QUESTION 1

1. Downloading Scopus Dataset

Steps to download data (may vary between the databases):

a. Go to the homepage of SCOPUS https://www.scopus.com
b. Type your relevant topic (KEYWORD=Quantum Computing) in Search box
c. Click source type- check to JOURNALS.
d. Document type- ARTICLE
e. Click Subject Area (i.e Computer Science , Information systems etc)
f. After all the steps- Click "LIMIT TO" or refine in web of science
g. find out best articles by click sort based on Number of citations
h. Click Export button after selecting articles based on method adopted, Choose CSV (comma separated value) Excel to export the data and bibtext to export the data.

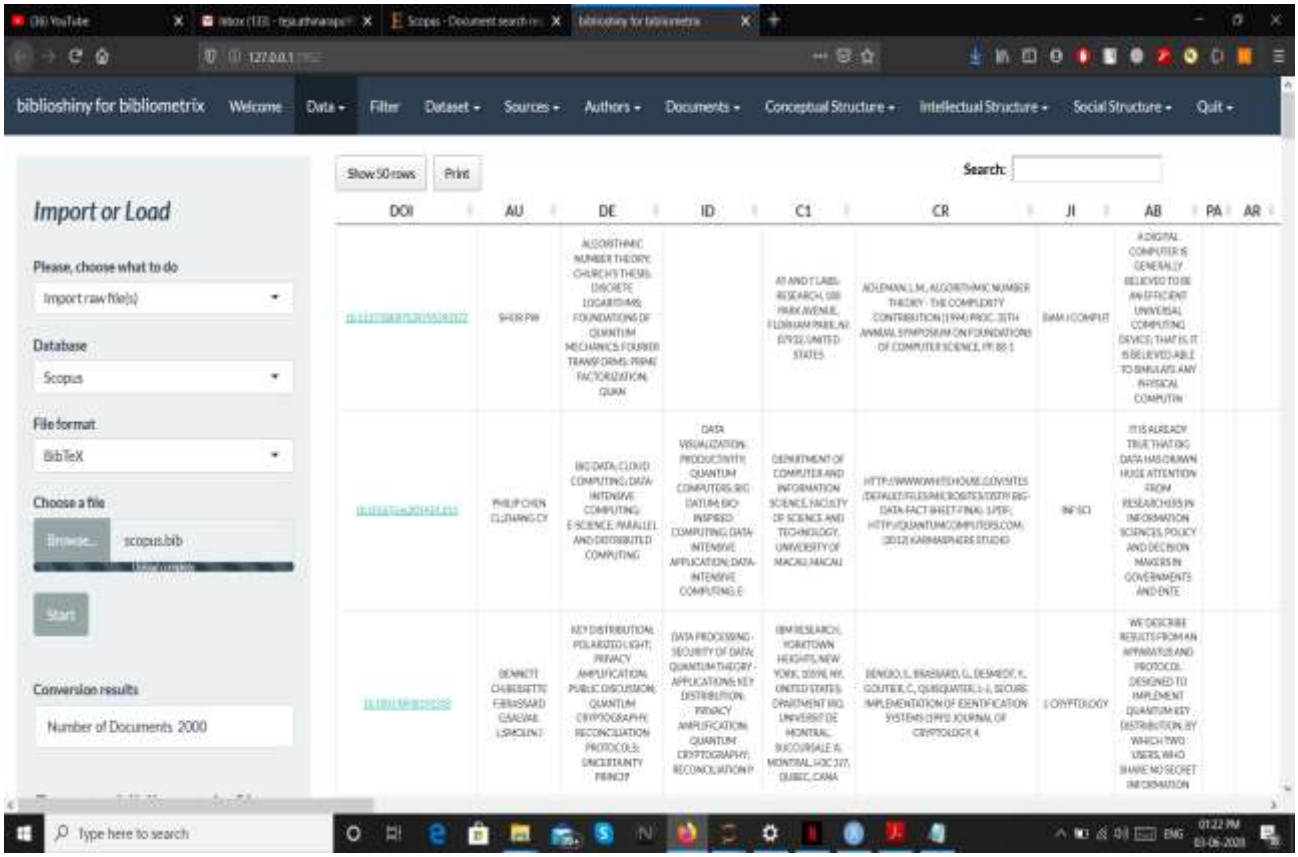
2. Using R to Analyze.

- In R studio, R script is used for executing R commands
- Library sqldf is used for executing sql queries in R.

  The sqldf() function is typically passed a single argument which is an SQL select statement where the table names are ordinary R data frame names. sqldf() transparently sets up a database, imports the data frames into that database, performs the SQL select or other statement and returns the result using a heuristic to determine which class to assign to each column of the returned data frame

- A special R package bibliometric is used for analysing the scopus data directly in a web tool of this package called Biblioshiny

- Using this biblioshiny() ,directly in web browser different type of analysis can be done on scopus/csv/webofdocuments data.
- Biblioshiny for bibliometrix :



- Bibliometrix is developed by Massimo Aria and Corrado Cuccurullo.

  It is an open-source tool for quantitative research in scientometrics and bibliometrics that includes all the main bibliometric methods of analysis.With **biblioshiny**, the shiny app introduced from version 2.0, bibliometrix has become very easy to use even for those who have no coding skills. Bibliometrix package provides various routines for importing bibliographic data from SCOPUS, Clarivate Analytics' Web of Science, PubMed, Digital Science Dimensions and Cochrane databases, performing bibliometric analysis and building data matrices for co-citation, coupling, scientific collaboration analysis and co-word analysis.

QUESTION 2

1. Using VOSviewer :

VOSviewer is a software tool for creating maps based on network data and for visualizing and exploring these maps. The functionality of VOSviewer can be summarized as follows:

a. Creating maps based on network data. A map can be created based on a network that is already available, but it is also possible to first construct a network. VOSviewer can be used to construct networks of scientific publications, scientific journals, researchers, research organizations, countries, keywords, or terms. Items in these networks can be connected by co-authorship, co-occurrence, citation, bibliographic coupling, or co-citation links.To construct a network, data from Web of Science,Scopus, PubMed, RIS, or Crossref JSON files can be used.

b. Visualizing and exploring maps. VOSviewer provides three visualizations of a map: The network visualization, the overlay visualization, and the density visualization. Zooming and scrolling functionality allows a map to be explored in full detail, which is essential when working with large maps containing thousands of items

c. Using bib file exported from scopus ,a map network is created to visualize co-authorship links using different weights : documents, links and total link strengths

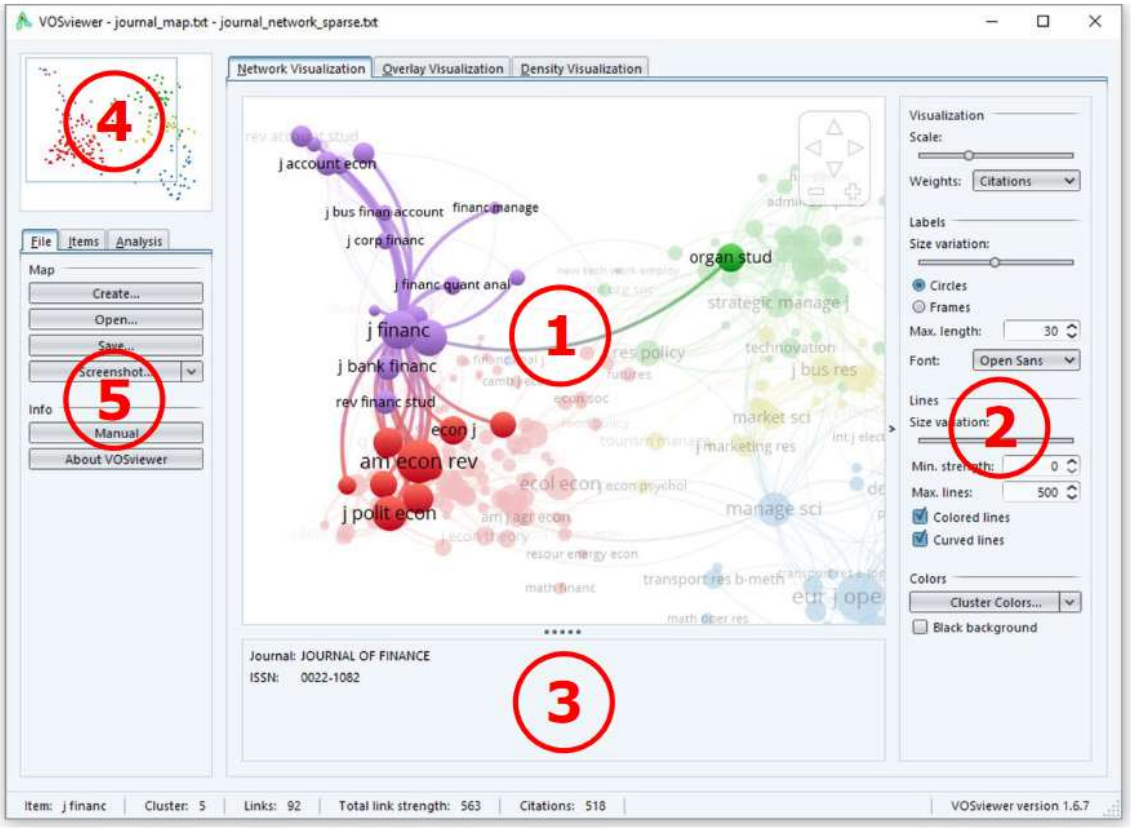


Figure 1. Main window of VOSviewer. The numbers designate (1) the main panel, (2) the options panel, (3) the information panel, (4) the overview panel, and (5) the action panel.

QUESTION 3

1. **Co-Authorship relation Prediction :**

   - Language : python(Anaconda Jupyter Notebook)
   - Based on the Author's Source title, correspondence address and Article title his relationship with other authors is measured.
   - Using cosine Similarity between these features , the relationship measure is predicted.
   - To use cosine similarity we need to tokenize the text in those features, so natural language processing techniques such as stemming, tfidf are used to create a matrix used for cosine similarity.

2. **Co-Citation Prediction :**

   - Language : Python(Anaconda Jupyter Notebook)
   - In que 1, we discussed about R package Biblioshiny ,using this package "Annual_Total_Citation_per_Year.csv" is created using scopus dataset.
   - Which consists of Year wise mean ciation per year and the year.
   - Using Time series ML algorithm, we can predict the no. of citations or mean citation per year in the next coming year.
   - To implement, we used Autoregression(imported directly using python statsmodels.tsa modules)to train and test the data .

❖ The following are the results of the given 3 project questions, implemented using above techniques. (code, data, and output files are attached along with this pdf in a zipfile)

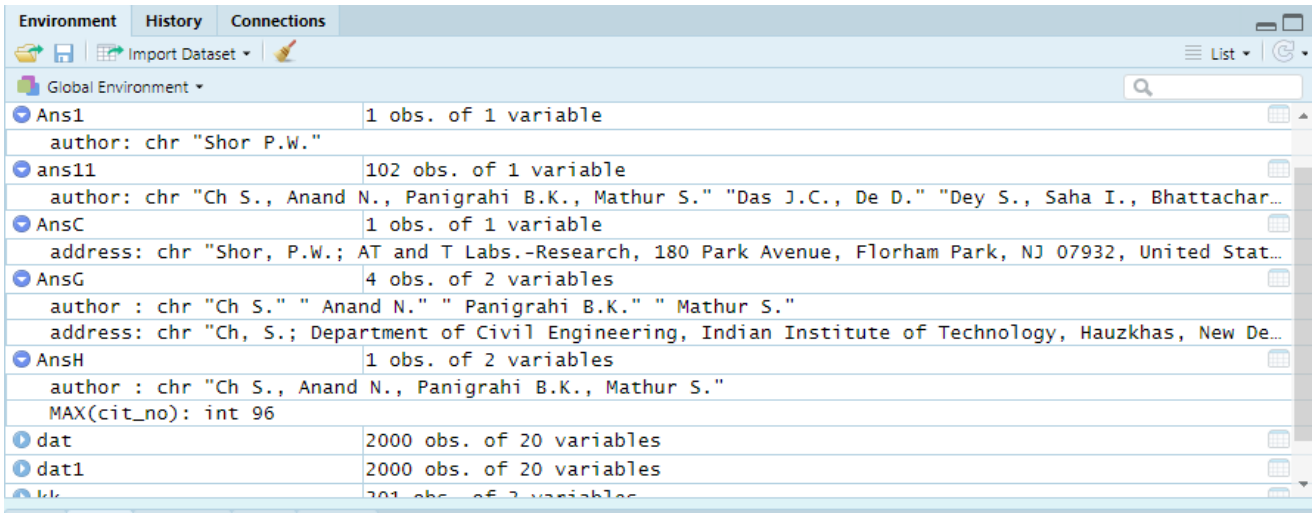
RESULTS

1. The task is to create and analyse Co-Authorship network using Scopus or Web of science databases

   Keyword : Quantum Computing.
   Dataset : Scopus

   a. Highest cited author and his h-index (from the world)

   Ans : **Shor P.W. , 2**



b. Highest publication author

Ans : **HAGHPARAST M**

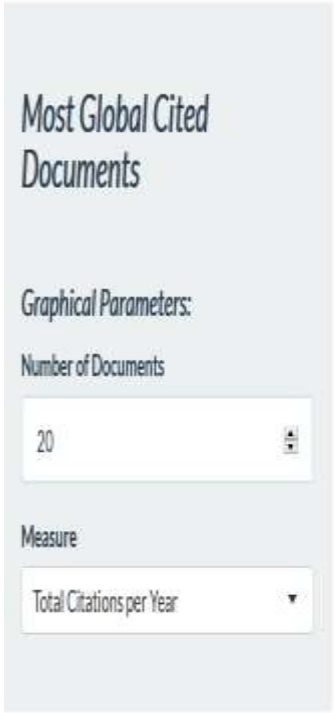
Plot Table

Show 20 rows Copy CSV Excel PDF Print Search:

Authors	Articles	Authors-Frac	Articles Fractionalized
HAGHPARAST M	17	AL-RABADI AN	9.5000
QIU D	16	HAGHPARAST M	7.5833
GAO H	15	QIU D	7.2333
WANG Y	15	DE WOLF R	6.6167
ZHANG Y	15	AARONSON S	6.1667
LI L	14	AMBAINIS A	5.9500
WANG X	14	ZHANG Y	5.2000

c. Highest cited authors avg. citations, and the country name.

Ans : **Shor, P.W, 141.2917, United States**



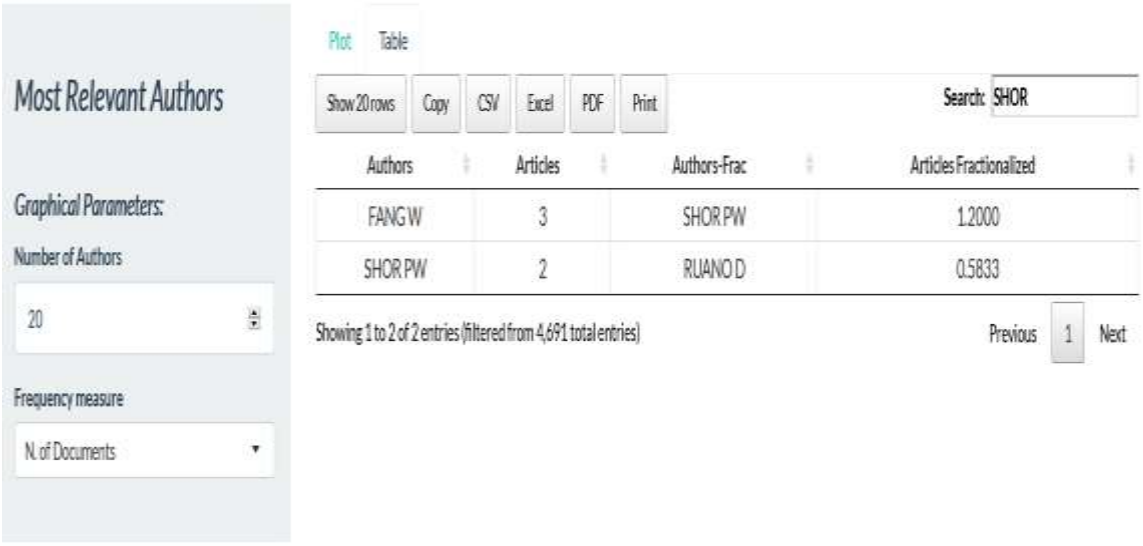
Plot Table

Show 20 rows Copy CSV Excel PDF Print Search:

Paper	Total Citations	TC per Year
SHOR PW, 1997, SIAM J COMPUT	3391	141.2917
PHILIP CHEN CL, 2014, INF SCI	1378	196.8571
BENNETT CH, 1992, J CRYPTOLOGY	1290	44.4828
HAN KH, 2002, IEEE TRANS EVOL COMPUT	1100	57.8947
KENDALL RA, 2000, COMPUT PHYS COMMUN	663	31.5714
BENNETT CH, 1997, SIAM J COMPUT	627	26.1250
WEIDEMAN JAC, 2000, ACM TRANS MATH SOFTWARE	532	25.3333
HAN KH, 2004, IEEE TRANS EVOL COMPUT	445	26.1765
BLANCO MA, 2005, J CHEM THEORY COMPUT	375	23.4375

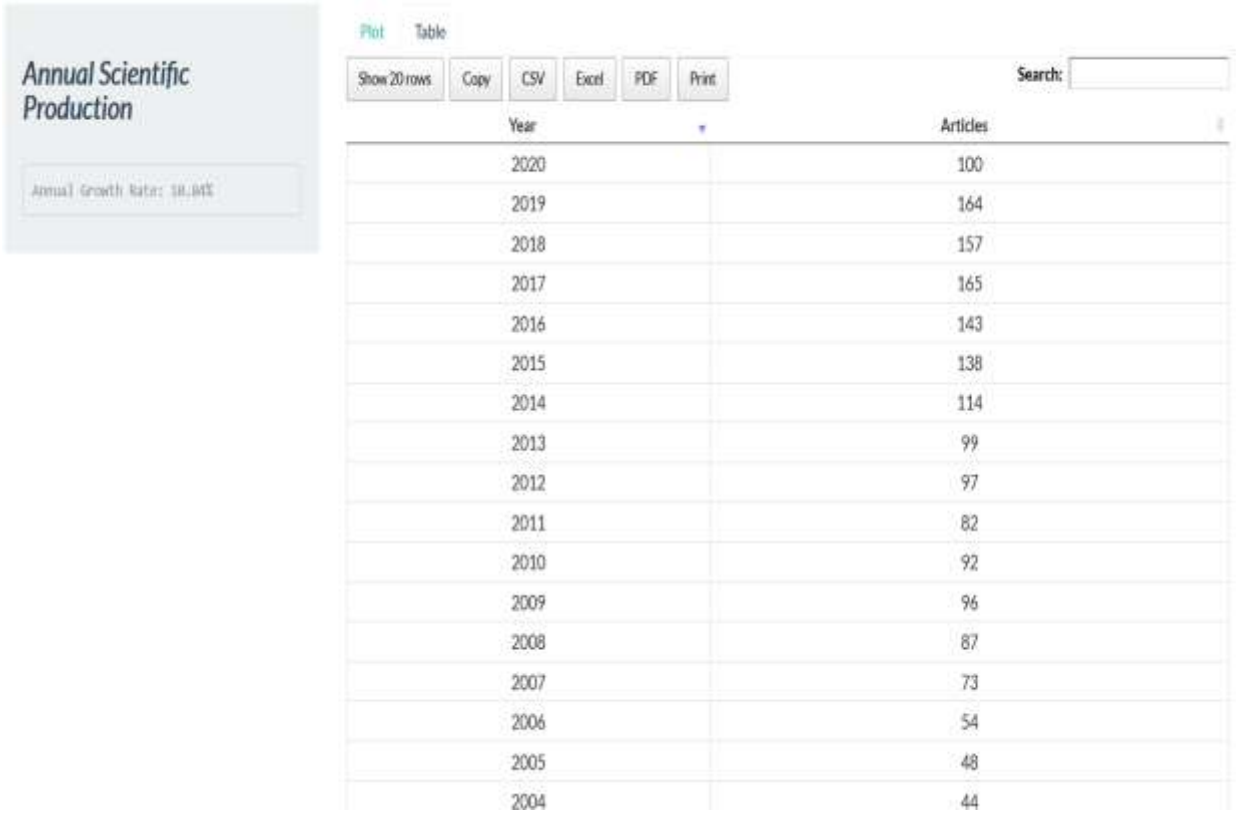
d. Total number of publications of the highest cited author

Ans : **2**



e. Total publication in year

Ans:



f. Total citation per year

Ans :

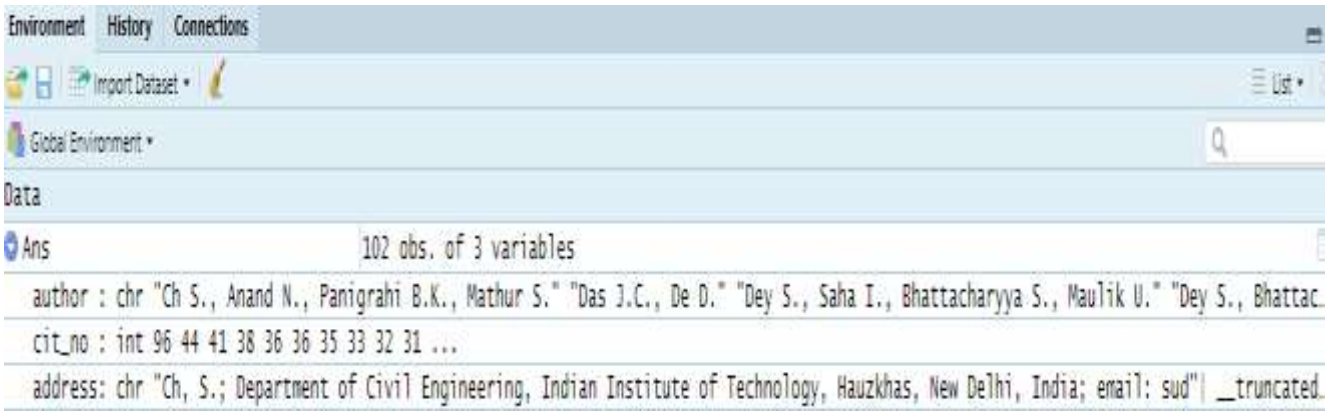
Plot Table

Show 20 rows Copy CSV Excel PDF Print Search:

Paper	Total Citations	TC per Year
SHOR PW, 1997, SIAM J COMPUT	3391	141.2917
PHILIP CHEN CL, 2014, INF SCI	1378	196.8571
BENNETT CH, 1992, J CRYPTOLOGY	1290	44.4828
HAN KH, 2002, IEEE TRANS EVOL COMPUT	1100	57.8947
KENDALL RA, 2000, COMPUT PHYS COMMUN	663	31.5714
BENNETT CH, 1997, SIAM J COMPUT	627	26.1250
WEIDEMAN JAC, 2000, ACM TRANS MATH SOFTWARE	532	25.3333
HAN KH, 2004, IEEE TRANS EVOL COMPUT	445	26.1765
BLANCO MA, 2005, J CHEM THEORY COMPUT	375	23.4375
BENNETT CH, 2014, THEOR COMPUT SCI	319	45.5714
AKTULGA HM, 2012, PARALLEL COMPUT	315	35.0000

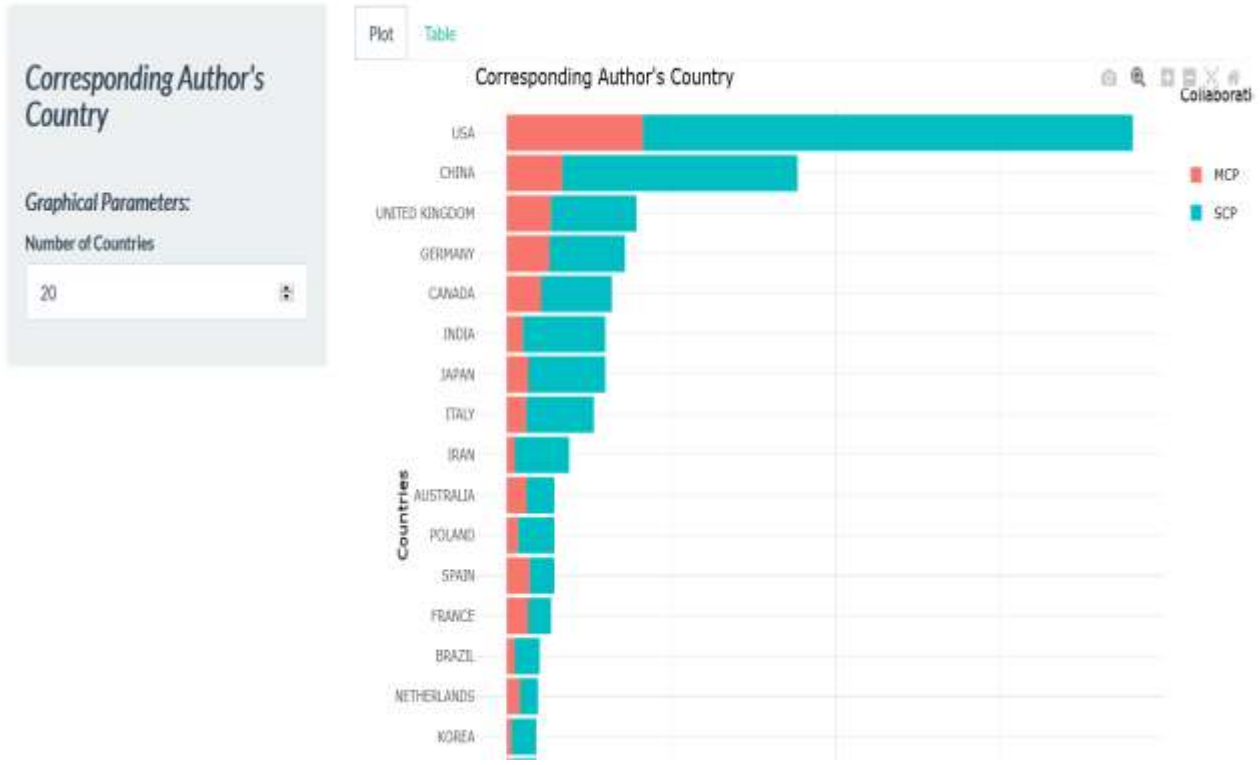
g. Highest cited author from india and the university

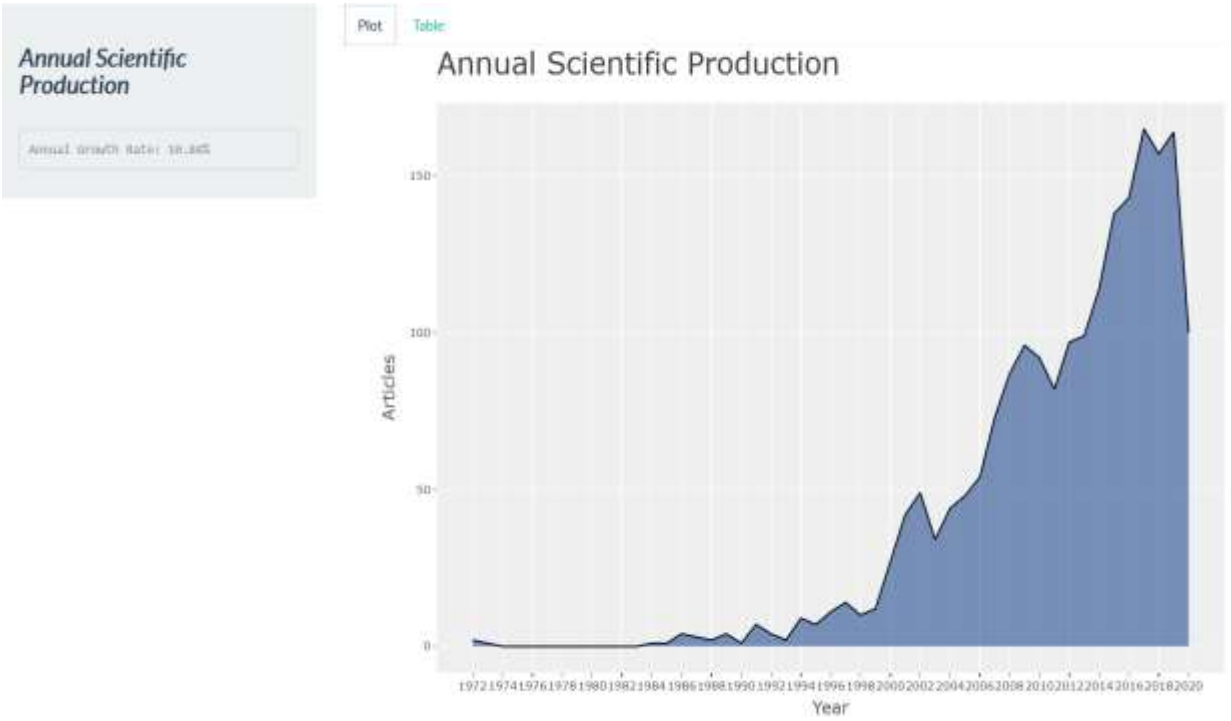
Ans : **Ch S., Department of Civil Engineering, Indian Institute of Technology , Hauzkhas, New Delhi, India**



i. Comparative year wise article publication analysis of india, china and usa.

Ans :





k . Country wise total number of publication

Ans:

Plot Table

Show 20 rows | Copy | CSV | Excel | PDF | Print | Search:

Country	Articles	Freq	SCP	MCP	MCP_Ratio
USA	381	0.273314	298	83	0.218
CHINA	177	0.126973	143	34	0.192
UNITED KINGDOM	79	0.056671	52	27	0.342
GERMANY	72	0.051650	46	26	0.361
CANADA	64	0.045911	43	21	0.328
INDIA	60	0.043042	50	10	0.167
JAPAN	60	0.043042	47	13	0.217
ITALY	53	0.038020	41	12	0.226
IRAN	38	0.027260	33	5	0.132
AUSTRALIA	29	0.020803	17	12	0.414
POLAND	29	0.020803	22	7	0.241
SPAIN	29	0.020803	15	14	0.483
FRANCE	27	0.019369	14	13	0.481
BRAZIL	20	0.014347	15	5	0.250
NETHERLANDS	19	0.013630	11	8	0.421
KOREA	18	0.012912	15	3	0.167
TAIWAN	18	0.012912	13	5	0.278

7:0.0:1:3052/#tab-7389-2

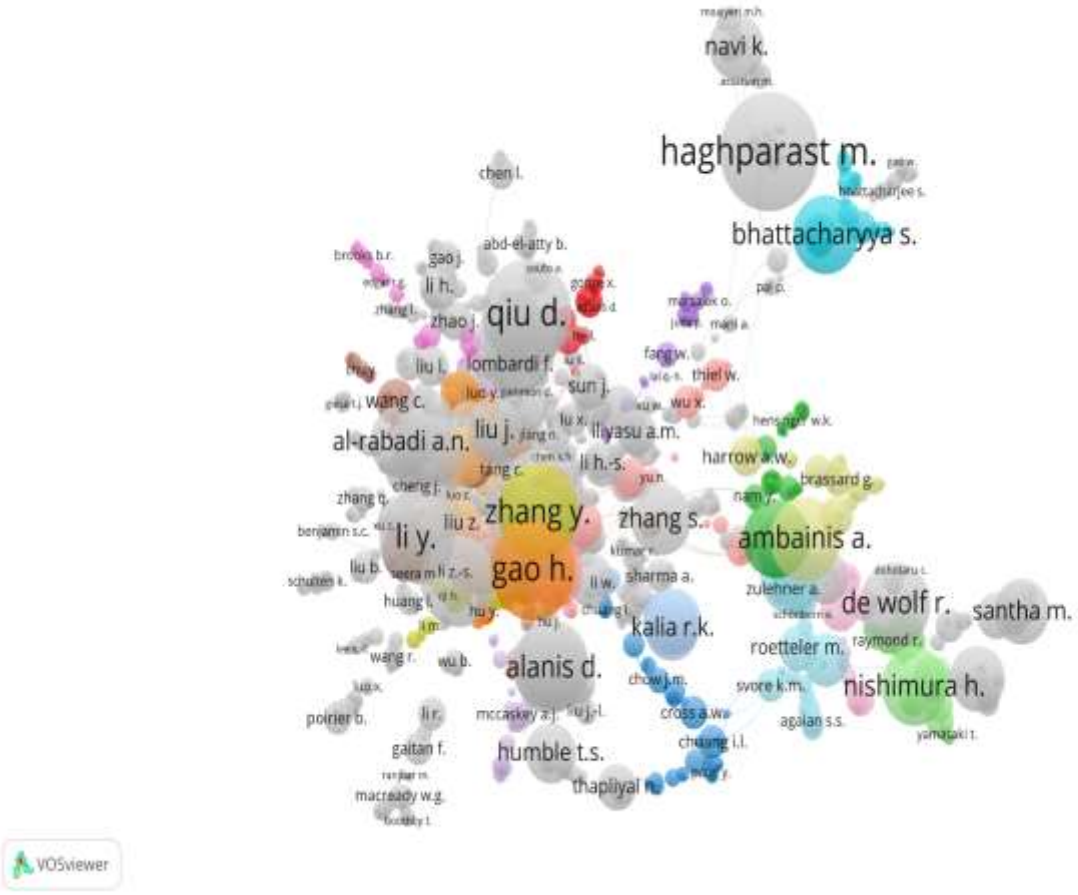
2. Analyse the data(eg. co-authorship links) using visual tools like Gephi or VOSviewer

Ans : Visual Tool Used : VOS viewer

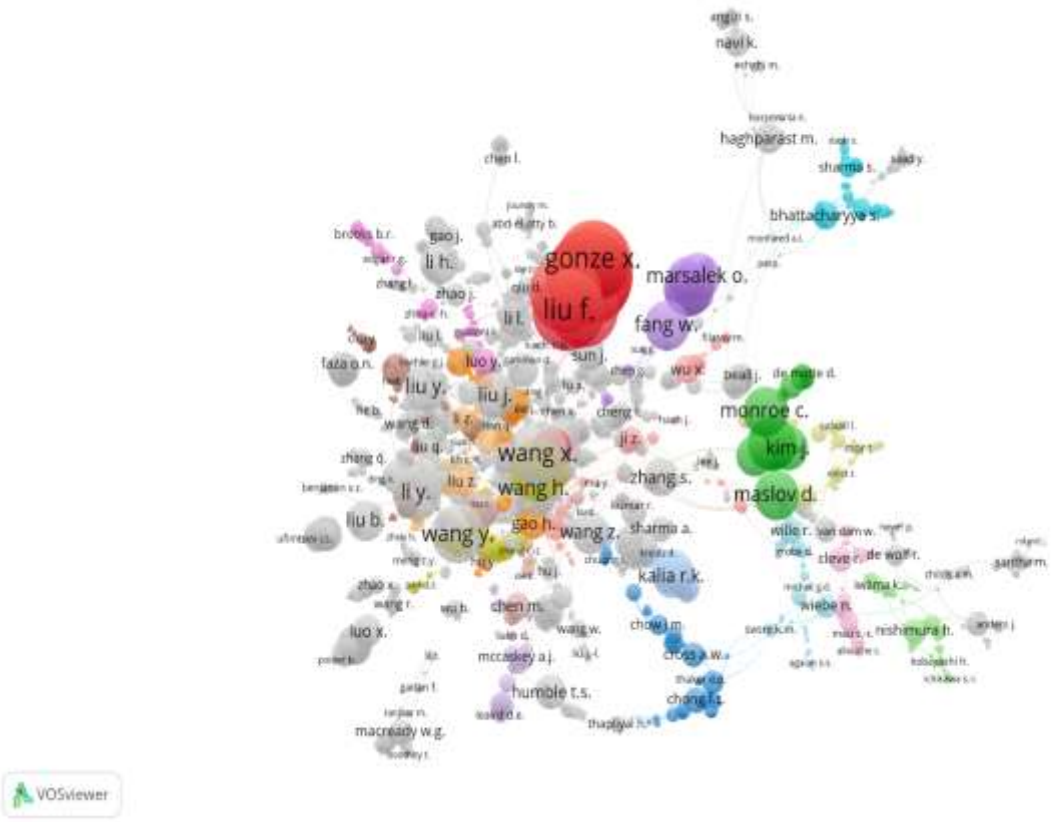
Analysis of co-authorship links :

Nodes : Authors
Normalization : Association method

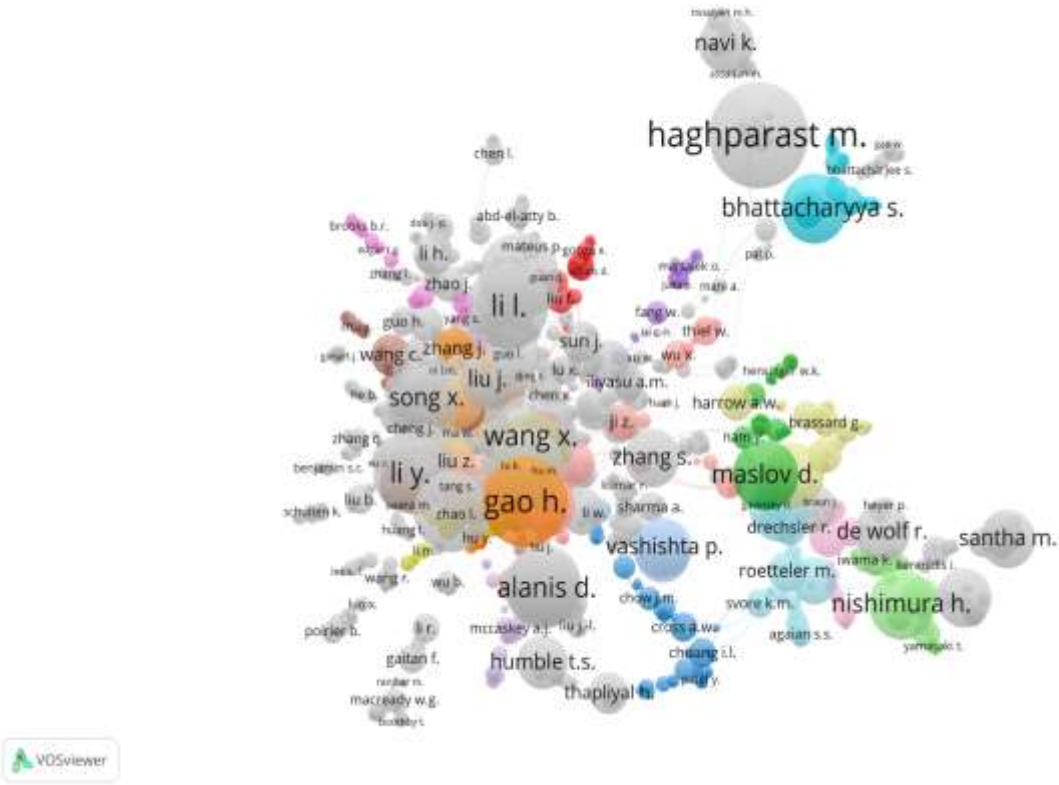
1. Based on weights : documents



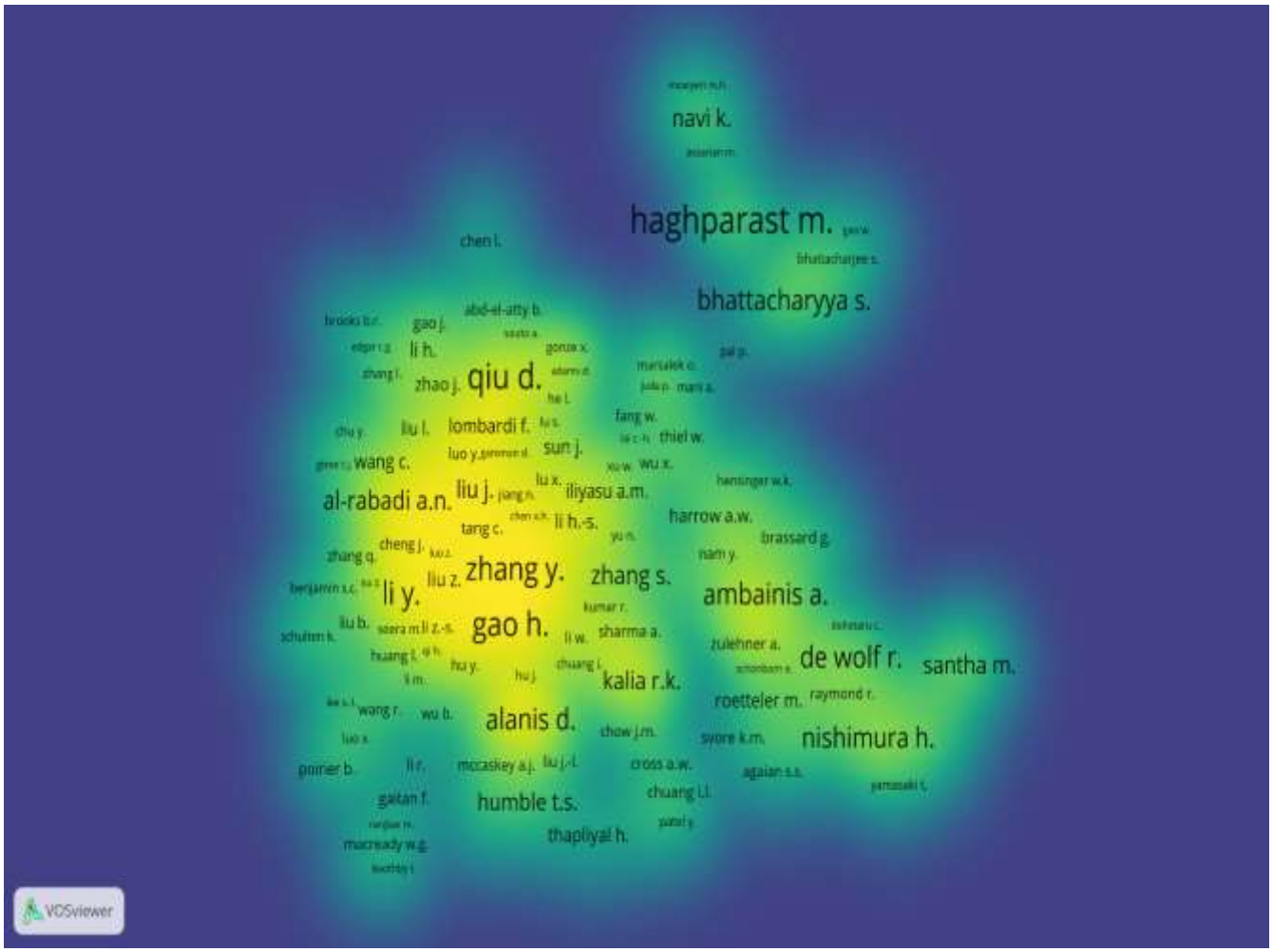
2. Based on weights : Links b/w Authors



3. Based on weights : Total Link Strength



4. Density Visualaization :



3. Co-Author Relationship Prediction and Citation prediction using known machine learning techniques

Ans : 3.1 : Co-Author Relationship Prediction

```
In [38]: co_authors = []
for key,value in enumerate(sorted_similar_Author_scores):
    if value[1]>=0.97:
        co = get_name_from_index(value[0])
        if co != author and co != '[]' :
            co_authors.append(co)
            print('Author {}-----already in Co_Author Relationship----------> {}'.format(author,co))
    if value[1]>0.5 and value[1]<0.97:

        co = get_name_from_index(value[0])
        if co != author and co != '[]' :

            co_authors.append(co)
            print('Author {}----- Co_Author Relationship----------> {}'.format(author,co))
```

```
Author ['Degregorio N., Iyengar S.S.']----- Co_Author Relationship----------> ['Jakowski J., Sumner I., Iyengar S.S.']
Author ['Degregorio N., Iyengar S.S.']----- Co_Author Relationship----------> ['Li J., Li X., Iyengar S.S.']
```

```
In [ ]:
```

3.2 : Citation Prediction

```
In [35]: from statsmodels.tsa.ar_model import AutoRegResults
import numpy
# load model
model = AutoRegResults.load('ar_model.pkl')
data = numpy.load('ar_data.npy')
last_ob = numpy.load('ar_obs.npy')
# make prediction
predictions = model.predict(start=len(data), end=len(data))
# transform prediction
yhat = predictions[0] + last_ob[0]
print('Prediction: %f' % yhat)
```

```
Prediction: 4.342005
```