



Blasting the Big Deal

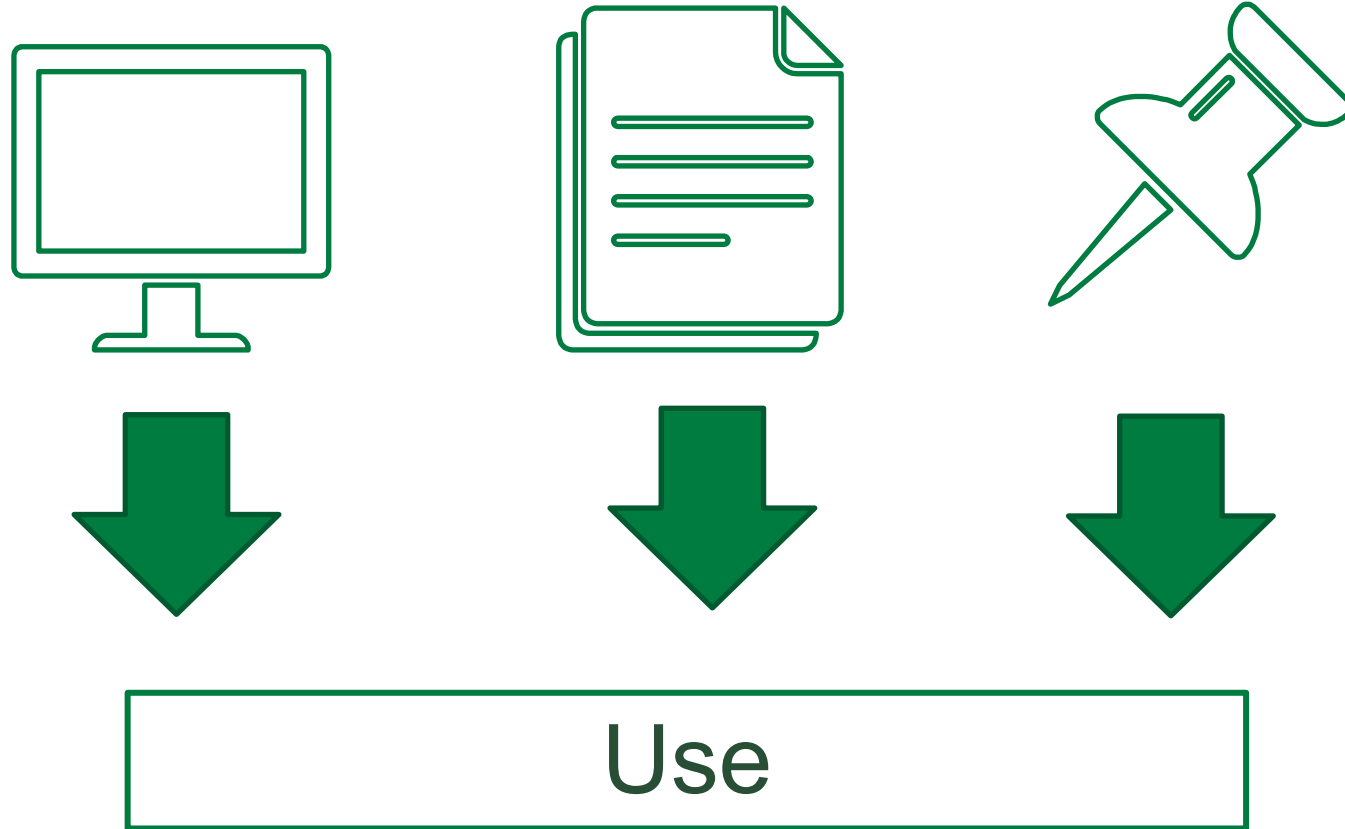
Incorporating Citation Analysis to Support Decision Making
Around Journal Cancellation

27 Apr 2021

Alison Henry

ahenry@ualberta.ca

Background



Goals of this analysis

- Group journals into clusters based on their use, where use includes **views/downloads, publication, and citation.**
- Facilitate a comparison between the Springer journals to which the University of Alberta Library subscribes

Data Gathering

Views/Downloads

Total Item Requests as provided by the vendor

Journal Categories

Category assignments from Web of Science and Scopus

Publication

Web of Science:
UAlberta authors, 2015-2019.

Scopus:
Titles not indexed by Web of Science.

Citation

Web of Science: Cited References.

Data Cleaning and Preparation

- Cited Reference titles are in abbreviated format. ISS staff collected abbreviations from the Web of Science journal abbreviation table. They found abbreviations for **1,819** journals out of a list of **3,024** Springer titles.
- All source titles from the three lists were converted to lower case and white space was trimmed to facilitate matching.
- Data is very positively skewed, so log transformation was performed to improve cluster differentiation.

Summary of preparatory steps

- Total number of publications retrieved from Web of Science: **42,215**
- Number of Springer publications retrieved from Web of Science: **3,262**
- Additional publications added from Scopus: **102**
- Total number of publications: **3,364**
- Number of unique journal titles represented: **817**
- Number of citations: **1,700,867**
- Citations to Springer journals: **107,727**

Data Completeness

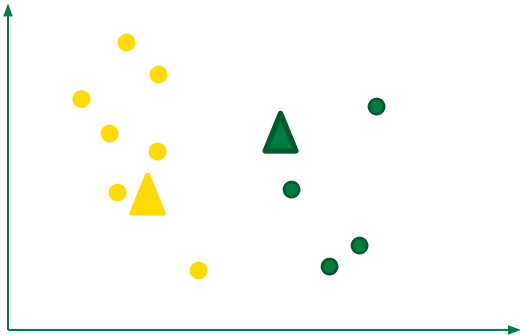
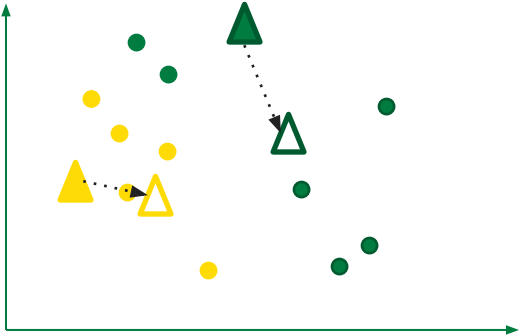
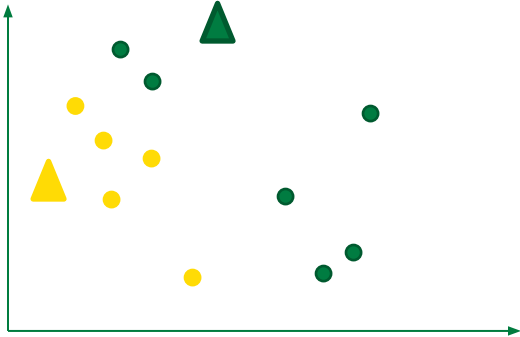
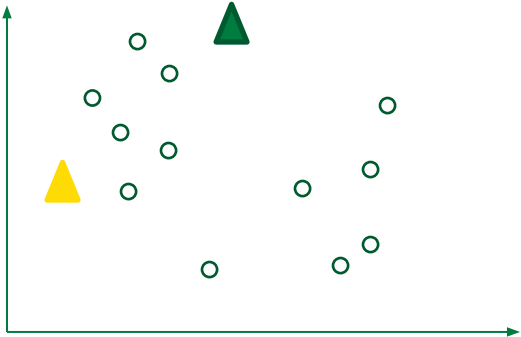
Once the lists were merged, it became apparent that we did not have evidence of all three types of use for all titles. A zero value was entered where no publication, citation, or view/download use was found.

- Titles in which UAlberta authors published: **817**
- Titles for which citations were found: **1,625**
- Titles for which view/download activity was reported: **2,577**

Limitations of this analysis

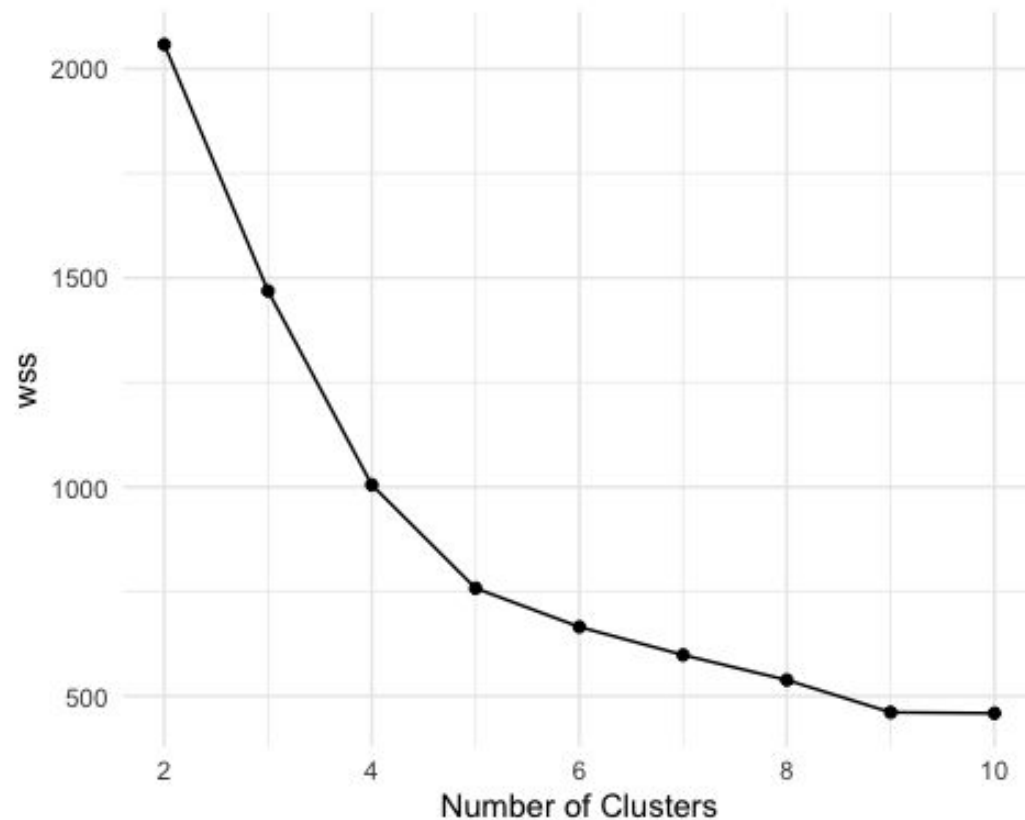
- Single source of citations
- 5 years of publication and citation data, but only one year of Item Request data. Publications and Citations were summed over the 5-year period.
- Text matching for titles rather than using ISSNs
- Missing Category assignments for some journals

K-Means Clustering



Elbow graph

- ▷ In order to determine the optimal number of clusters, we ran the clustering algorithm times multiple times, with different values of k , and plotted the within clusters sum of squares.



Graph to determine optimal number of clusters

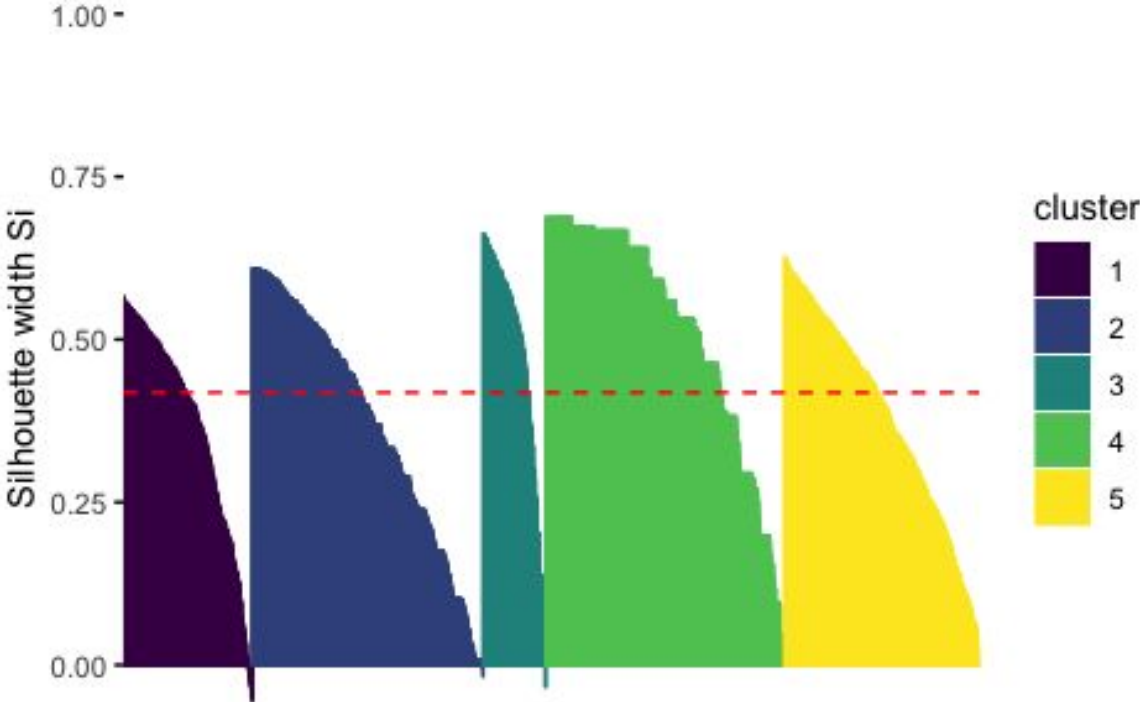
Assigning Journals to a Cluster

We re-ran the clustering algorithm, and assigned each journal to one of 5 clusters. The cluster sizes worked out as follows:

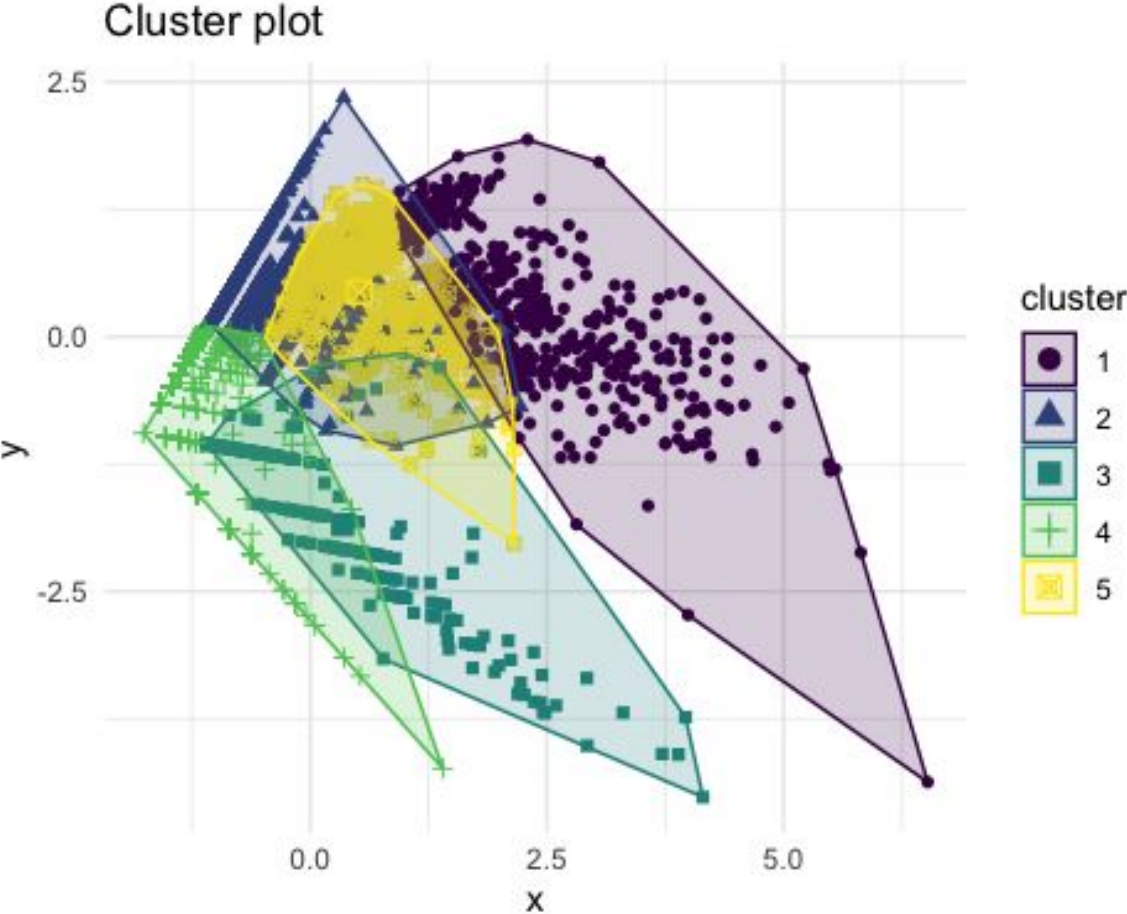
Cluster	Number of Titles
1	437
2	782
3	212
4	810
5	661

Silhouette Width Evaluation

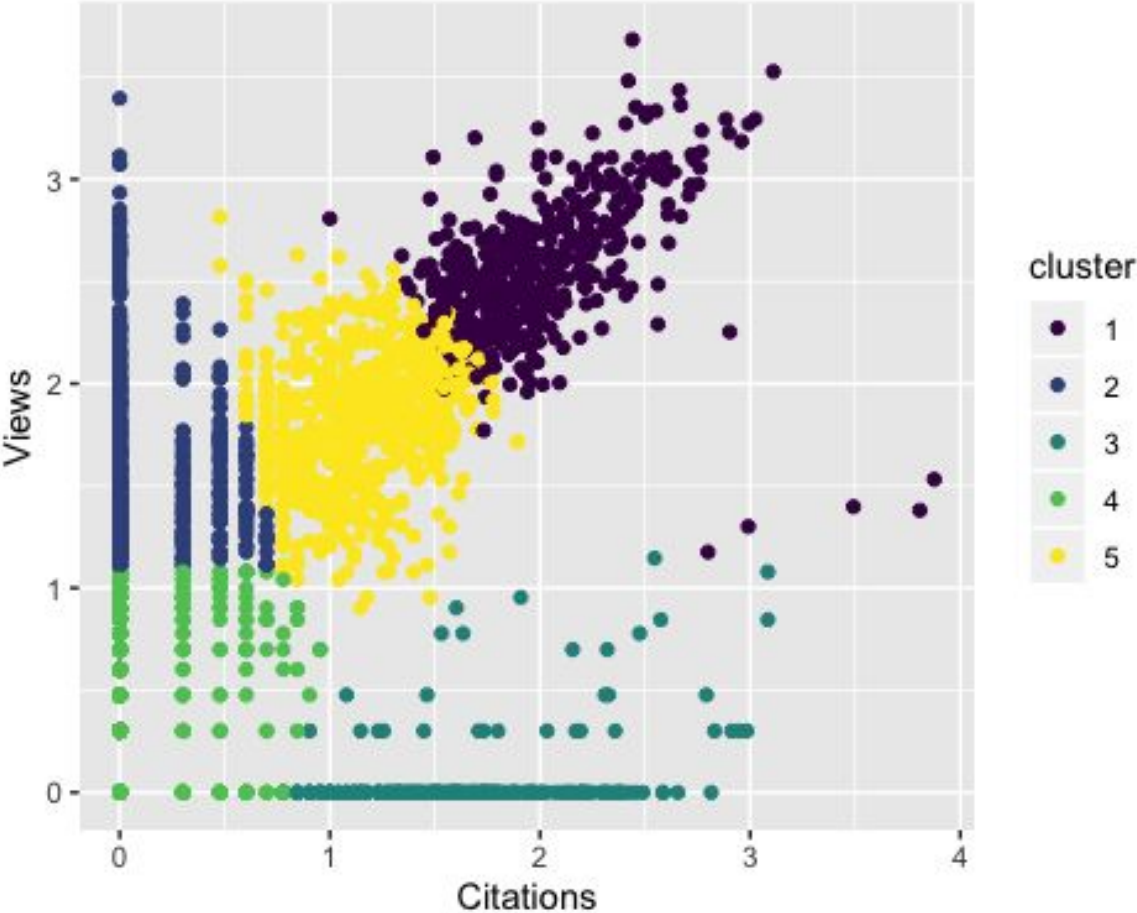
Clusters silhouette plot
Average silhouette width: 0.42



Visualizing the Clusters



Cluster View of Citations



Citations vs. views, by cluster

Additional Plots

- ▷ [Journal Analysis project output testing](#)



Thank you

Please feel free to get in touch if you have questions

ahenry@ualberta.ca