Metrics and Visualizations in the Astrophysics Data System

Edwin A. Henneken (ehenneken@cfa.harvard.edu)
Smithsonian Astrophysical Observatory, Cambridge MA, USA

The SAO/NASA Astrophysics Data System (ADS) is the digital library portal used on a daily basis by all professional astronomers worldwide to stay up-to-date. The ADS also maintains a citation database which, for the discipline of astronomy, is complete to a very high degree, because of the inclusion of many conference proceedings, smaller journals and grey literature. This makes the ADS an environment that is very suitable for calculating accurate metrics, both citation and usage based. The ADS has developed both metrics and visualization services, to augment the discovery process. These support both calculation of metrics and visualizations for individual researchers, but also for programs and instruments (see figures 1 and 2). All visualizations are based on the JavaScript d3.js library. Users can further refine their query results by selecting terms or entities from the visualizations and use them to further refine their list of results.

Why did we spend significant resources developing these services? As far as the metrics service is concerned, there are various reasons. We, as ADS, have been involved in bibliometric research since the early days of our service, so we would greatly benefit from having a formal metrics service. In addition, we have had many requests from our users (researchers, project managers, librarians and publishers) for a service that would provide statistics for a set of publications. Researchers want to provide citation statistics with a job application or just check how well they are doing compared to their competitors. Managers of e.g. a fellows program sometimes use publication metrics as one of the criteria in selecting candidates. Projects managers and librarians use metrics for the quantification of scientific output and impact of a program or instrument (Winkelman 2012). The actual information generated by the metrics service is based on user requests and trends in the information science community. For example, the moment the Hirsch factor was introduced, users started to ask how they could (easily) generate the h-index using the ADS. Therefore, in addition to basic publication statistics, we supply an overview of indicators: *h-index* (Hirsch 2005), *g-index* (Egghe 2006), *iN-index* (the number of publications with at least N citations, N=10, 100) and the *tori* and *riq* indices (Pepe 2012). In addition to basic statistics and indicators, we also provide histograms and a time series diagram, to get a sense of distribution and evolution. In order to show the contribution of just the refereed publications, we also provide basic statistics, indicators and histograms for this class of publications.

Thanks to its powerful query syntax, it is also fairly easy to generate custom bibliometrics data using ADS 2.0 (http://adsabs.org, code name "Bumblebee"). For example, if we want to explore the citation network on the granularity level of journals and determine which refereed journals cite a particular journal the most for a given year, we could run a query like

```
citations(bibstem:AJ year:2012) year:2012 property:refereed
```

which finds all citations to articles published in The Astronomical Journal in 2012 that are refereed and themselves published in 2012. Running this query in the user interface immediately provides interesting results by looking at the facets in the results overview. Queries like these can also be executed in the API of the ADS, allowing convenient capture of results.

In terms of visualizations, the ADS offers an *Author Network* and a *Publication Network*. The author network takes the top 200 most frequently appearing authors within your result set, measures the frequency of collaboration between them, and displays color-coded groups of authors organized around a center point. The paper network creates groups of papers that share a significant number of references, and names those groups by looking for shared, unique words in their titles. Users can also generate a *Concept Cloud*, which is a word cloud based on words in your search results. A slider can be used to change the emphasis of these words: from *unique* to *frequent*. Finally, search results can also be used to generate a *Results Graph*. When you look at a list of ADS search results, you can sort by date published, by citation count, or by recent popularity of the article in ADS, but you cannot see all of these dimensions at once, and you cannot easily see outliers. The results graph is a customizable scatter chart that allows you to assign values to the x and y axes as well as to the radius of the circle representing a paper.

Visualizations provide insights that do not follow readily from looking at lists and numbers. They can provide qualitative metrics. The author network visualization, for an author query, can show how this author's collaborations have shifted and who his/her most prolific co-authors are. The paper network visualization, for an author query, can show how this author's interests have shifted over time and which of his/her interests have had the most impact in terms of citations or usage. In the case of a subject query, the author network visualization gives insight in who the major contributors are to this field and if there are collaborations that are particularly active. The paper network visualization for a subject query illustrates various aspects of this subject; perhaps there is both a strong theoretical and observational component, or maybe there are groups of people who look at different aspects.
The visualizations can also be used as filters. For example, after having done a subject query, you can use the author network visualization to select a group of authors, filter out only those publications with authors from this group and then calculate the metrics overview.

References

- Winkelman, S. and Rots, A. H. (2012). "Observatory Bibliographies: Not Just for Statistics Anymore", in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series 8448, 844829.
- Pepe, A., & Kurtz, M. J. (2012). "A Measure of Total Research Impact Independent of Time and Discipline". PLoS ONE, 7, 46428
- Egghe, L. (2006). "Theory and practise of the g-index". Scientometrics, vol. 69 (1), 131–152
- Hirsch, J. E. (2005). "An index to quantify an individual's scientific research output". PNAS 102 (46): 16569
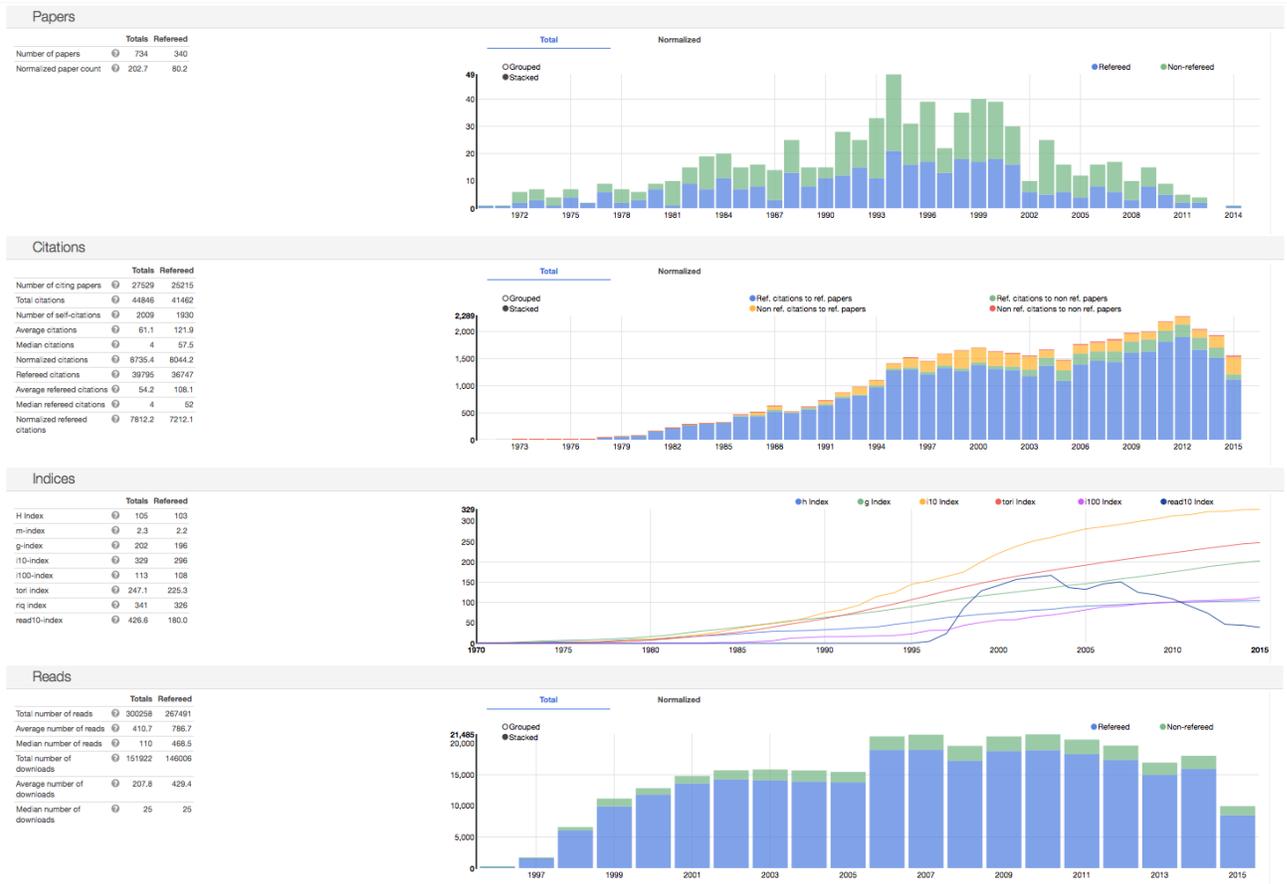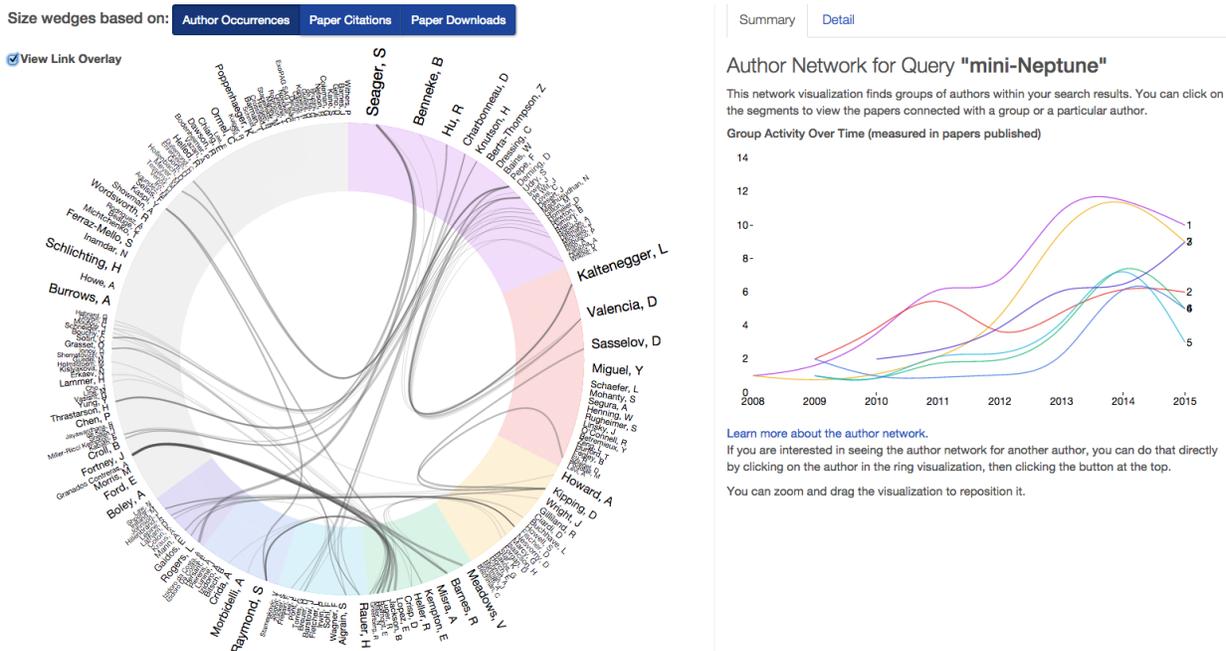
Figure 1. Example of metrics overview

Figure 2. Example of author network