a bibliometric study on the literature of Open Science and Open Access

Giannis Tsakonas[1] . Sergios Lenis[2] . Moses Boudourides[3,4]

[1] Library & Information Center, University of Patras, Greece
[2] University of Patras, Greece
[3] School of Professional Studies, Northwestern University, USA
[4] The Science of Networks in Communities (SONIC) Lab, Northwestern University, USA

Input

- 2,846 publications / March 2019
- Data retrieved from / structured data / OA versioning scheme
- Timespan: 1999-2018

- Basic query: `TITLE: ("Open Science") OR TITLE: ("Open Access")`
- Refinements: `NOT TOPIC: (endoscop*) NOT TOPIC: (fish*) NOT TOPIC: (enteroscop*) NOT TOPIC: (schedul*) ...`

Processing

- Processing in Apache Zeppelin, a Python-enabled platform for the deployment of multi-purpose research notebooks.
- Notebooks are hybrid interactive environments, designed for computational tasks, including ingestion, analysis and visualization of data.

- Three notebooks available on Zepl, a platform for data exploration:
- **Statistics**, general statistics overview
- **Timelines**, progress through time
- **Networks**, connections of authors, sources and access types

part 1: statistics

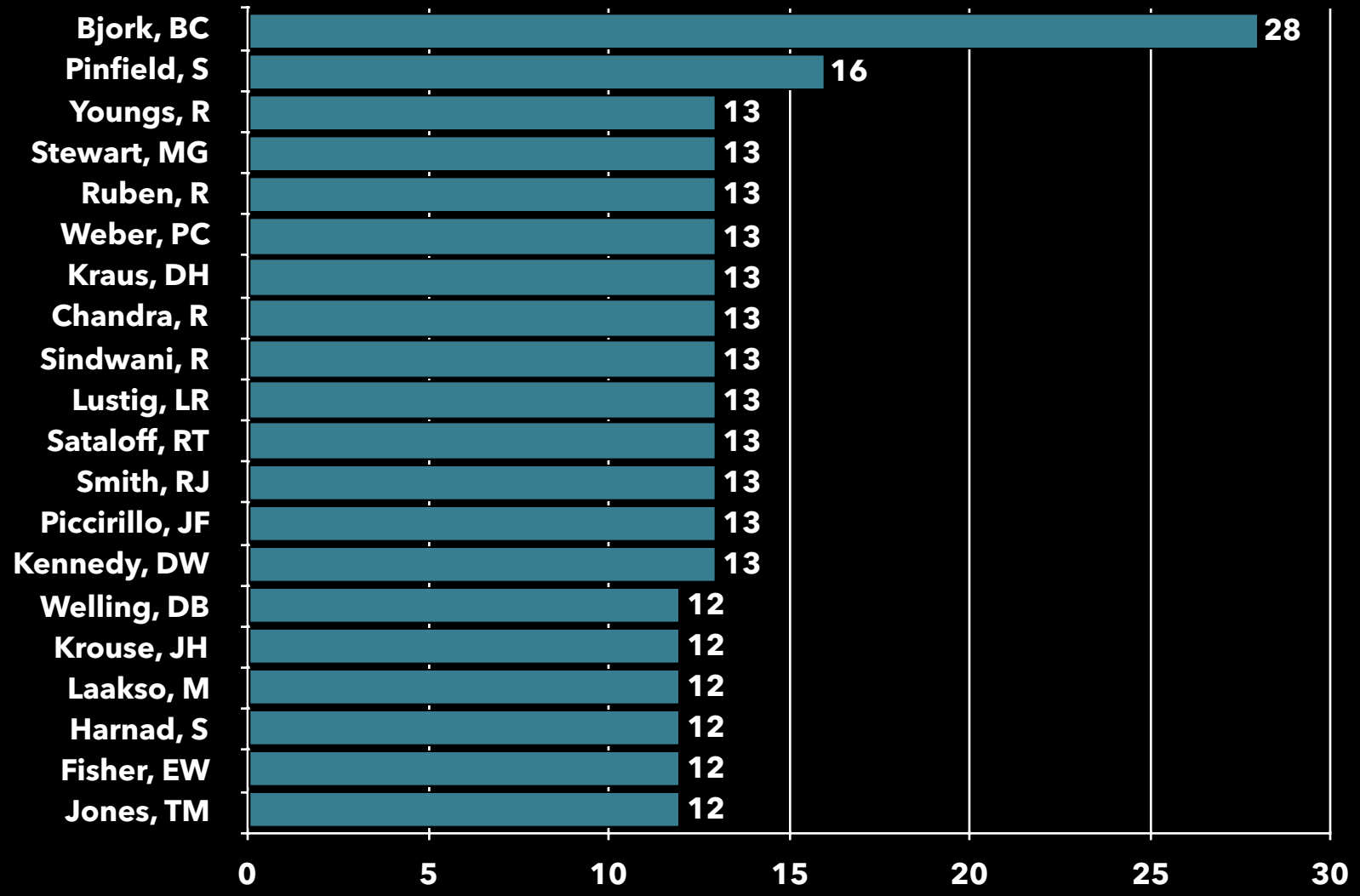
authors . journals . productivity . access type . funding .
document types . identifiers & descriptors .
languages . times cited

zepl notebook https://bit.ly/2L9TGW3

statistics: authors

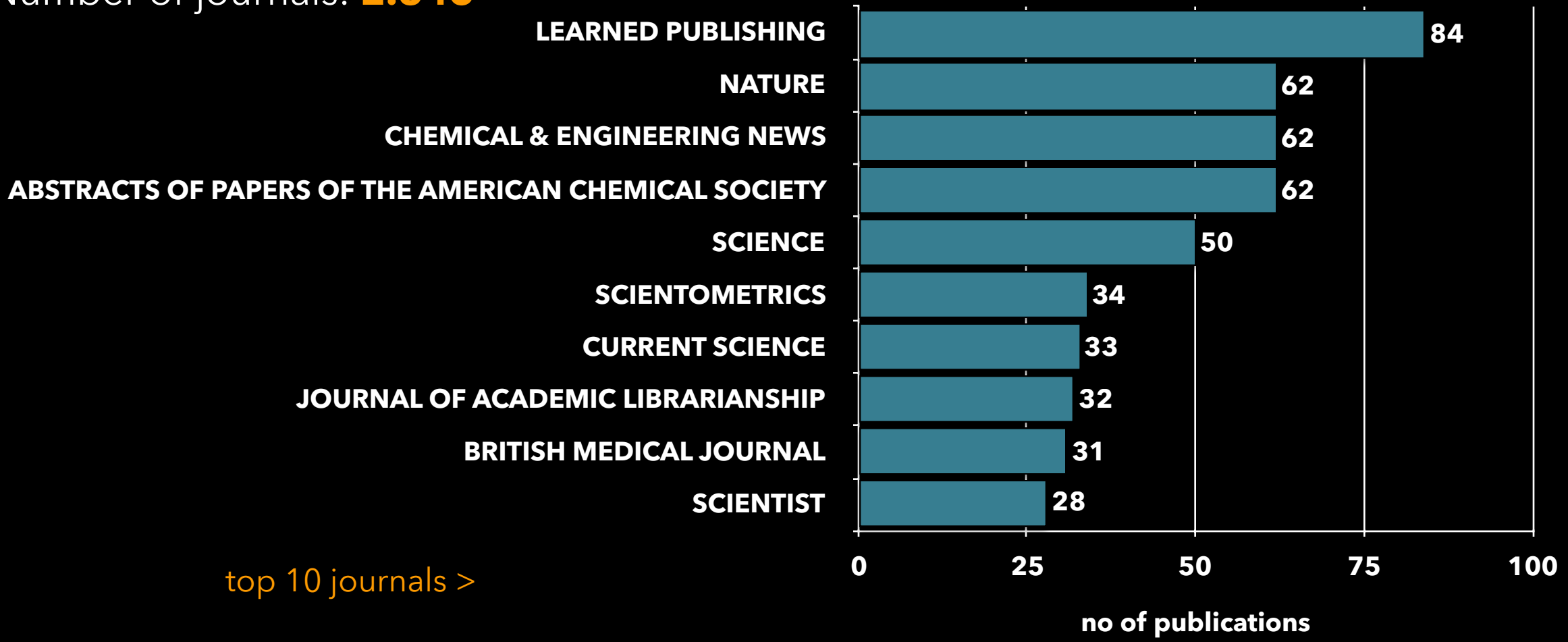
Unique authors: **5.659**
Authoring teams: **2.362**

top 20 authors >



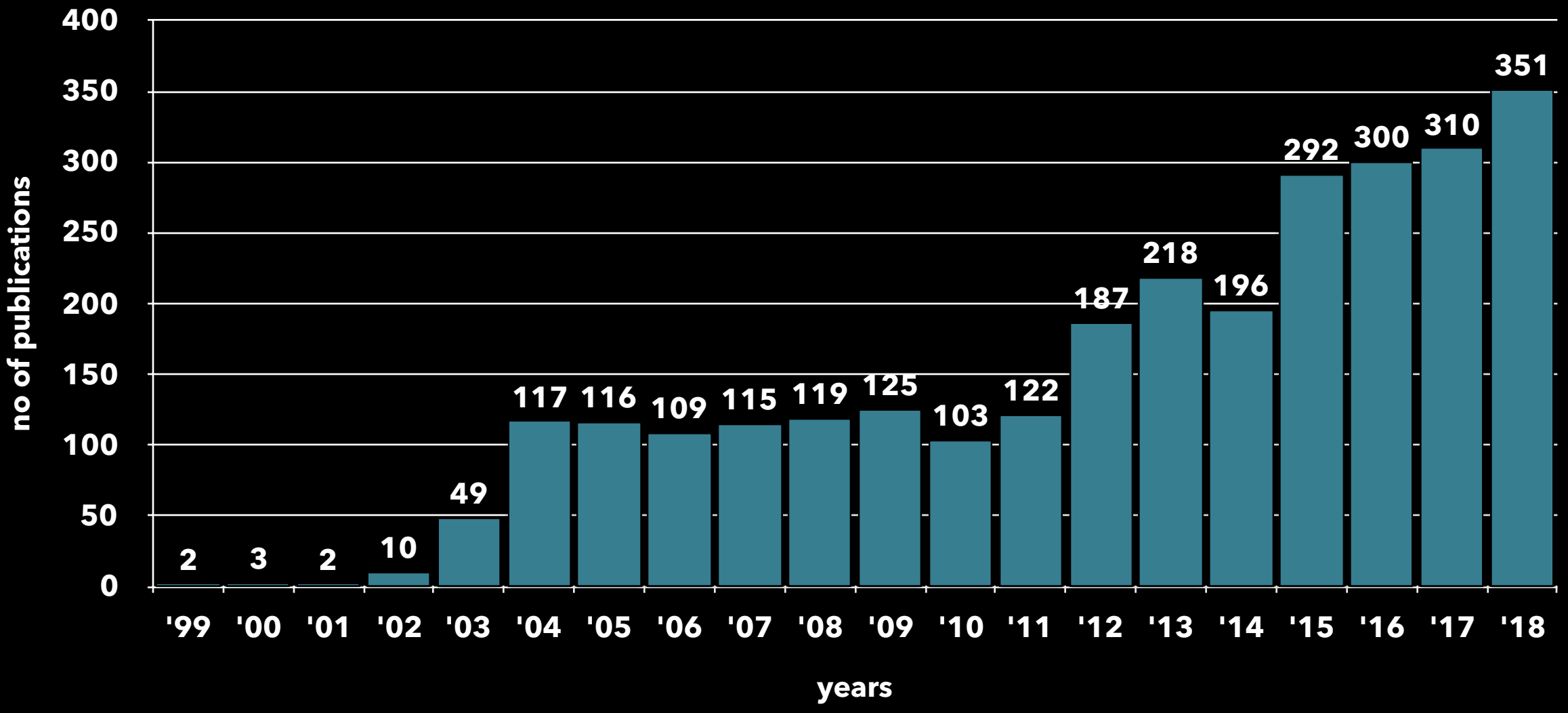
statistics: sources

Number of journals: **2.846**

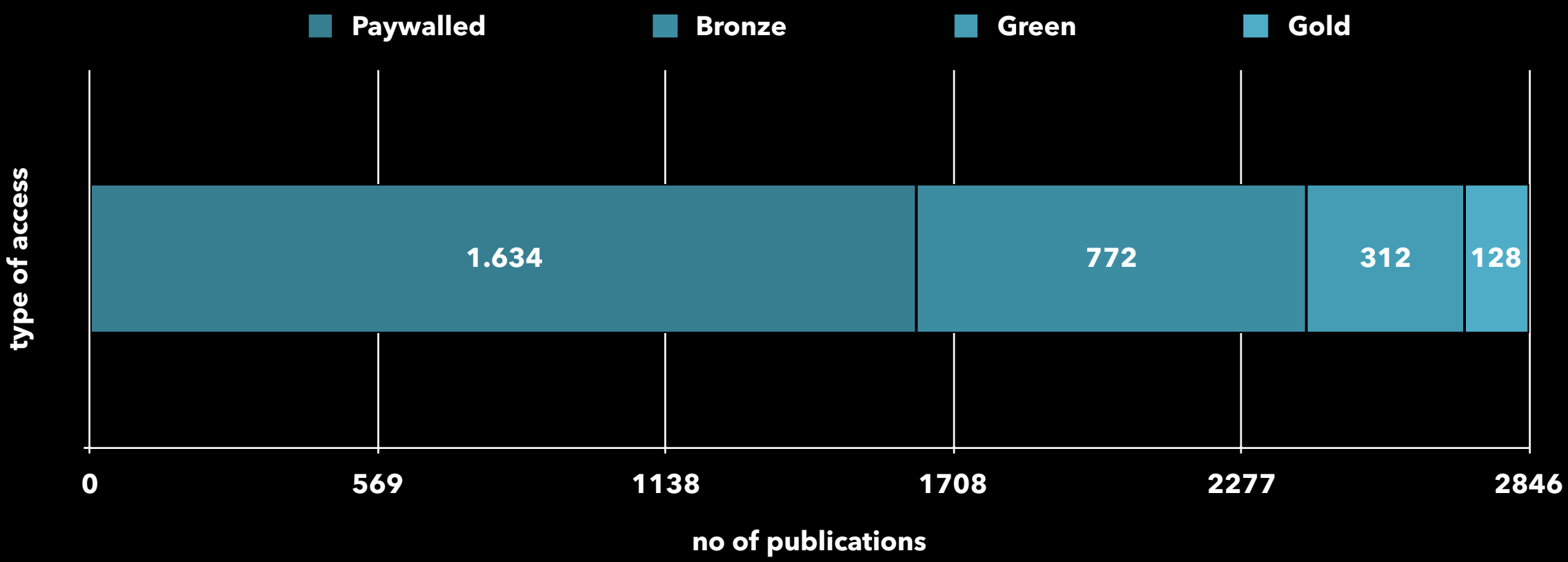


top 10 journals >

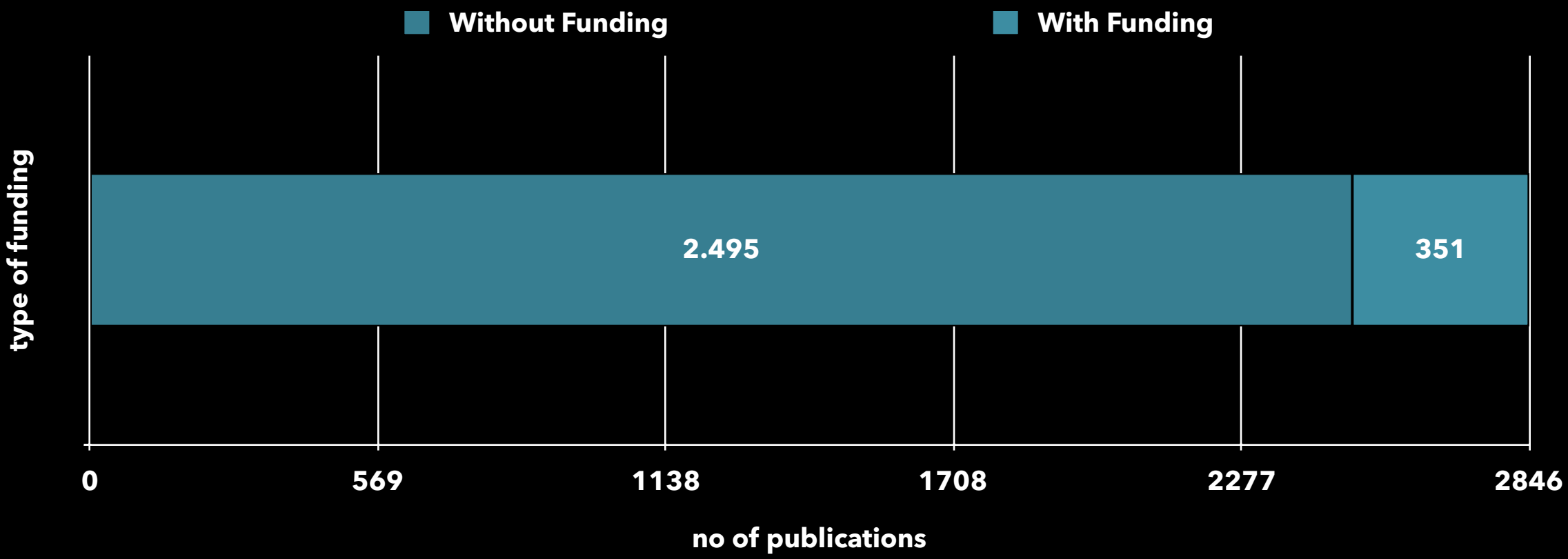
statistics: productivity



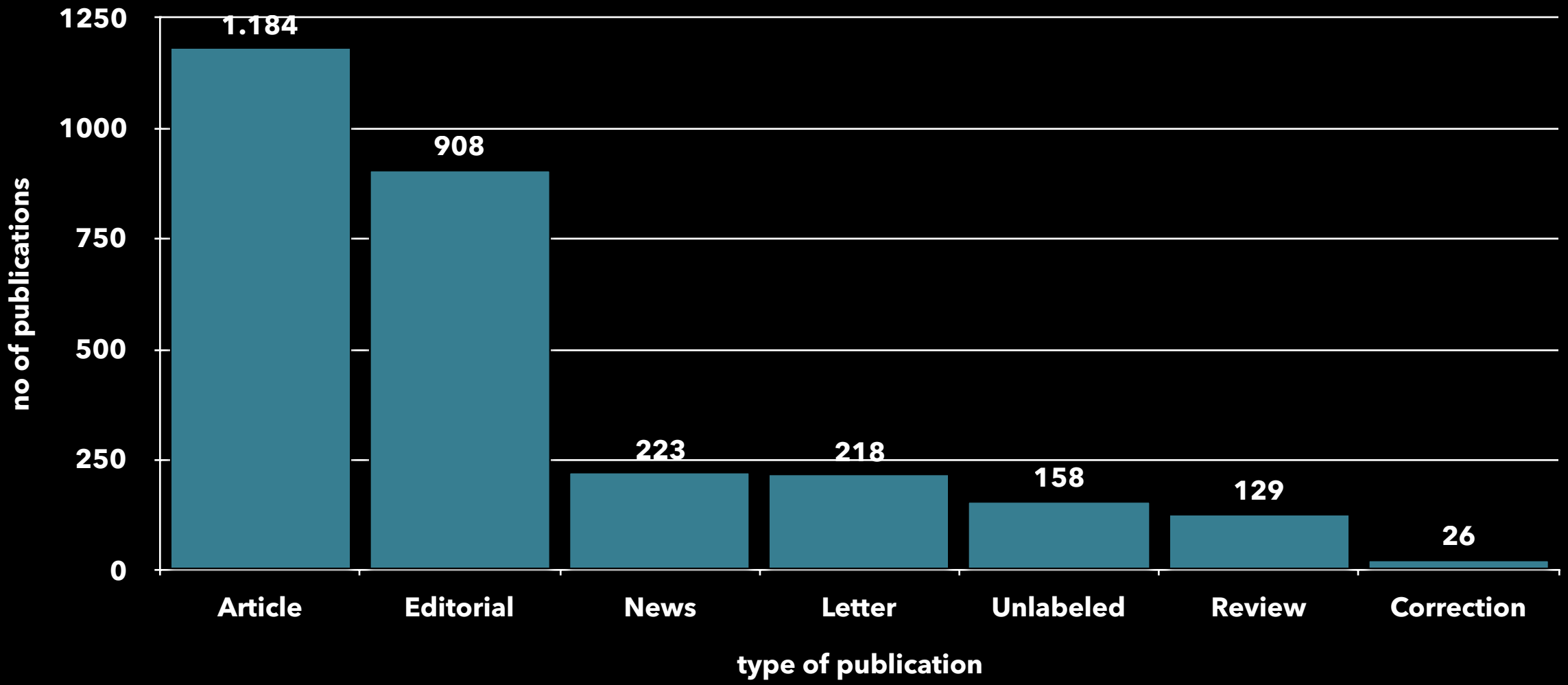
statistics: access type



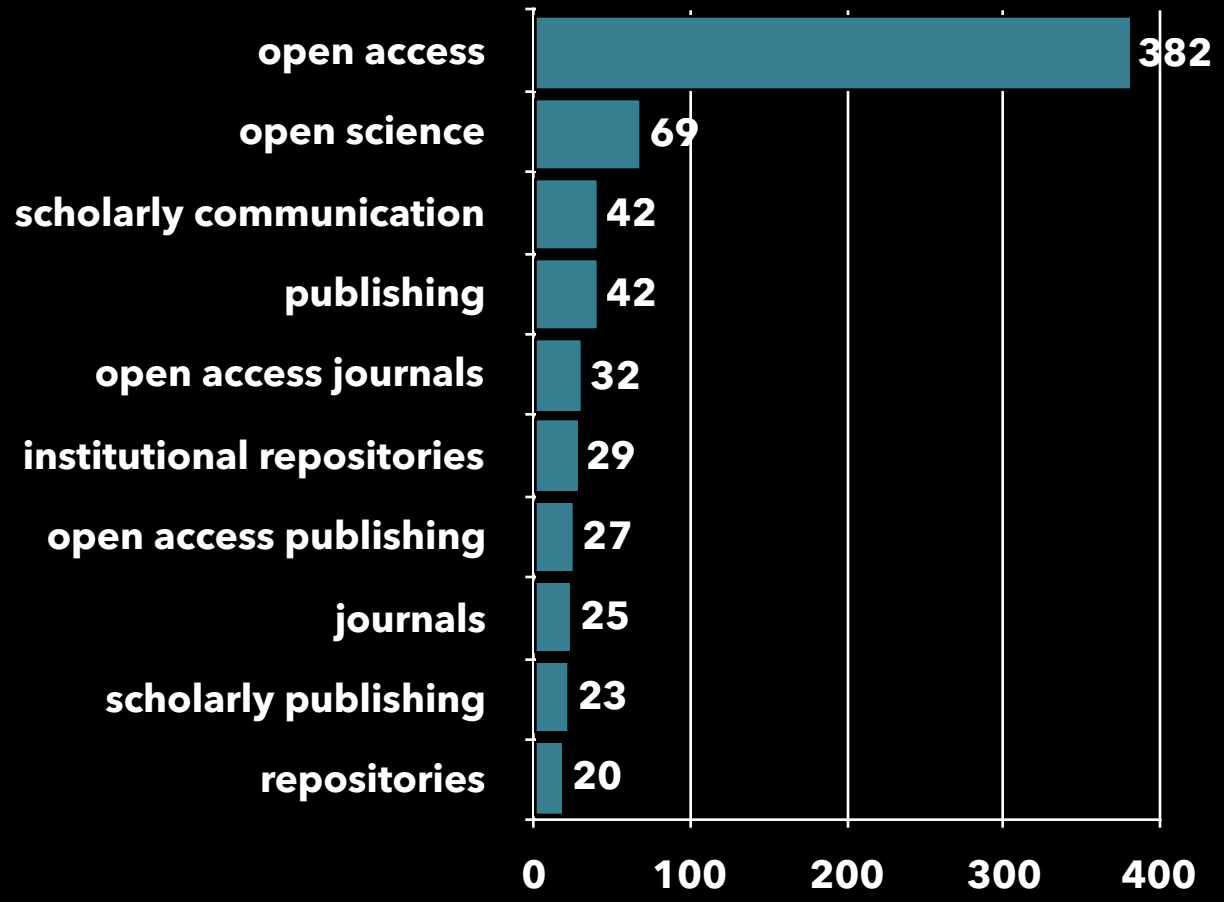
statistics: funding

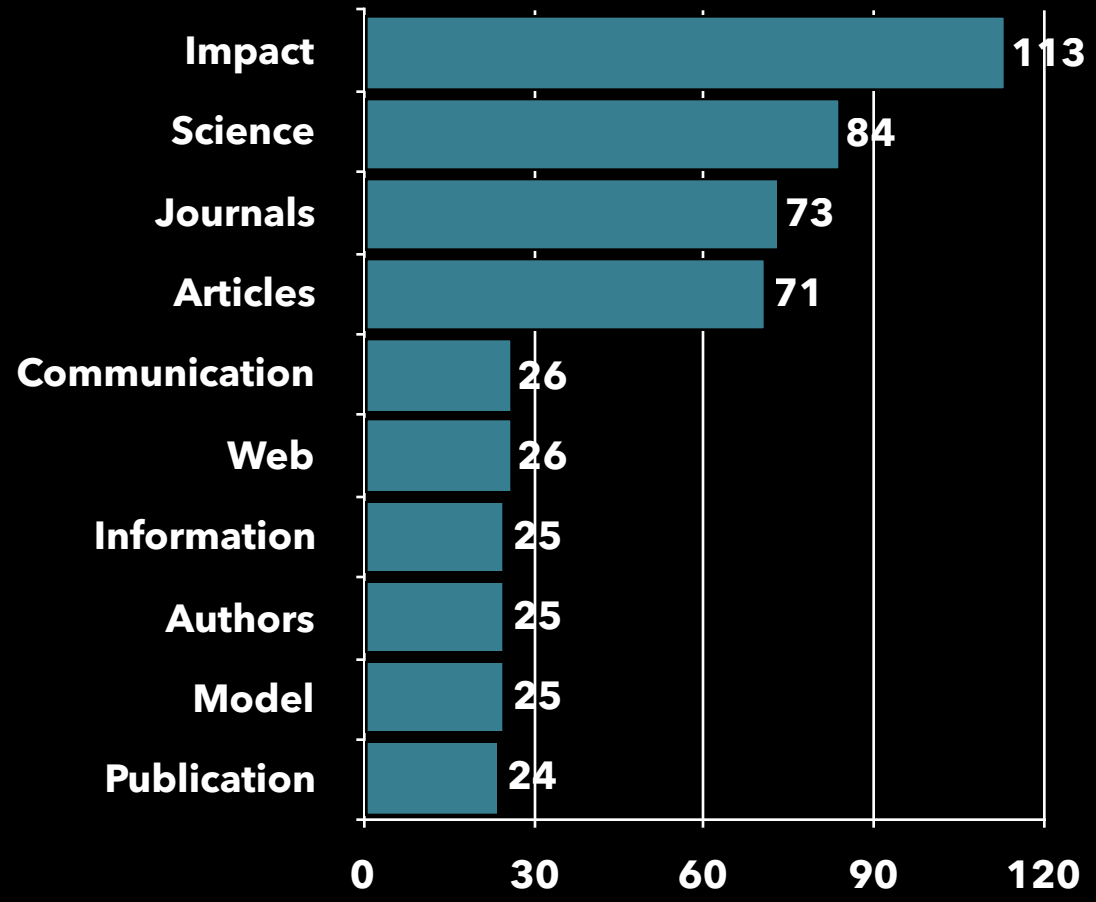


statistics: document type



statistics: descriptors & identifiers



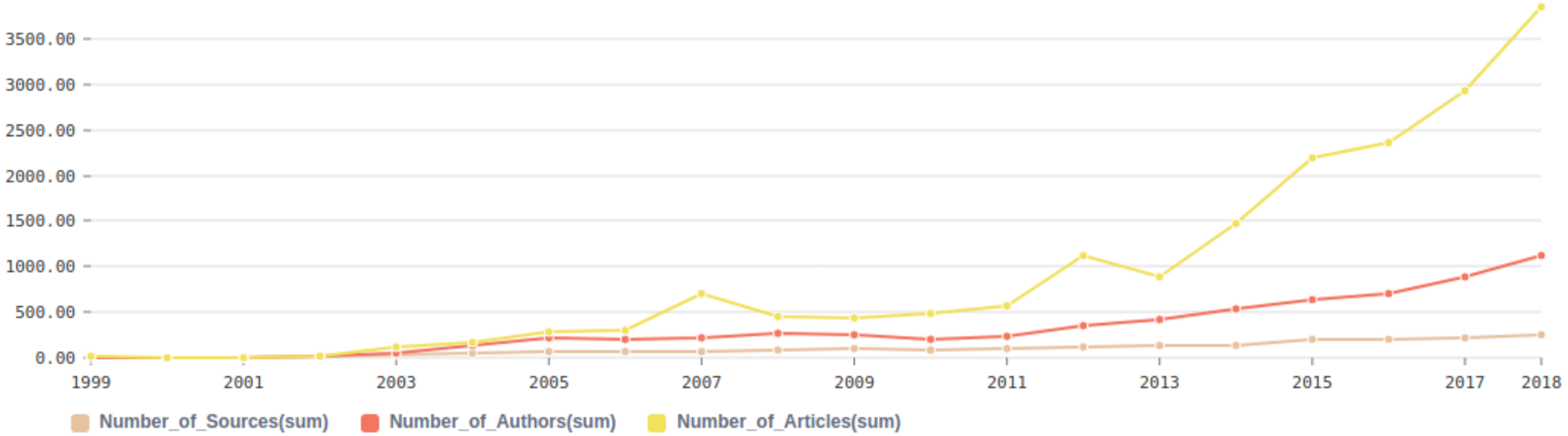


part 2: timelines

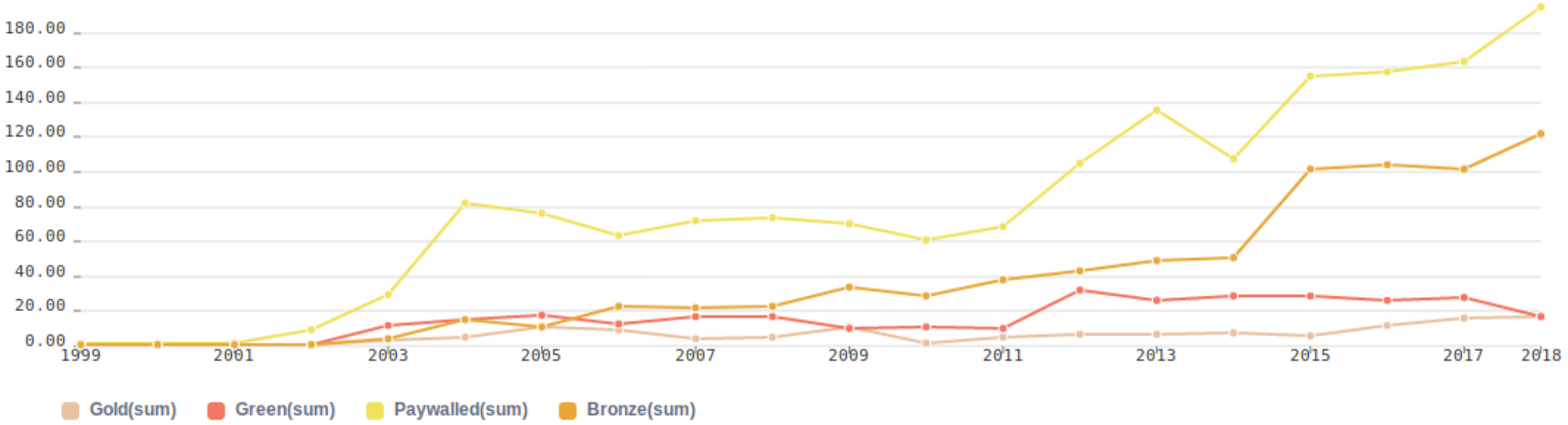
timelines of publications, authors & sources / access type / funding and document type . **life cycles of authors per access type . expansion rates of authors . time variations of authors & sources .** rates of continuity or lapse for authors & sources

zepl notebook https://bit.ly/31VON8I

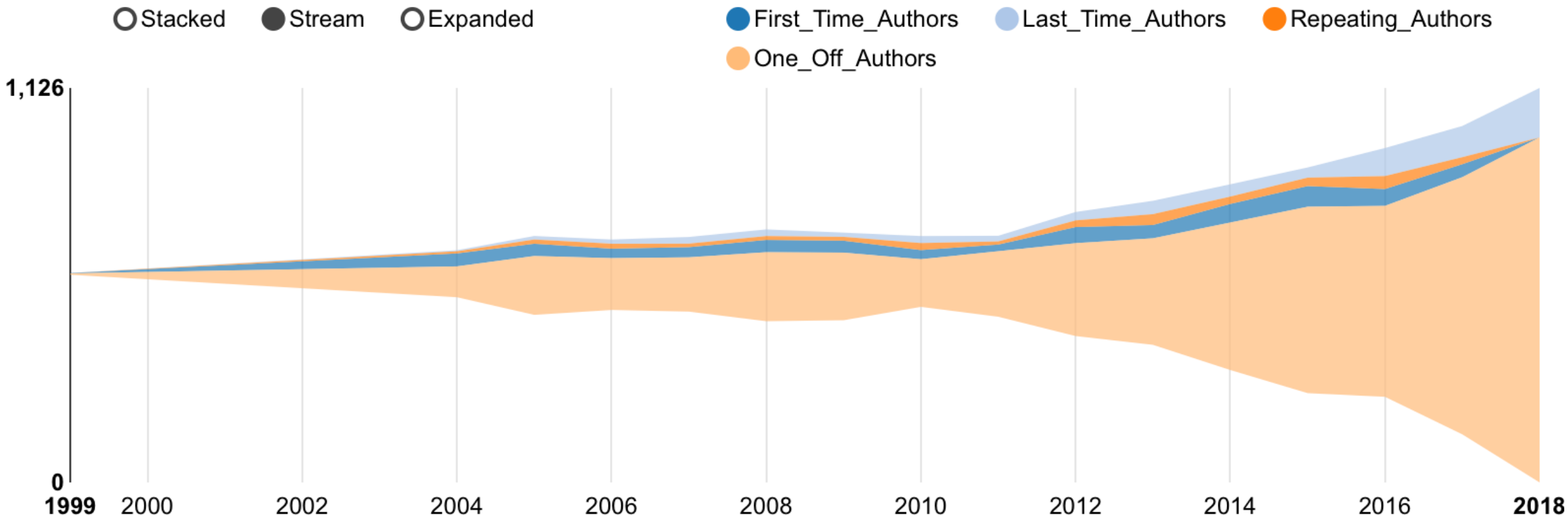
timelines: publications, authors & sources



timelines: access type

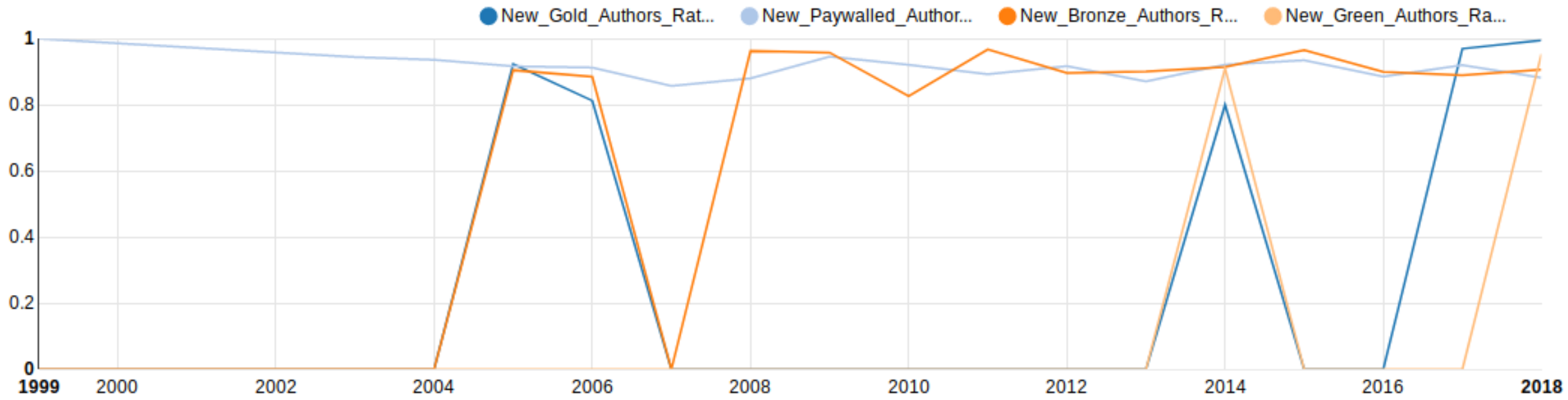


lifecycles: authors



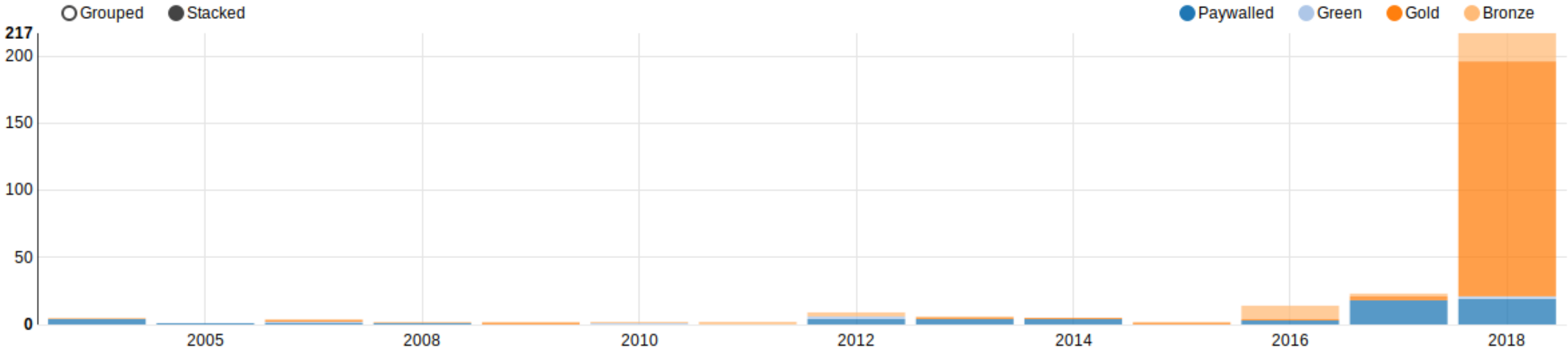
Excluding 2018, the majority of the authors in our sample publish only once and then disappear (83%, 2017). First time authors have a highest at (27.6%, 2004), where as 2016 was a "draining" year (11.3%).

expansion rates: authors



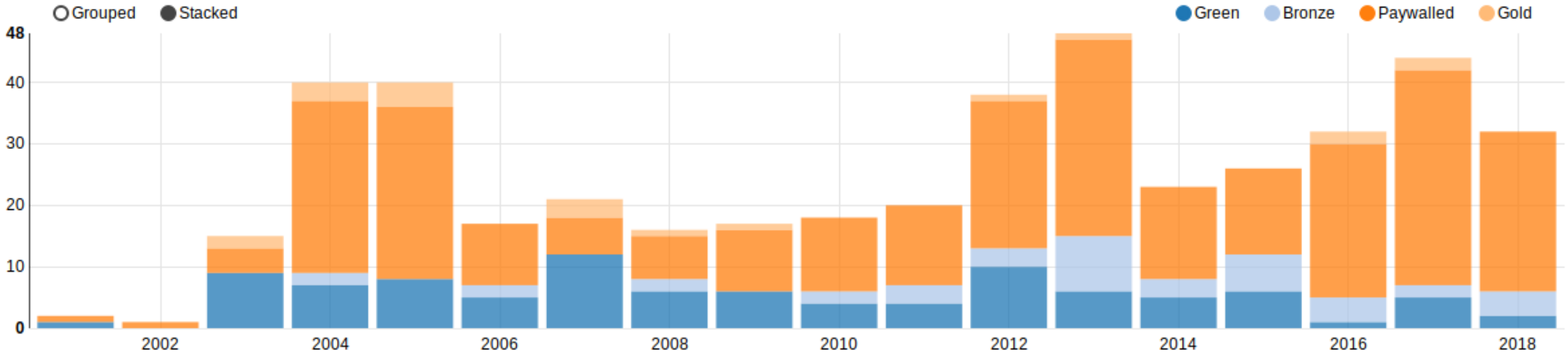
The ratio of First Time Authors and the One-Off Authors in a year to all existing authors in a year.

time variations: top authors



Access type of the publications of 22 authors with at least 10 publications each one. Paywalled and Bronze are consistent, while Gold and Bronze are preferred in the "burst year" of 2018.

time variations: top sources



Access type of publications in nine sources with at least 30 publications each one. Deeper views on Document Type show that Editorials and News Item play significant role.

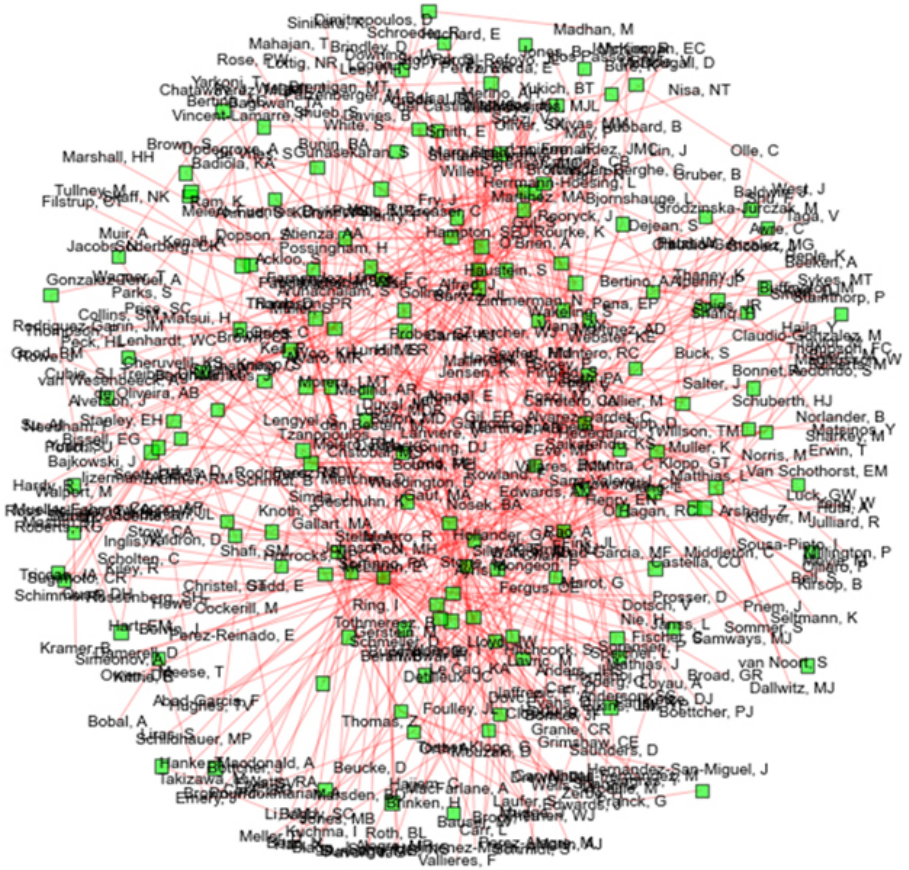
part 3: networks

bipartite graph of authors-publications . co-authorship graph . **bipartite graph of authors-access type** . access assortativity of the co-authorship graph

zepl notebook https://bit.ly/31VON8I

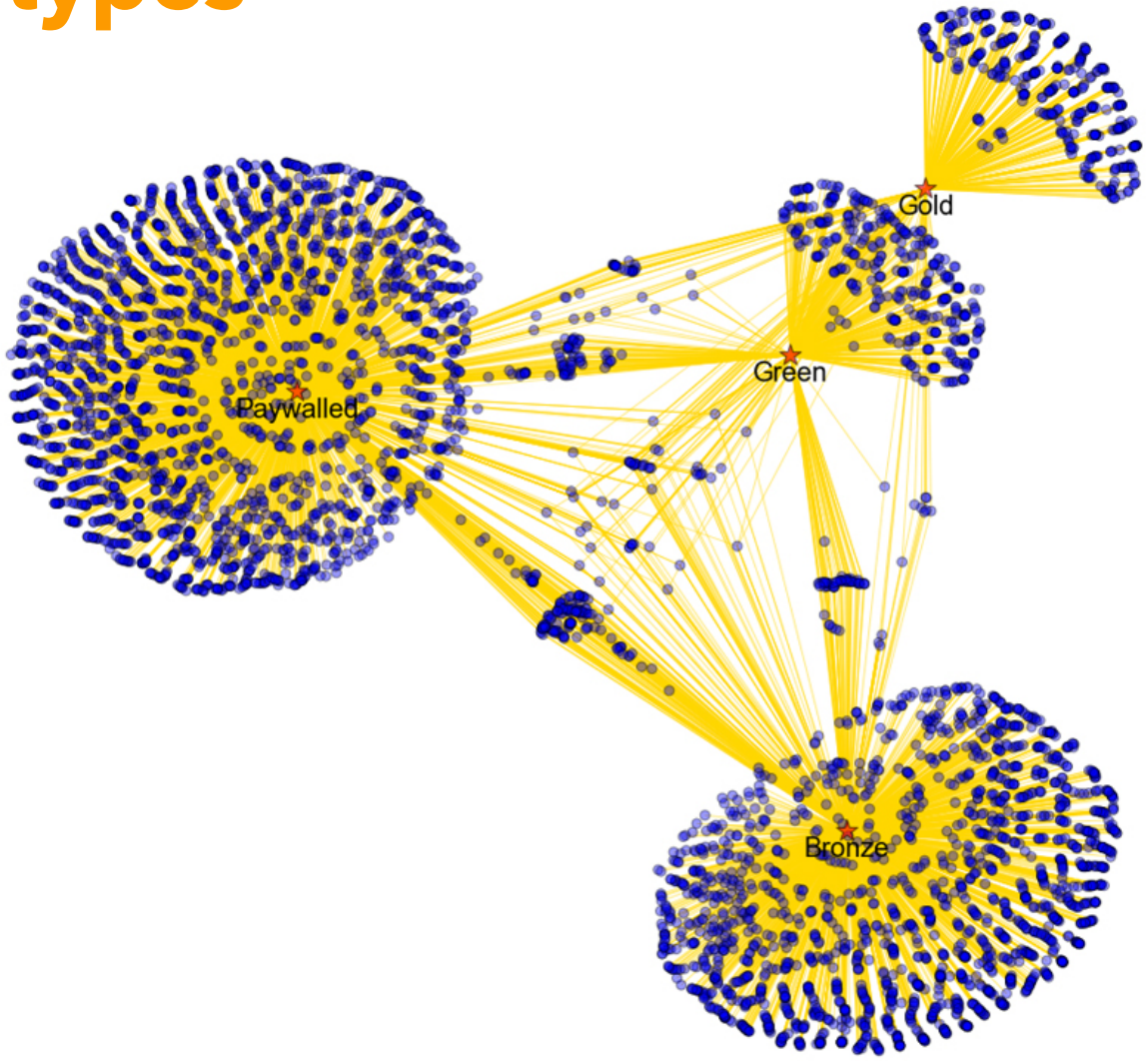
networks: authors & publications

- The bipartite graph of publications and authors has 8.505 nodes (5.659 authors and 2.846 publications) and 7.248 edges (authorships).
- The largest connected component has 590 nodes (447 authors and 143 publications) and 732 edges (authorships).



networks: authors & access types

- The bipartite graph of authors and access type has 5.663 nodes (5.659 authors and four access types) and 6.666 edges.
- Both Paywalled and Bronze group into large clusters of authors.
- A third cluster of authors is shared between Green and Gold.
- Gold OA authors have the least ties with other access types.



Some conclusions

- "Open Access" is semantically ambiguous term / similar terms in medical informatics, electrical engineering, rural planning and land management, drug policies, fishery, etc. / manual cleaning is needed, but not perfect.
- We observed similar growth trends like in other fields / steady growth after 2004 > upscaling after 2012.
- We also observed that the field is expanding / there are many occasional authors / few are the consistently publishing ones / fresh authors represent approx. 10% each year.
- Bronze publications appear as an opportunity to work further on training.

thank you for your attention

addendum at DOI 10.17605/osf.io/u7azn