

1. Go to the Scopus website and search for relevant papers
2. For example, we searched for each journal of interest by each year (example below: *Trends in Neurosciences*, 2020)

## 99 document results

SRCTITLE (trends AND in AND neurosciences) AND (LIMIT-TO (PUBYEAR, 2020)) AND (LIMIT-TO (EXACTSRCTITLE, "Trends In Neurosciences"))

Edit Save Set alert

CSV file exported. See your downloaded file for more details.

Search within results...

Refine results

Limit to Exclude

Open Access

Year

2020 (99)

Author name

Bayne, T. (2)

Massimini, M. (2)

Seth, A.K. (2)

Adalikkan, C. (1)

Documents Secondary documents Patents View Mendeley Data (263) Search other databases

Analyze search results Show all abstracts Sort on: Cited by (highest)

All CSV export Download View citation overview View cited by Save to list

	Document title	Authors	Year	Source	Cited by
1	The Brain's Glymphatic System: Current Controversies <i>Open Access</i>	Mestre, H., Mori, Y., Nedergaard, M.	2020	Trends in Neurosciences 43(7), pp. 458-466	181
View abstract > YALE > LINKS Related documents Paperpile					
2	Astrocytes and Microglia: In Sickness and in Health <i>Open Access</i>	Vainchtein, I.D., Molofsky, A.V.	2020	Trends in Neurosciences 43(3), pp. 144-154	165
View abstract > YALE > LINKS Related documents Paperpile					

3. Click "CSV export" and export the data:

Export document settings

You have chosen to export 99 documents

Select your method of export

Mendeley ExportLibris RIS Format EndNote, Reference Manager CSV Excel BibTeX Plain Text ASCII in HTML

What information do you want to export?

Citation information	Bibliographical information	Abstract & keywords	Funding details	Other information
<input checked="" type="checkbox"/> Author(s)	<input type="checkbox"/> Affiliations	<input type="checkbox"/> Abstract	<input type="checkbox"/> Number	<input type="checkbox"/> Tradenames & manufacturers
<input checked="" type="checkbox"/> Author(s) ID	<input type="checkbox"/> Serial identifiers (e.g. ISSN)	<input type="checkbox"/> Author keywords	<input type="checkbox"/> Acronym	<input type="checkbox"/> Accession numbers & chemicals
<input checked="" type="checkbox"/> Document title	<input type="checkbox"/> PubMed ID	<input type="checkbox"/> Index keywords	<input type="checkbox"/> Sponsor	<input type="checkbox"/> Conference information
<input checked="" type="checkbox"/> Year	<input type="checkbox"/> Publisher		<input type="checkbox"/> Funding text	<input type="checkbox"/> Include references
<input checked="" type="checkbox"/> EID	<input type="checkbox"/> Editor(s)			
<input checked="" type="checkbox"/> Source title	<input type="checkbox"/> Language of original document			
<input checked="" type="checkbox"/> volume, issue, pages	<input type="checkbox"/> Correspondence address			
<input checked="" type="checkbox"/> Citation count	<input type="checkbox"/> Abbreviated source title			
<input checked="" type="checkbox"/> Source & document type				
<input checked="" type="checkbox"/> Publication Stage				
<input checked="" type="checkbox"/> DOI				
<input checked="" type="checkbox"/> Open Access				

Cancel Export

4. Run `sc_by_pair.py`, modifying the folder structure as necessary. In our case, we had three folders ("All\_Neuro", "All\_Neurology", "All\_Psychiatry"). Within each of those three folders were folders for each journal, which included the exported .csv files. (**Note:** the download requirements differ from those in the "raw\_data\_analysis" folder because here the database is at the level of a citing/cited pair, not at the whole-article level).

5. Use [genderize.io](https://genderize.io) to probabilistically assign gender to each name. We recommend using the .csv tool, but you may also use their Python API. Please note that this requires a paid subscription.
6. Run `model_building.py` to build logistic regression models. You may enter the "--y" argument as 1) "pairs," in which case it runs the citing/cited pairs model, or 2) "extreme," in which case it runs the highly self-citing model