

Visualizing Patterns of Citation After a Retraction Event

Brett Williams
Senior Research Analyst

Purpose

1. Looking for a way to visualize what happens after a specified 'event'
2. Increasing concern in many areas around replication and academic misconduct
3. Develop tools and analysis practices to identify key people and organizations

The Event

Retraction Watch

January to September 2012 a total of 20 papers with Dr Das as a key author were retracted.

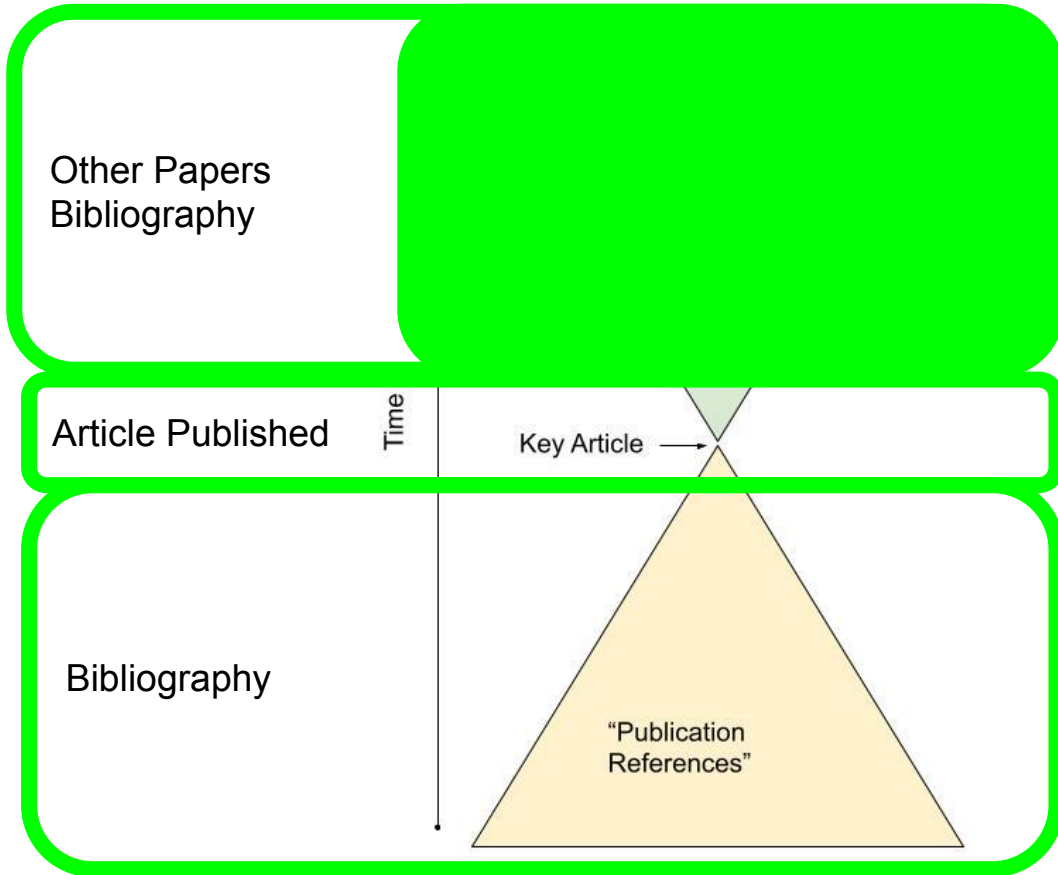
As of 2023, the total is 22. The Retraction Watch database has classified these papers as

Falsification/Fabrication of Data Intentional changes to data so that it is not representative of the actual finding

<https://retractionwatch.com/category/by-author/dipak-das/>

The Sample





20 Articles Retracted March- Sep 2012

1. News Release January 11, 2012
2. Das was a highly cited author and director of a lab
3. Single year 'event' enables a before and after comparison.
4. Das claims about heart health impact of broccoli and red wine were widely reported in the media

Visualization 1 - Baseline Prior to 2012

Methods

Using the Base Sample CSV & my custom formatting tool, I created properly formatted Dimension searches for the incoming citations for each of the 20 sample papers

Time: 2002-2012. 19 of 20 articles had incoming citations

These were exported from Dimensions and mapped using VosViewer

Example Dimensions Search

Search created using in-house URL formatting tool.



2002 OR 2003 OR 2004 OR 2005 OR 2006 OR 2007 OR 2008 OR 2009 OR 2010 OR 2011 OR 2012 ×
Publication Year

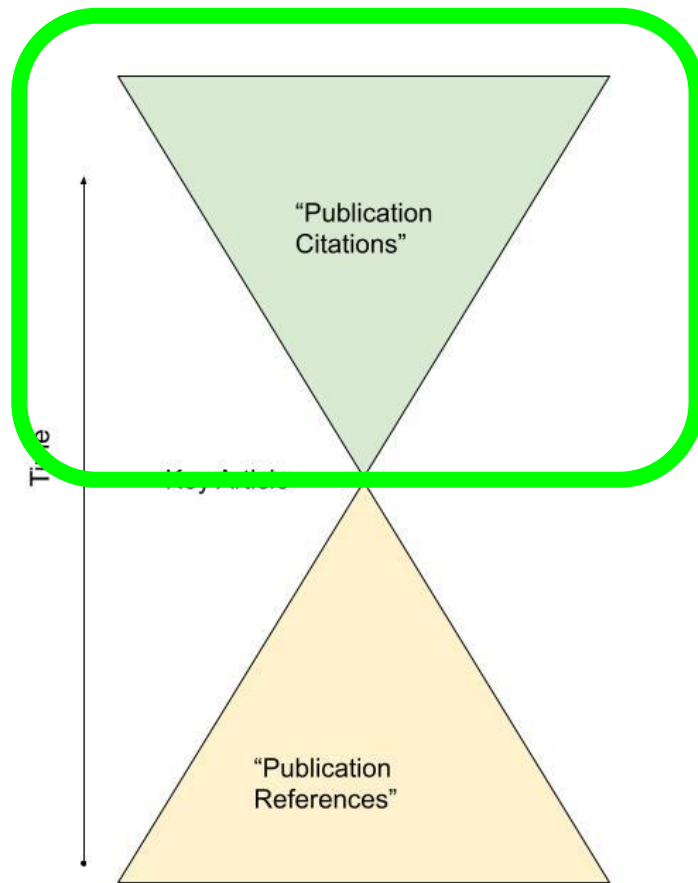
Publication: RETRACTED: Role ... ×
Publication citations

e.g. plastic AND instrument

Search in: Full data Title and abstract DOI



2002 - 2012



Citation Network

Of the incoming citations to the 20 sample articles from 2002 to 2012 there are 673 incoming citations.

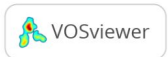
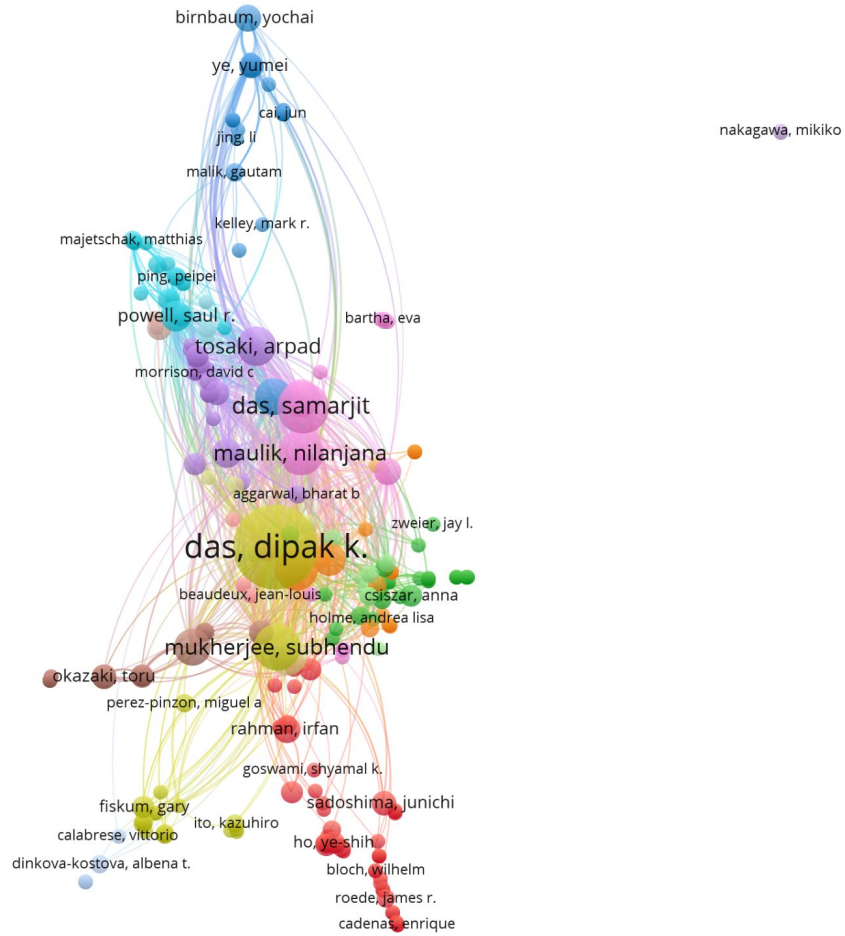
This is the citation network before the retraction event.

<https://tinyurl.com/2342vqb5>

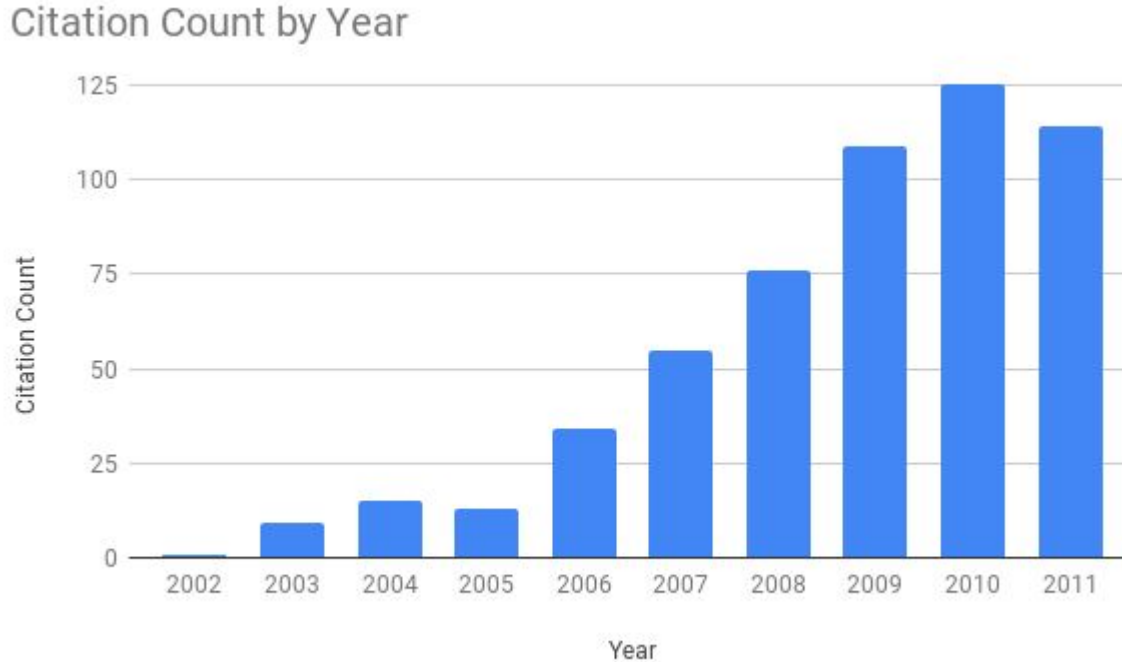


Author network

2 citation to a retracted Das paper limit, the largest cluster of authors



Citation Count by Year 2002-2012



What do you predict happens after 2012?

1. How will the citations behave?
2. Where do you think the citations will be coming from?

Visualization 2 - 2013 to Present

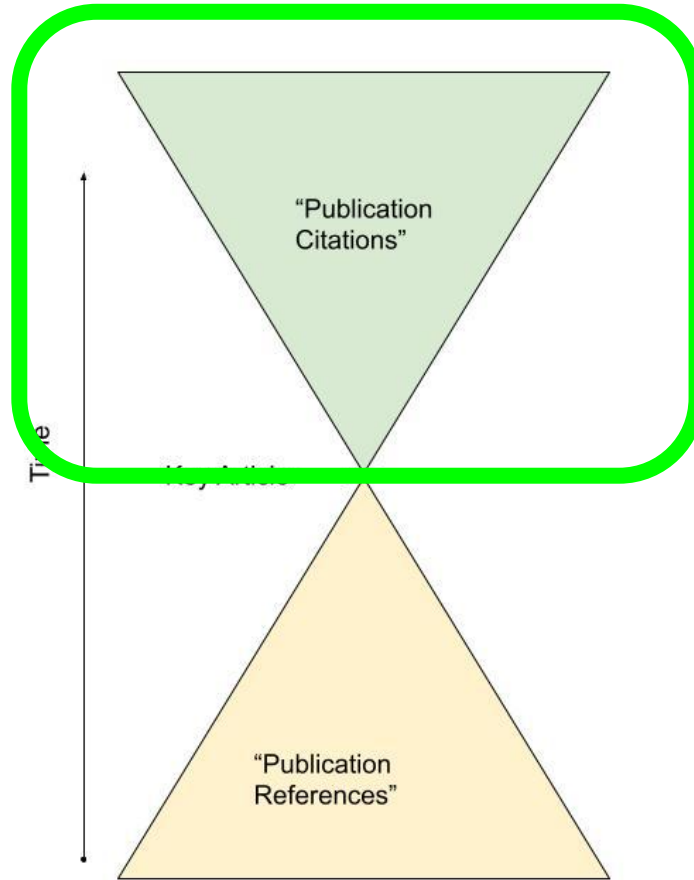
Methods

Using the Base Sample CSV & my custom formatting tool, I created properly formatted Dimension searches for the incoming citations for each of the 20 sample papers from the period 2013 to 2023.

These were exported from Dimensions and mapped using VosViewer



2013 - 2023



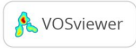
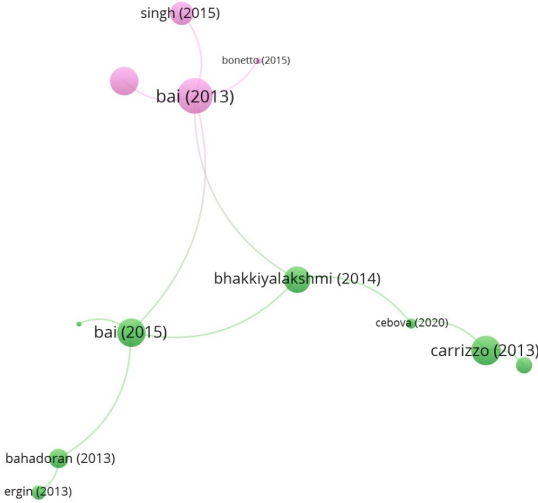
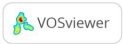
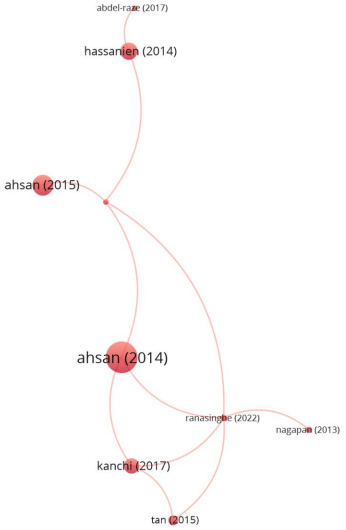
Citation Network

Of the incoming citations to the 20 sample articles from 2013 to 2023 there are 357 incoming citations.

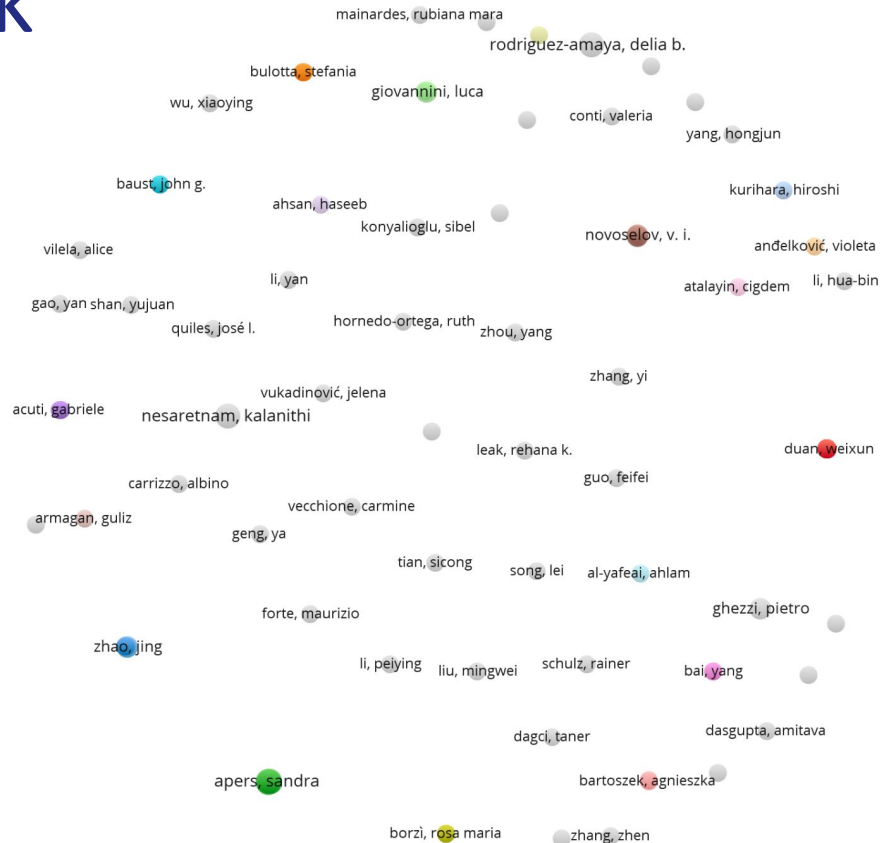
<https://tinyurl.com/253nb4oc>



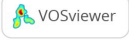
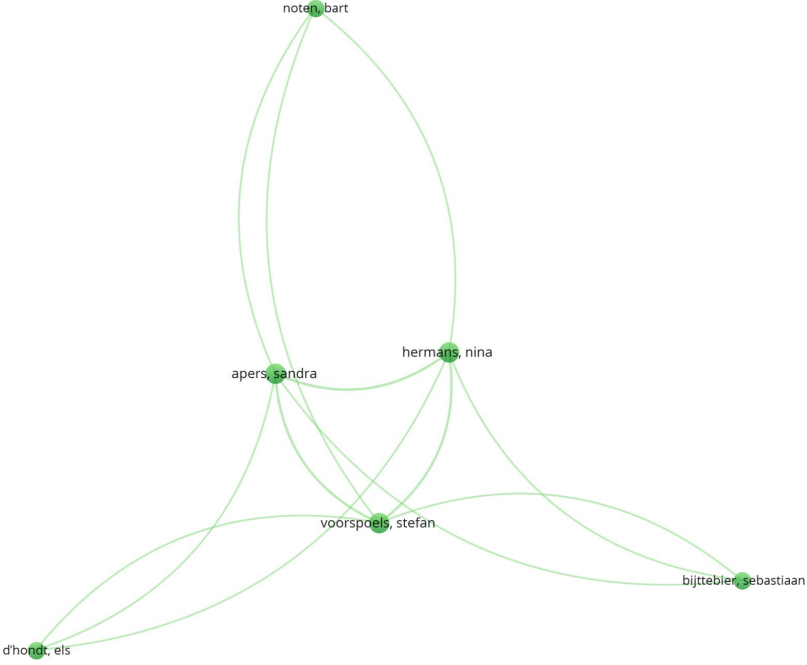
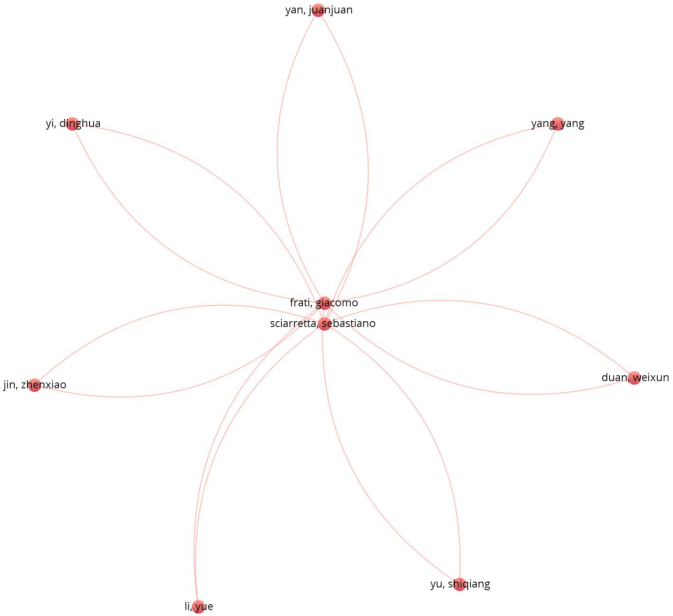
Largest Clusters



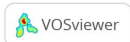
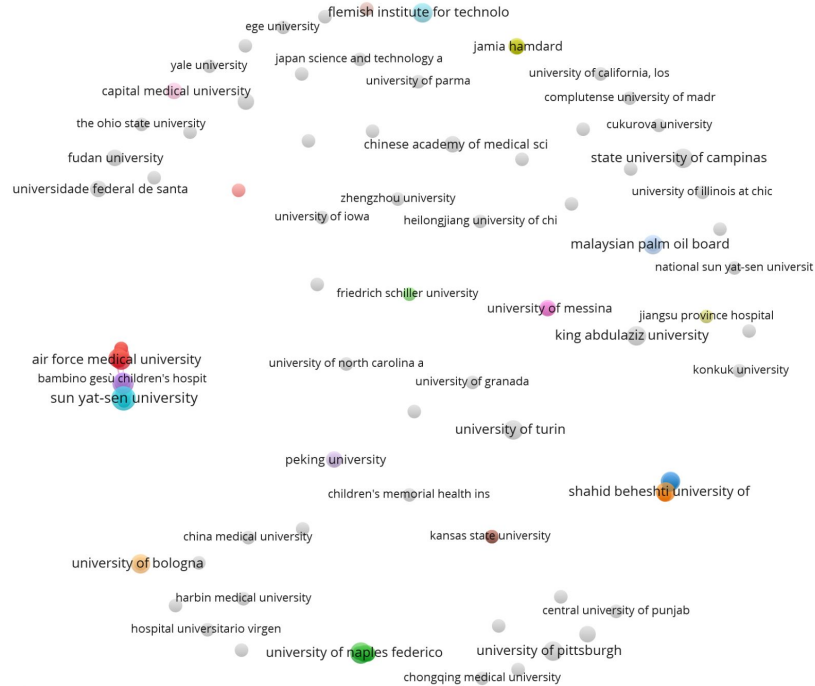
Author Network



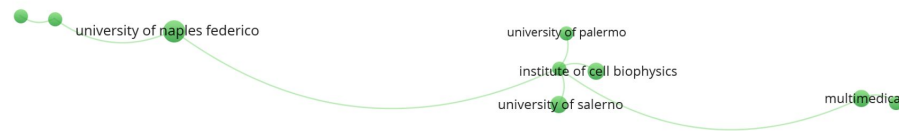
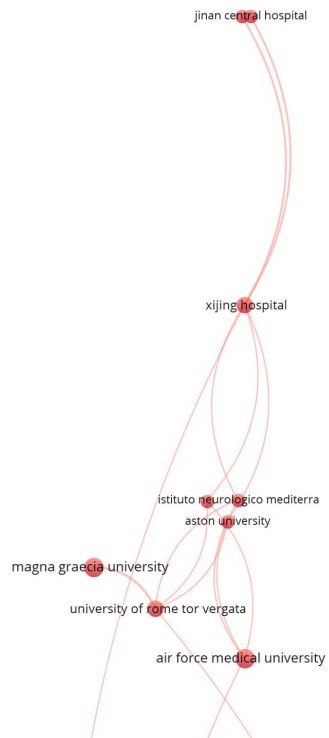
Largest Two Clusters



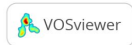
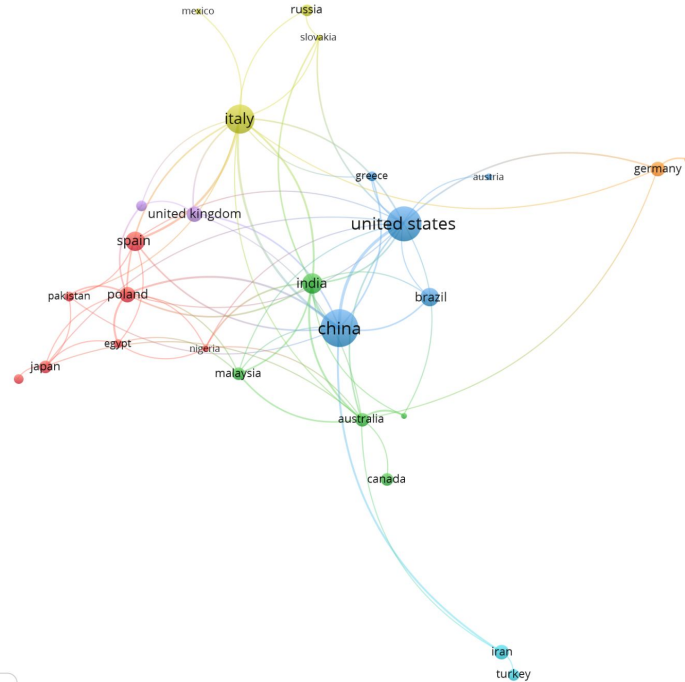
Affiliated Institutions



Largest Two Clusters

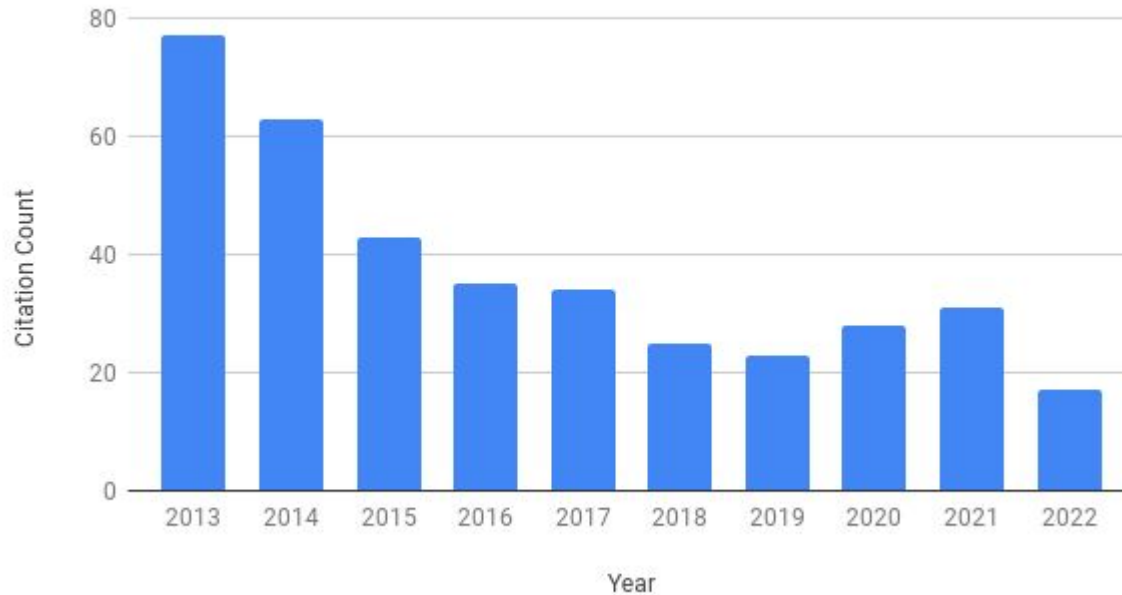


Countries



Citation Count by Year 2013-2023

Citation Count by Year



Lessons Learned

1. There is a baseline of naive references in this sample to retracted papers, even after a very public retraction.
2. Post-retraction citation networks in this sample are
 - a. Less connected at citation level
 - b. Less connected at the institute level
 - c. More international
3. Dimensions data is very rich and can be processed in a variety of different ways with unique insights at the institute & country level.

Additional Steps

1. Confirm results with Publish & Perish with Google Scholar data
2. Build identifying key sources of citation (institute, author & country) into complex research query processes

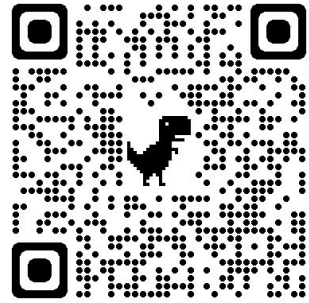
Parallel Research Projects

1. What effect do poor retraction indicators have?
 - a. This research has led to a very strong emphasis on getting researchers to use Zotero in our collaborations because of the convenience of the highlighting feature
2. What effect do article sharing sites like SciHub have on post-retraction citation?
3. Can similar or parallel techniques be applied to grey literature? (exploring with Overton exports)

Key Tools For The Visualization Process

- **Zotero** - Centralize collection and export results in BibTex & CSV
 - Scite Plugin
 - DOI Manager
- **Dimensions** - Free access to mapping data with a verified university email address
- **VOSViewer** - Accurate Map (More accurate, more complicated)
- **Custom Formatting Tools** to speed up search strategy
 - Google Sheets
- **Chrome Plugins**
 - LinkGrabber - grabs direct links to sites for later processing
 - HighlightThis - Lists of DOIs to spot key articles
 - OpenMultipleURLs - Does what it says on the label
 - Unpaywall - Another way to grab PDFs
 - Scite

Tools & techniques detailed here



Williams, B. (2020). Dimensions & VOSViewer Bibliometrics in the Reference Interview. *The Code4Lib Journal*, 47.

<https://journal.code4lib.org/articles/14964>

The VOSviewer software provides easy access to bibliometric mapping using data from Dimensions, Scopus and Web of Science. The properly formatted and structured citation data, and the ease in which it can be exported open up new avenues for use during citation searches and reference interviews. This paper details specific techniques for using advanced searches in Dimensions, exporting the citation data, and drawing insights from the maps produced in VOS Viewer. These search techniques and data export practices are fast and accurate enough to build into reference interviews for graduate students, faculty, and post-PhD researchers. The search results derived from them are accurate and allow a more comprehensive view of citation networks embedded in ordinary complex boolean searches.

Thank you.

@BrettLWilliams on Twitter

brett.williams@oma.org

oma.org

