Methods for Bibliometric Institutional Profiles for the Practitioner

Christina K. Pikas, Christina.Pikas@jhuapl.edu

This poster presents a case study of how a librarian can prepare a bibliometric profile of a research organization. The case study updated and revisited an earlier study by Berl (1986) profiling the Johns Hopkins Applied Physics Laboratory (APL). Assuming access to either Scopus from Elsevier or Web of Science (WoS) from Clarivate, the remaining steps can all be completed with tools and knowledge available to the sophisticated practitioner.

BACKGROUND

Profiling multidisciplinary organizations or even geographic regions over time can be done to compile a retrospective or for competitive intelligence, strategic planning, or to assess potential collaboration partners (cf Kostoff et al., 2007). This requires normalization over disciplinary areas as well as over time because the various fields represented within an organization may have dramatically different citing behaviors. The citation source normalization method suggested by Leydesdorff and Opthof (2010) is used for discipline and binning is used for time.

METHODS

Retrieve article metadata from citation database

- Use database organization roll-up feature (*Organizations- Enhanced* field) , and search Berl(1986) articles individually
- Retrieve full record in plain text format for APL's and citing articles

Use R scripts to

- Calculate fractional times cited
- Determine top venues, most prolific authors, and collaborating organizations
- Map collaborating organizations
- Perform k-means longitudinal clustering to view citation trajectories
- Perform Latent Dirichlet Allocation (LDA) to categorize major topics

RESULTS

Table 1: Top cited articles by fractional counting (TC_F)

Citation	TC	↓TC_F
Mirowski,M., Reid,P., Mower,M., et al. (1980) Termination of Malignant Ventricular Arrhythmias with an Implanted Automatic Defibrillator in Human-Beings. *New England Journal of Medicine 303*,322-324. doi:10.1056/NEJM198008073030607	850	42.43
Kanungo,T., Mount,D., Netanyahu,N., Piatko,C., Silverman,R. and Wu,A. (2002). An Efficient k-Means Clustering Algorithm: Analysis and Implementation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 24*, 881-892. doi:10.1109/TPAMI.2002.1017616	789	40.56
Spall,J.(1992). Multivariate Stochastic-Approximation using a Simultaneous Perturbation Gradient Approximation. *IEEE Transactions on Automatic Control, 37*, 332-341. doi:10.1109/9.119632	568	32.98
Raney,R., Runge,H., Bamler,R., Cumming,I. and Wong,F. (1994) Precision SAR Processing using Chirp Scaling. *IEEE Transactions on Geoscience and Remote Sensing, 32*, 786-799. doi:10.1109/36.298008	315	25.63
Ottman,G., Hofmann,H., Bhatt,A. and Lesieutre,G. (2002). Adaptive Piezoelectric Energy Harvesting Circuit for Wireless Remote Power Supply. *IEEE Transactions on Power Electronics, 17*, 669-676. doi:10.1109/TPEL.2002.802194	407	23.72
Brown,M., Burschka,D. and Hager,G. (2003). Advances in Computational Stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 25*, 993-1008, doi:10.1109/TPAMI.2003.1217603	414	22.14
Franson,J (1989). Bell Inequality for Position and Time. *Physical Review Letters, 62*, 2205-2208, doi:10.1103/PhysRevLett.62.2205	471	20.16
Sharpe,W., Yuan,B. and Edwards,R.(1997) A New Technique for Measuring the Mechanical Properties of Thin Films. *Journal of Microelectromechanical Systems 6*,193-199. doi:10.1109/84.623107	285	16.85
Murphy,J. and Aamodt,L.(1980) Photothermal Spectroscopy using Optical Beam Probing - Mirage Effect. *Journal of Applied Physics, 51,* 4580-4588, doi:10.1063/1.328350	276	16.57
Ott,E. and Sommerer,J.(1994) Blowout Bifurcations - the Occurrence of Riddled Basins and on Off Intermittency. *Physics Letters A, 18*, 39-47, doi:10.1016/0375-9601(94)90114-7	378	15.34

Table 2: Top cited articles by fractional counting (TC_F) in year bins.

	Citation	Broad Topic	TC	↓TC_F
2011-2015	Boutet,S., Lomb,L., Williams,G.J., et al. (2012). *Science*	Protein Imaging	155	4.71
	A'Hearn,M. F., Belton,M.J.S., Delamere,W.A., et al. (2011). *Science*	Comets	109	3.92
	Russell,C. T., Raymond,C.A., Coradini,A., et al. (2012). *Science*	Minor Planets	135	3.43
2006-2010	Fujiwara,A., Kawaguchi,J., Yeomans,D., et al. (2006). *Science*	Asteroids	243	11.45
	Immel,T. J., Sagawa,E., England,S.L., et al. (2006). *Geophysical Research Letters*	Ionosphere	260	7.78
	Waite,J., Combi,M., Ip,W., et al. (2006). *Science*	Moons	245	6.97
1996-2005	Kanungo,T., Mount,D., Netanyahu,N., et al.(2002). *IEEE Transactions on Pattern Analysis and Machine Intelligence*	Clustering (computer science)	789	40.56
	Ottman,G., Hofmann,H., Bhatt,A. and Lesieutre,G. (2002). IEEE Transactions on Power Electronics	Power Supply	407	23.72
	Brown,M., Burschka,D. and Hager,G. (2003). *IEEE Transactions on Pattern Analysis and Machine Intelligence*	Audio Processing	414	22.14
1980-1995	Mirowski,M., Reid,P., Mower,M., et al. (1980). *New England Journal of Medicine*	Automatic Defibrillators	850	42.43
	Spall,J. (1992). *IEEE Transactions on Automatic Control*	Stochastic Approximation	568	32.98
	Raney,R., Runge,H., Bamler,R., et a. (1994). *IEEE Transactions on Geoscience and Remote Sensing*	SAR	315	25.63

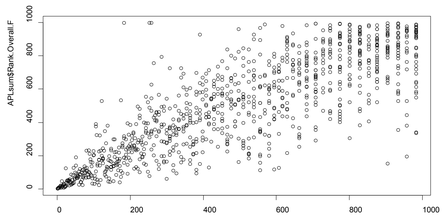

Figure 1: Graph of Fractional Count vs. Standard

Did fractional counting change the order? Yes. Although Spearman rank correlation r_S= 0.82 ($p<0.001$), we can see from Figure 1 and Tables 1 and 2 that many articles were shifted in the rankings.

Table 3: Most prolific authors (fractional counting)

Author	Articles$_F$
Lui, ATY	75.66
Meng, CI	64.64
Krimigis, SM	58.25
Newell, PT	49.70
Cheng, AF	41.56
Roelof, EC	41.01
Potemra, TA	40.35
Anderson, BJ	39.33
Sibeck, DG	36.51
Franson, JD	35.75
Greenwald, RA	33.00
Mitchell, DG	32.75
Lorenz, RD	32.14
Goldhirsh, J	31.35
Williams, DJ	30.27
Ohtani, S	28.97
Mauk, BH	27.92
Monchick, L	27.58
Paxton, LJ	27.48
Spall, JC	26.42

Descriptive information about the articles, venues, authors, and collaborating organizations can easily be exported from R using the bibliometrix package (Aria & Cuccurullo, 2017). Table 3 shows the most prolific authors published from 1980-2015. As with fractional counting of citations, fractional counting of authorship is used to rank independent and small science authors (highlighted in yellow) more highly when competing with PIs of big science instruments.

Topical analysis over time using LDA showed the shifting emphases of APL's work (Figure 2). Physical chemistry has dropped in importance while planetary science has remained at a high level with spikes as data are returned from major missions.

Articles Per LDA Topic Per Year

(by % of that year's articles, topics with >95 articles)

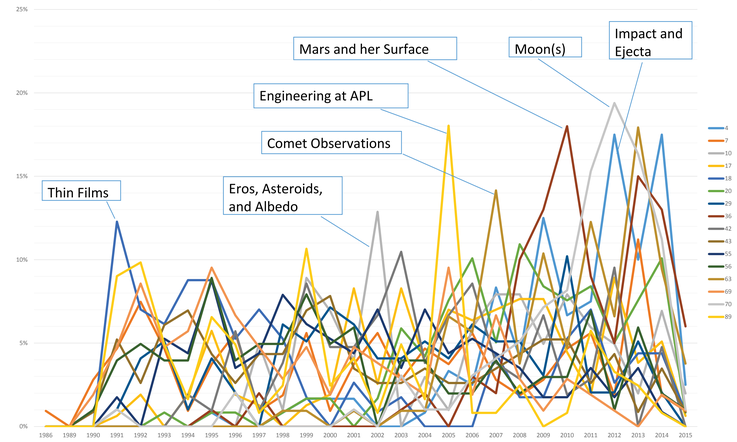

Figure 2: Topics Of Articles Over Time

Looking retrospectively, count of citations says little about whether the papers have continuing impact or if they were important of their time. Using longitudinal clustering (Genolini et al, 2016), we can identify articles that continue to be cited many years later and may still shape the lab's legacy (See Figure 3).

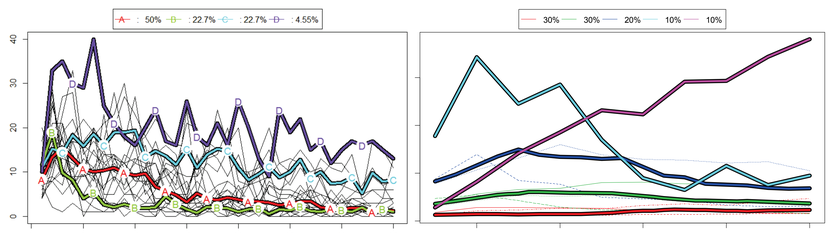

Figure 3: Citation clusters over time. (l) Berl (1986) articles over 35 years. (r) 1980-2005 articles over 10 years.

CONCLUSIONS

Once citation data are obtained, freely available tools can be used to perform analyses and visualizations that capture many different facets of the organization's scholarly output. Librarians can apply these tools in a valid and reliable way that is responsive to the requestor's needs.

ACKNOWLEDGEMENTS

I would like to thank Loet Leydesdorff for providing valuable pointers, insights, and guidance in completing this analysis, the APL Cool R group for helpful tips, and AG & SB for editing pointers. Some funding was provided by the APL Technical Digest office.

REFERENCES & LINKS

Scripts can be found on https://github.com/cpikas/institutionalprofiles Poster can be found on https://speakerdeck.com/cpikas

Aria, M. & Cuccurullo, C. (2017) bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics, 11,* 959-975. doi:10.1016/j.joi.2017.08.007

Berl, W. G. (1986). The 22 most frequently cited APL Publications. *Johns Hopkins APL Technical Digest, 7*, 221-232. http://www.jhuapl.edu/techdigest/views/pdfs/V07_N2_1986/V7_N2_1986_Berl.pdf

Genolini, C., Ecochard,R., Benghezal,M. Driss, R., Andrieu, S., & Subtil, F. (2016). kmlShape: An efficient method to cluster longitudinal data (time-series) according to their shapes. *PLoS One, 11*, e0150738. doi:10.1371/journal.pone.0150738

Kostoff, R. N., del Rio, J. A., Cortes, H. D., Smith, C., Smith, A., Wagner, C., ... Tshiteya, R. (2007). Clustering methodologies for identifying country core competencies. *Journal of Information Science, 33*, 21. doi:10.1177/0165551506067124

Leydesdorff, L., & Opthof, T. (2010). Scopus's source normalized impact per paper (SNIP) versus a journal impact factor based on fractional counting of citations. *Journal of the American Society for Information Science and Technology, 61*, 2365-2369. doi:10.1002/asi.21371

R Core Team (2017). R: A language and environment for statistical computing [Computer software]. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.