasters, a tool for exploring hidden patterns that underlie the research activity of a master's program. Such activity can be seen as a self-organizing system that evolves complex structures and dynamics mainly as a result of the interaction between its faculty and students. ShinyMasters leverages the advantages of bibliometric analyses on the metadata from completed dissertations during a given observation window to help revealing such patterns.

ShinyMasters follows a bibliometric workflow intended to obtain insights about production growth, dominant and emerging topics of research, keyword usage trends and collaboration networks, among other facets of the research landscape developed by the Master's degree ([see paper for further details](#)). These insights can be useful as a guidance to better assess the strengths of the programme and to prioritise strategic goals. The bibliometric techniques included in ShinyMasters are implemented using the [bibliometrix R package](#).

Instructions: Use the selector input to choose the dataset with the collection of dissertations to analyse. Choices include preloaded datasets for two engineering Masters' programmes (in spanish) and an option to upload a custom BIB file ([see paper for further details](#)). Then press 'Setup' and navigate to the analyses tabs in the top of the page to run the bibliometric techniques.

Choose dataset:

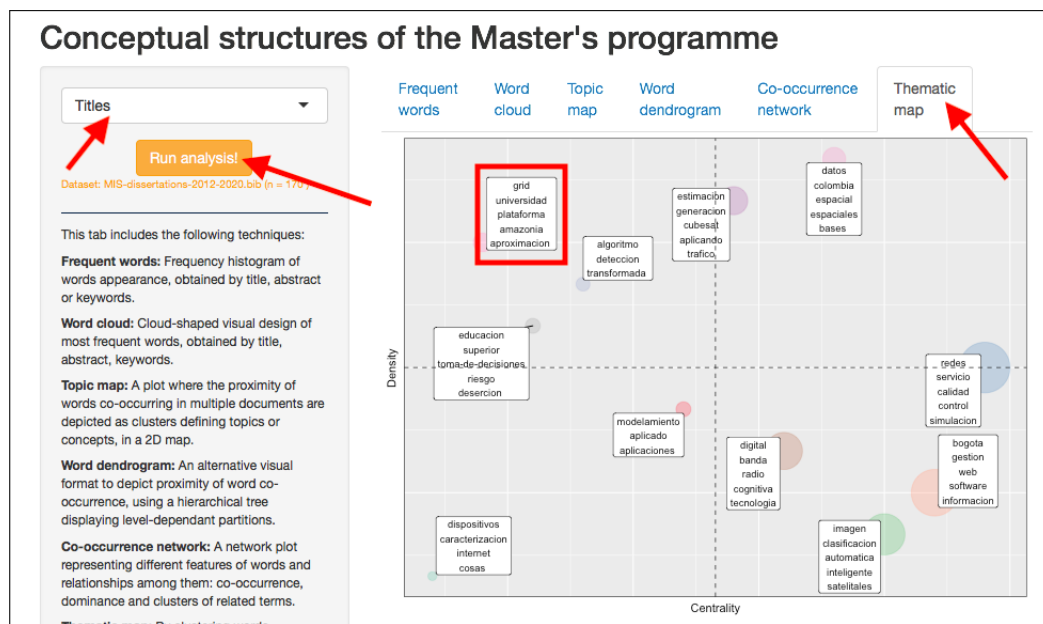
- MSc. Industrial Engineering, UDFJC (2010-2020)~
- MSc. Industrial Engineering, UDFJC (2010-2020)
- MSc. Industrial Engineering, UDFJC (2016-2020)
- MSc. Information Sciences, UDFJC (2012-2020)
- MSc. Information Sciences, UDFJC (2016-2020)
- Other BIB file

[See pre-print paper](#)

[See Github repo with R scripts, datasets and lists for text cleanup](#)

ShinyMasters v1.0
Copyright (c) 2021 Sergio Rojas-Galeano
Universidad Distrital Francisco José de Caldas
srojas@udistrital.edu.co

Once setup is ready, we turn to the "Conceptual structures" tab panel, select the "Titles" option, press the "Run analysis!" button, and click on the "Thematic map" panel. The result now shows that the *grid* theme node does not include the *empleando* any more:



Chapter 6

Software license

ShinyMasters version 1.0
Copyright © 2021 Sergio A. Rojas.

This program is free software: you can redistribute it and/or modify it under the terms of the GNU Affero General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU Affero General Public License along with this program. If not, you can download it from:

<https://www.gnu.org/licenses/agpl-3.0.txt>.