

Article-level metrics: a comparison between publishers

Colin Batchelor
Data Science


Bologna, September 2018

The logo of the Royal Society of Chemistry, featuring a stylized circular emblem with various colored segments (teal, yellow, grey) and the text "ROYAL SOCIETY OF CHEMISTRY" in white capital letters.

ROYAL SOCIETY
OF CHEMISTRY

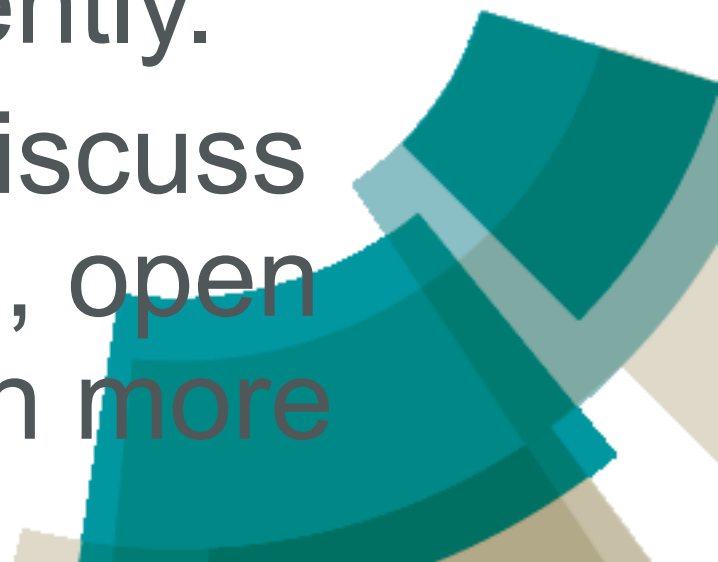


Overview

- Cambridge Metrics group
 - Article-level metrics
 - How do they correlate?
 - How do they relate to Snowball metrics?
- 



Cambridge Metrics group

- First meeting January 2017
 - Initially RSC, PLoS, eLife, CUP and EBI
 - Company of Biologists, Royal Society and IoP have joined subsequently.
 - Regular meetings where we discuss metrics, UX, machine learning, open source, usage and much much more
- 



Company of Biologists

Royal Society of Chemistry

Public Library of Science


eLife

Cambridge University Press





Article-level metrics by publisher

- Number of citations: **all**
 - Number of downloads: **some**
 - Age of article: **all**
 - Mendeley reads by status and country: **all**
 - AltMetric scores by medium: **all**
 - Number of authors: **all**
 - Number of affiliations: **some**
 - OA or not?: **all** (eLife/PLoS implicitly)
 - Article type: **all**
 - Journal: **all** (eLife implicitly)
 - Interdisciplinarity: **RSC**
- 



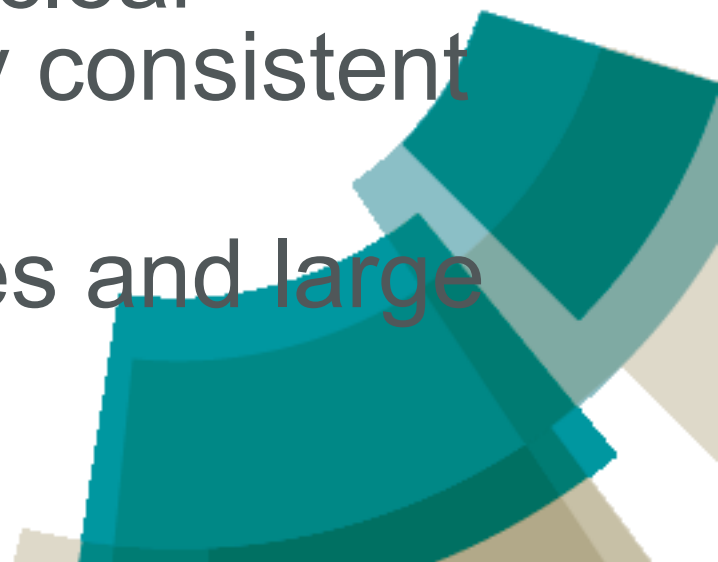
Dataset sizes

Sizes are for articles where all of the metrics were present:

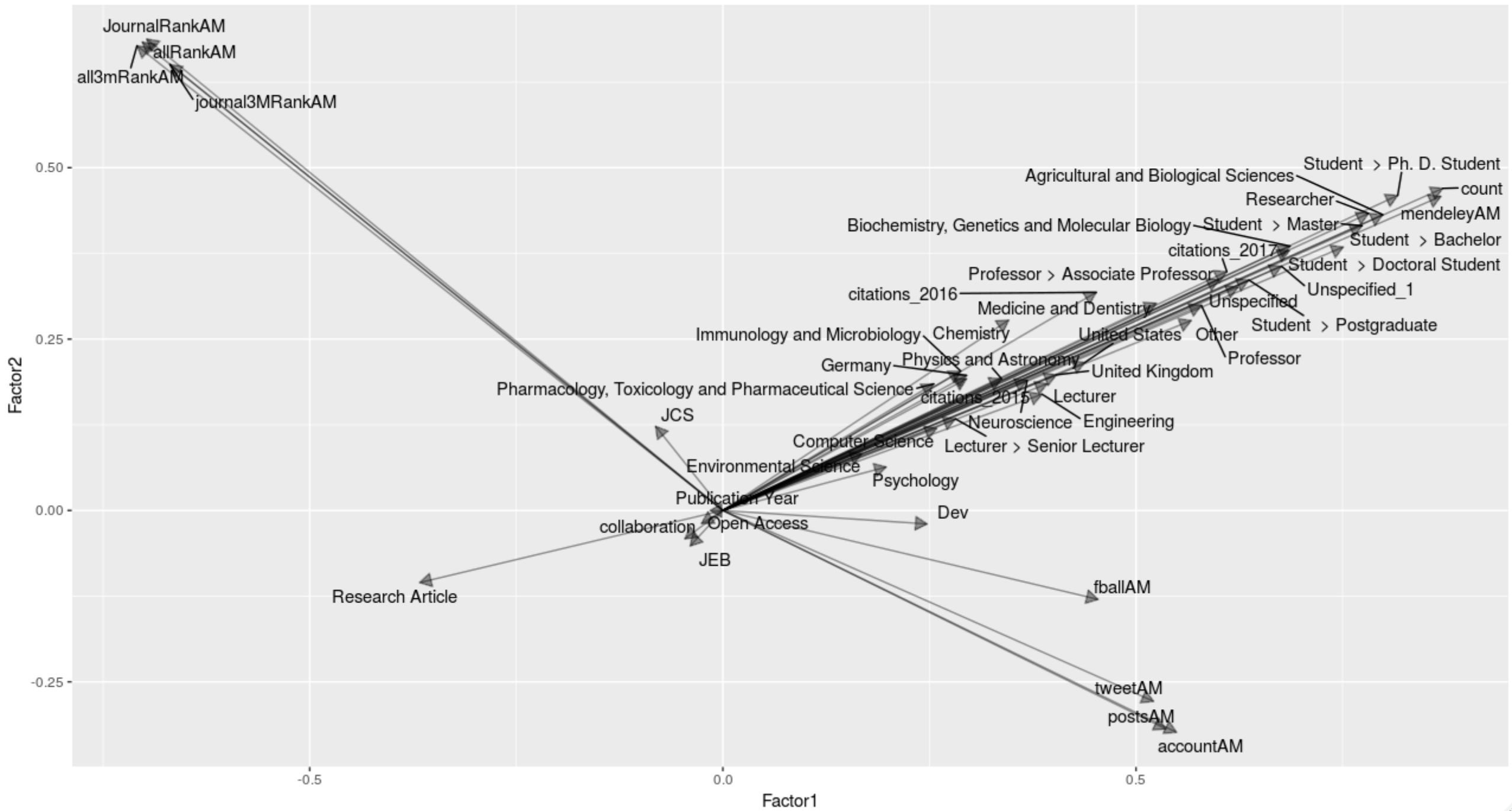
- RSC: 86821 articles
 - PLoS: 9872 articles
 - eLife: 4578 articles
 - CuP: 3661 articles
 - CoB: 2694 articles
- 

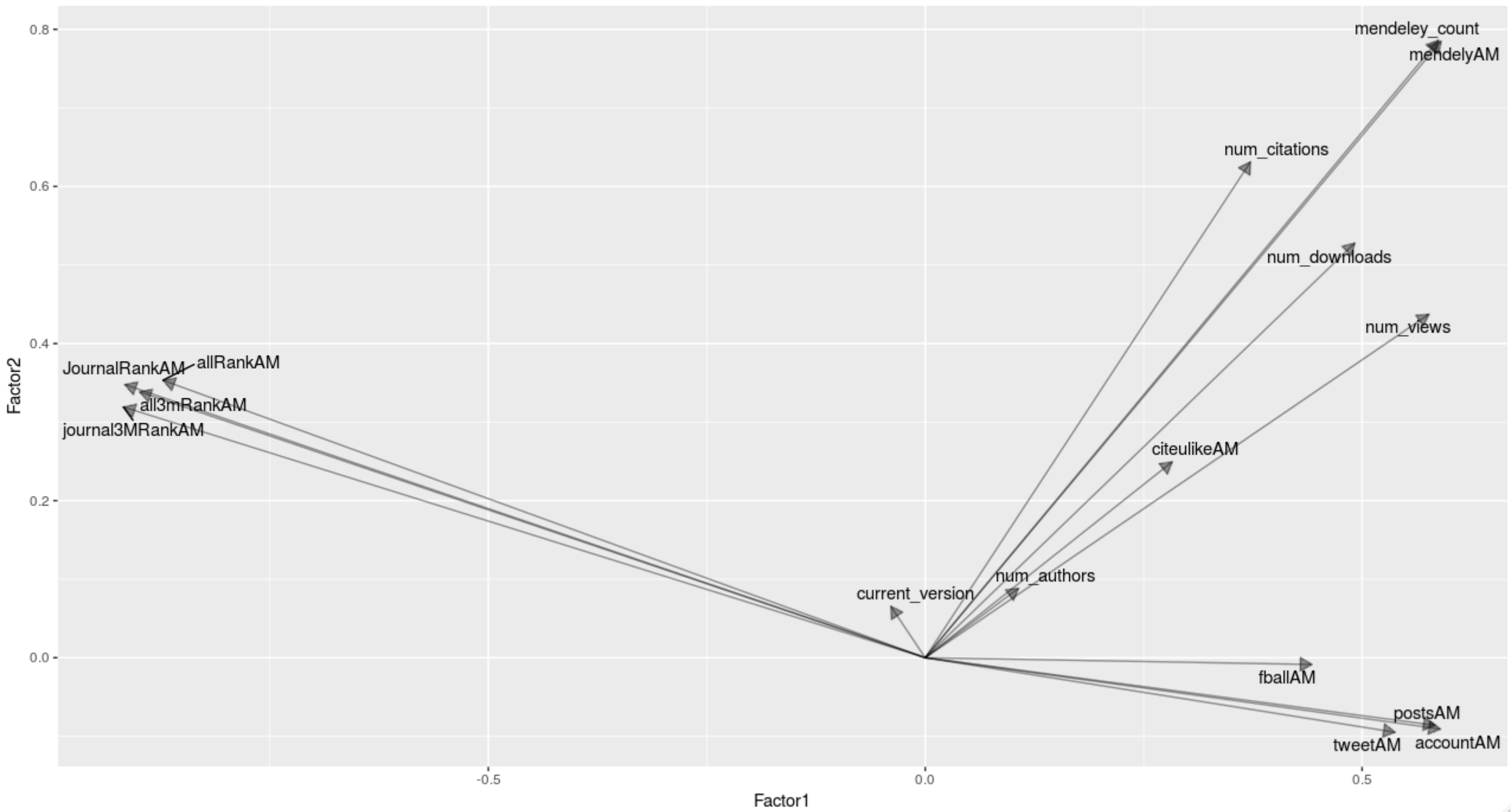


Dimensionality reduction

- We're looking for metrics that measure something
 - We're looking for metrics that are statistically distinct
 - Exploratory factor analysis showed clear groupings of ALMs that were largely consistent between publishers.
 - **However:** fits showed small p -values and large χ^2 -values; bad in this context.
- 


Company of Biologists



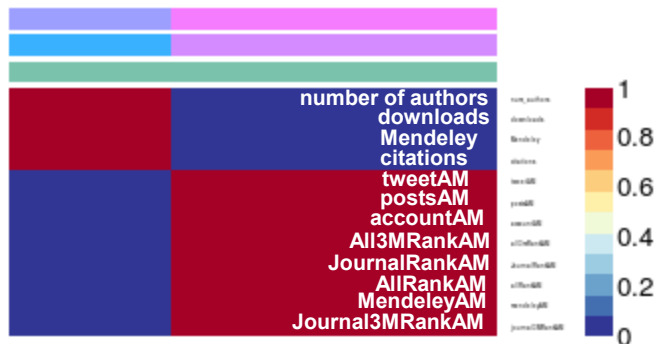




Non-negative matrix factorisation

- Another clustering method!
 - Popular in genetic analyses to obtain “metagenes”, image processing to obtain “basis images”.
 - R package (NMF) has (largely) excellent visualisation tools.
- 

rank = 2



basis

- 1
- 2

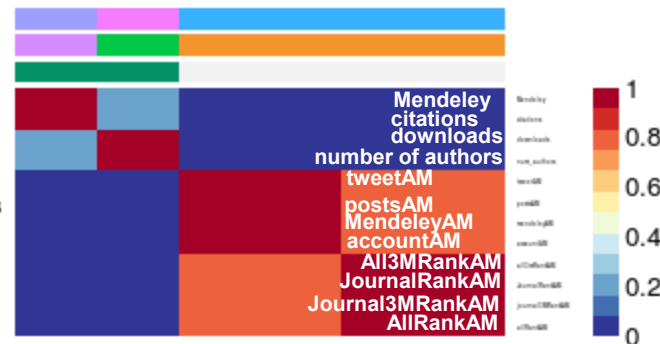
consensus

- 1
- 2

silhouette

- 1
- 0

rank = 3



basis

- 1
- 2
- 3

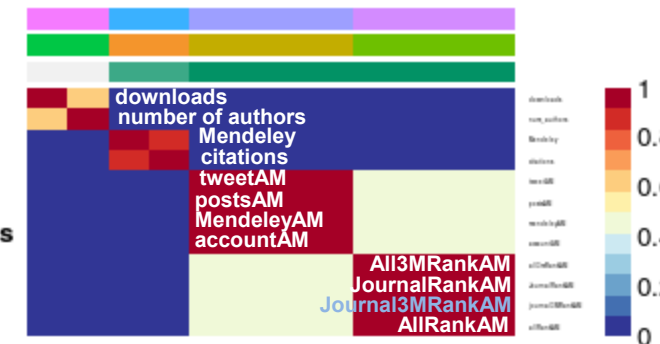
consensus

- 1
- 2
- 3

silhouette

- 1
- 0.89

rank = 4



basis

- 1
- 2
- 3
- 4

consensus

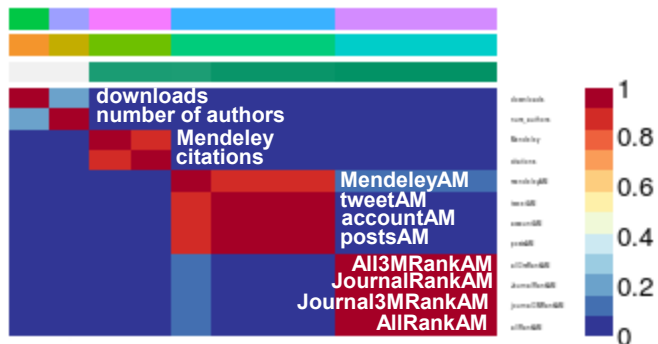
- 1
- 2
- 3
- 4

silhouette

- 1
- 0.6

Cambridge University Press

rank = 5



basis

- 1
- 2
- 3
- 4
- 5

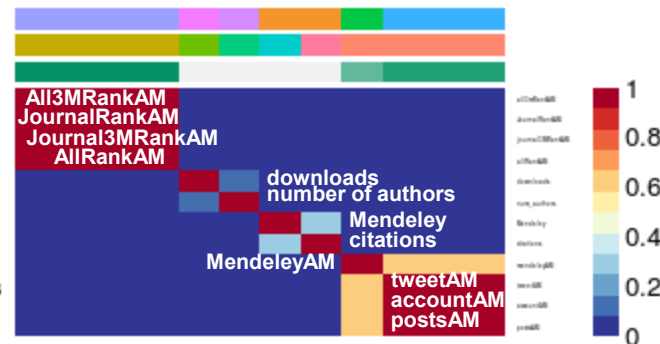
consensus

- 1
- 2
- 3
- 4
- 5

silhouette

- 1
- 0

rank = 6



basis

- 1
- 2
- 3
- 4
- 5
- 6

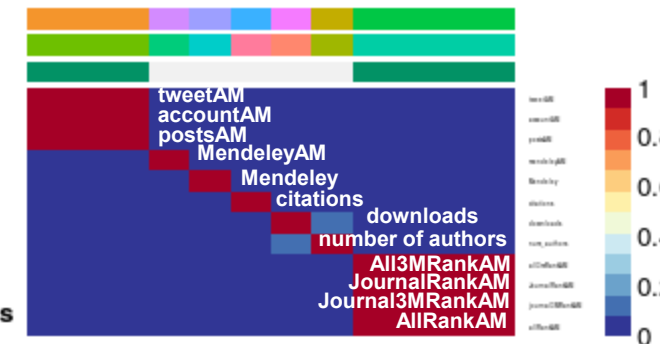
consensus

- 1
- 2
- 3
- 4
- 5
- 6

silhouette

- 1
- 0

rank = 7



basis

- 1
- 2
- 3
- 4
- 5
- 6
- 7

consensus

- 1
- 2
- 3
- 4
- 5
- 6
- 7

silhouette

- 1





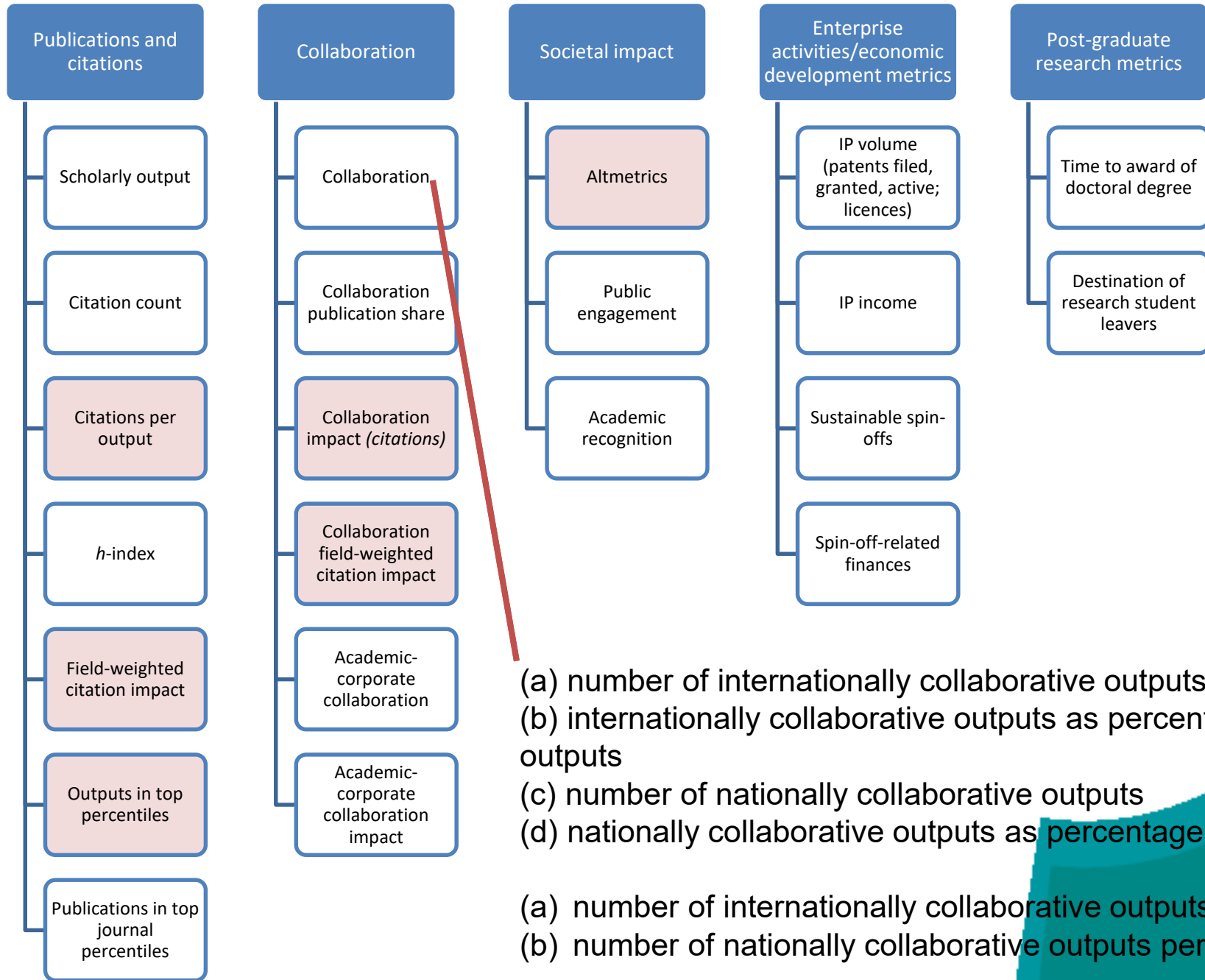
Snowball Metrics

- Intended for internal use within institutions.
- Divided into Input Metrics (grant applications), Process Metrics (income) and Output Metrics (scholarly output and metrics thereof).

<https://www.snowballmetrics.com/>




Snowball metrics





Conclusions

- The Cambridge Metrics group shows publishers working together on problems of general interest.
 - Some ALMs group together:
 - Scientific interest (downloads, collaboration)
 - Scientific relevance (Mendeley, citations)
 - Journal ranking (AM)
 - Social interest (AM)
 - Interdisciplinarity is hard to track.
- 



Thanks

Company of Biologists, Cambridge
University Press, eLife, Public Library of
Science, Data Science at the Royal
Society of Chemistry and the organisers

