How to select an area of research?

Trisha Sen

Georgia Institute of Technology

Web Scraping

term frequency –inverse document frequency

Tokenized Journal Abstracts

tf-idf to convert abstracts to numerical vectors

Topic Extraction

2 step Non-Negetive Matrix Factorization (NMF)

15 topics were extracted

33,082 articles from 1996-2019
32MB of data analyzed
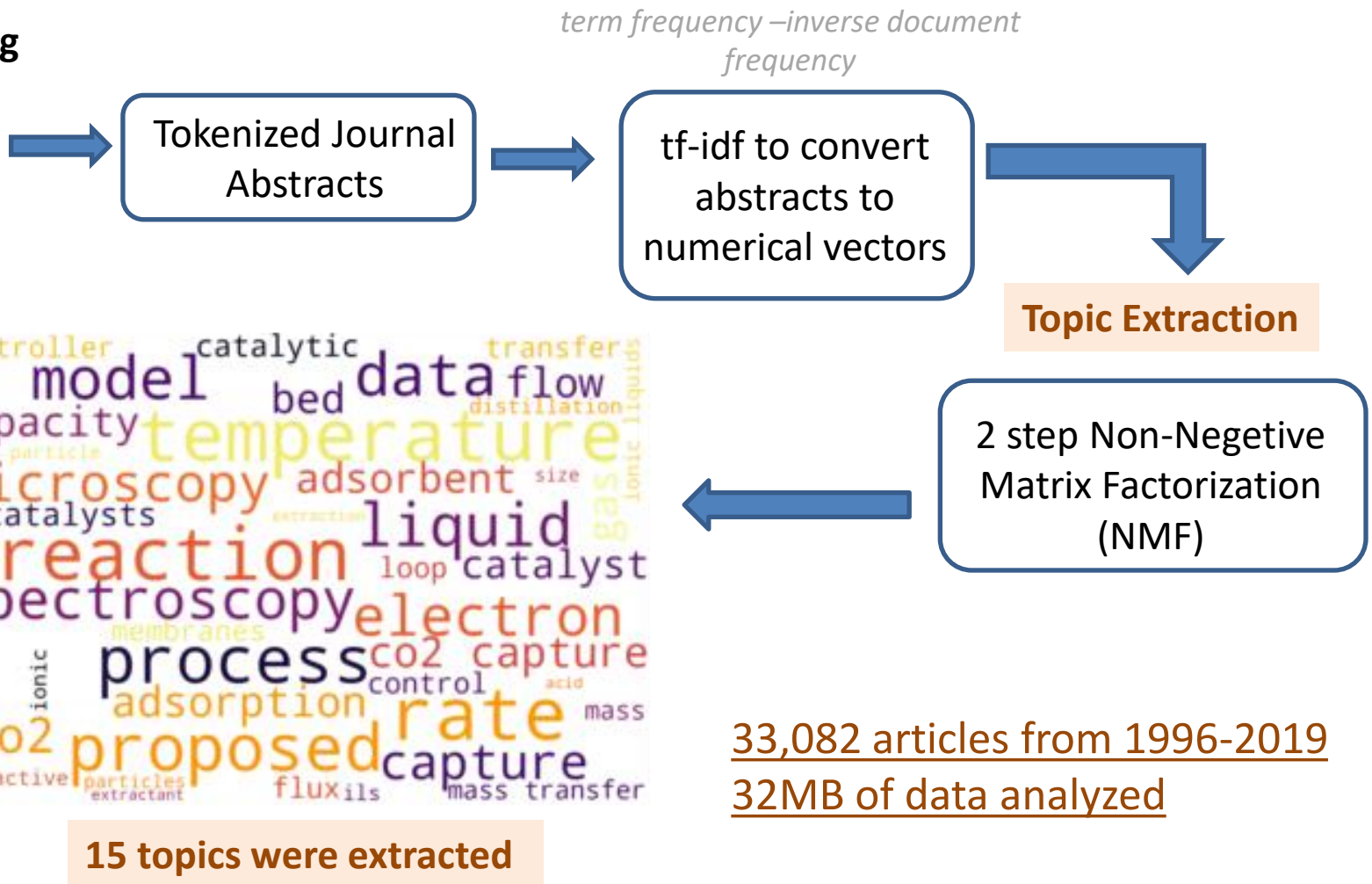
The WordCloud shows the top 3 words per topic and their relative frequency in the journal

Popularity is a measure of the percentage of research dealing with a given topic
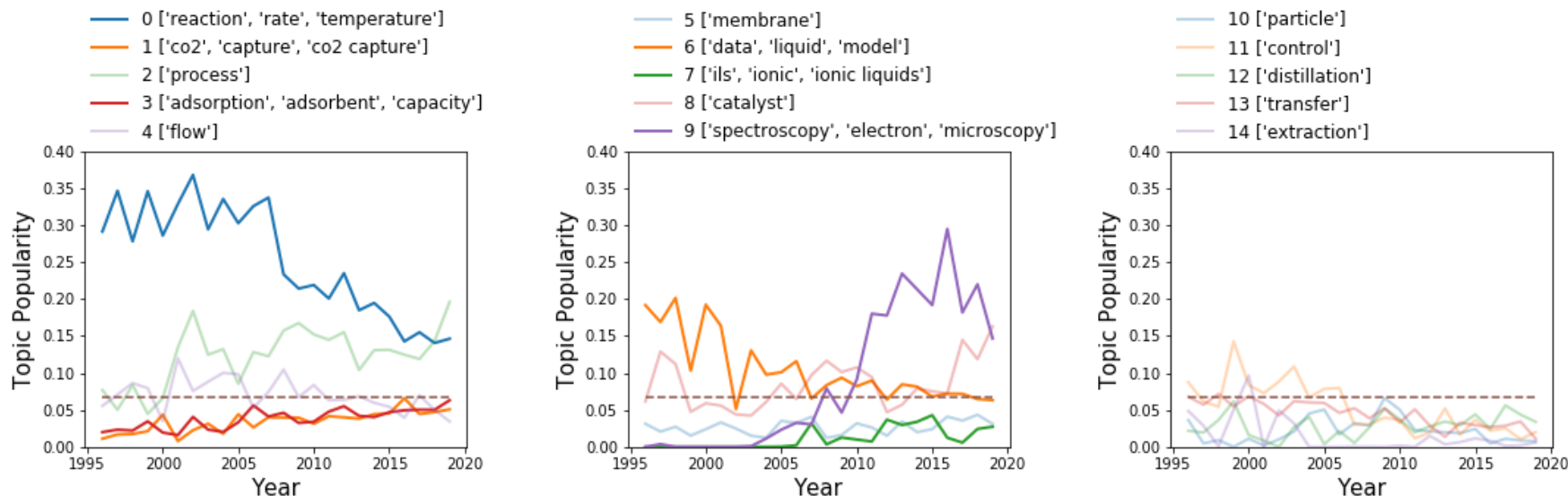
[9] *"Spectroscopy & Electron Microscopy"* , [7] *"Ionic Liquids"* emerged suddenly around 2005 and are plateauing now.
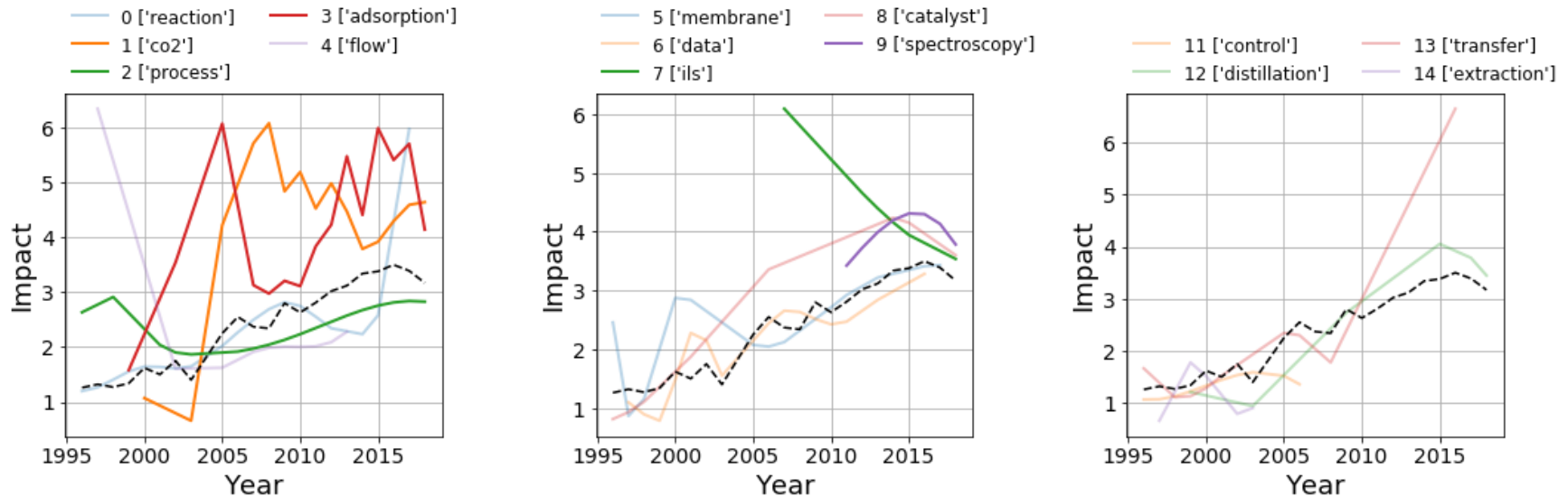
[1] *"CO_2 capture"* , [3] *"Adsorption"* , are seeing steady growth

[0] *"Reaction Engineering"* , [6] *"Fluid Mechanics"* are seeing a dramatic and steady decline

[12] *"Distillation"* , [5] *"Membranes"* , are small but steady fields

Impact is a measure of general outreach & relevance of research findings in a field


[9] *"Spectroscopy & Electron Microscopy"* , [7] *"Ionic Liquids"* started strong but are seeing steep decline in *Impact*

[1] *"CO_2 capture"* , [3] *"Adsorption"* , are the **"Hot Topic"** fields getting cited significantly more than the average (although number of contributions is lesser than average)

[2] *"Process Modeling"* has a lower than average outreach although the number of publications is well above the average

Backup Slides

What percentage of authors end up being cited in each research area?
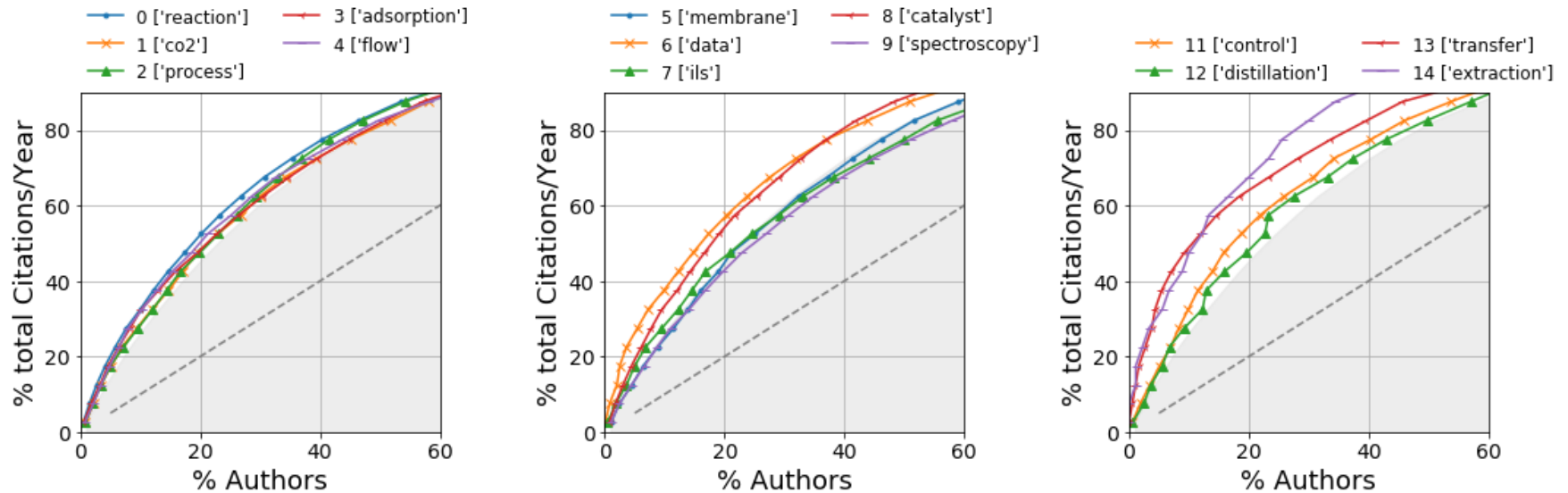
Degree of collaboration overall