1-10-W

Bibliometric Analysis of Resting State Literature

Matthew K. Doherty[1], Ayesha Anwar[1], Sam Barberie[1], Caitlin Hinz[1], Michelle Kaplan[1], Anna Rachlin[1], Michael P. Milham[1,2], Stephen M. LaConte[3,4] and R. Cameron Craddock[4]

1 Child Mind Institute, New York, New York,
2 Nathan Kline Institute for Psychiatric Research, Orangeburg, New York
3 School of Biomedical Engineering and Sciences, Virginia Tech, Blacksburg, Virginia
4 Virginia Tech Carilion Research Institute, Roanoke, Virginia

Introduction

- Researchers are increasingly facing the challenges of searching, sorting, and digesting rapidly growing literatures within and across scientific disciplines.
- The Child Mind Institute (CMI) Librarian[1] initiative was created to address the growing need for comprehensive, hand-vetted reference libraries, tagged with descriptive labels capable of facilitating the rapid sorting and review of evolving literatures of interest.
- Mining of resting state (RS) literature is thus enabled by the CMI Librarian database. We performed a quantitative analysis of publication patterns within the resting state literature (a bibliometric analysis) to gain insight into
  - trends of publication rates,
  - prolific working groups of authors,
  - experimental methods and areas of focus, and
  - highest impact publications.

Publication Rates

- PubMed[2] queries were used to find growth of publication volume for resting state and all of fMRI.
- Piecewise exponential functions model growth of fMRI domain as well as its RS subdomain.

Working Group Analysis

- Developed greedy algorithm using Dice coefficient-based heuristics to identify working groups WG.
  - initialize WG to $\emptyset$
  - for each seed author s in order of nonincreasing $|\text{pubs}(s)|$,
  - set $a = s$
  - do
  - add a to A
  - find new author a with $\mathbf{arg\,max}_{a \notin A}(|\text{pubs}(a) \cap \text{pubs}(A)|)$
  - loop until $|\text{pubs}(a)| < \mathbf{10}$ or $|\text{pubs}(a) \cap \text{pubs}(A)|/|\text{pubs}(A)| < \mathbf{0.3}$
  - if $\sum |\text{pubs}(A_i) \cap \text{pubs}(A)| >= \mathbf{50}$ and $\sum I(|A \cap WG_i|/|WG_i| > \mathbf{0.2}) = \mathbf{0}$
  - add A to WG

Word Frequency Analysis

- Performed on n-grams from neuroimaging methods, cognitive ontology[3,4], and the PubBrain lexicon[5].
- Term frequency (tf) was computed as the ratio of occurrences of each term to the number of words in the corpus.
- Inverse document frequency (idf) was computed as $-\mathbf{log}(|\text{documents containing the term}|)$.
- Intuitively, high tf means the term is popular across the corpus, while high tf*idf means the term is popular in a small subset of the corpus.
- tf was plotted against tf*idf for the terms with the highest product of the two.

Citation Analysis

- Generated directed graph of citations directly from binary PDF publication files.
  - Performed fuzzy search for every possible publication title in each file using Python difflib.
  - Resulted in graph of 1,196 publications (nodes) and 17,183 citations (edges).
- Found publications with highest pagerank (impact estimate) using NetworkX[6] implementation.
- Found mean and standard error of pageranks using jackknife procedure.

Results

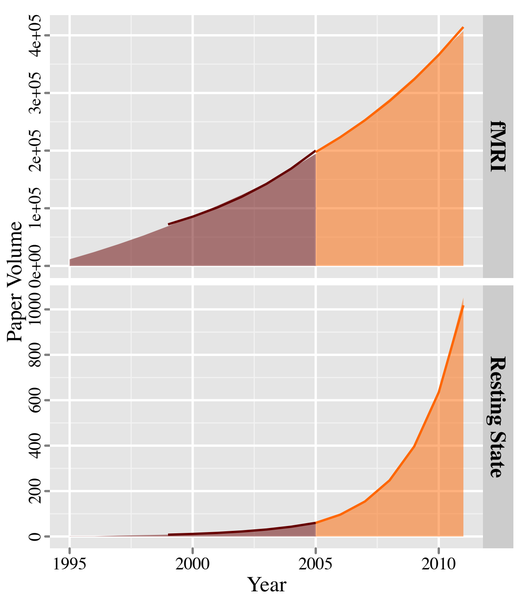

Figure 1 : Growth rates of fMRI and RSN volume

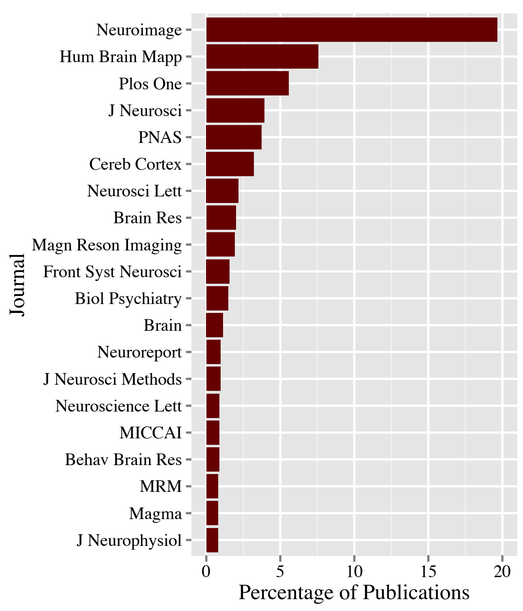

Figure 2 : Top journals by resting state publication count

- Growth of RS literature is piecewise exponential with 33% growth prior to 2005 and 47% after.
- In comparison, growth of fMRI literature was 29% before 2005 and 18% after.
- 27% of RS literature was published in Neuroimage and HBM.
- 26% of RS publications are open access, compared to 22% in all of fMRI.

- 23 authors (0.6% of 3704 total) in 4 working groups cover 17.5% of resting state publications.
- Mean papers by authors in these groups: 21.9. Overall mean: 0.310.
- Mean papers in common between pairs of authors in same/different working groups: 10.7/0.11.
- Working group authors (number publications in corpus by author):
  - Milham, M P (41) and Kelly, A M (36) and Castellanos, F X (37) and Biswal, B (36)
  - Jiang, T Z (35) and Li, Kun-Cheng (27) and Yu, Chunshui (19) and Tian, Li-Xia (18) and Zhou, Yuan (15) and Liu, Yong (15)
  - Gong, Qi-Yong (35) and Liu, Yi-Jun (21) and Liao, Wei (20) and Chen, Hua-fu (20) and Lu, Guang-Ming (18) and Zhang, Zhiqiang (17) and Zhong, Yuan (13)
  - Schlaggar, Bradley L (19) and Petersen, Steven E (16) and Cohen, Alexander L (13) and Fair, Damien (13) and Dosenbach, Nico U F (10) and Miezin, Fran M (10)

Results (continued)

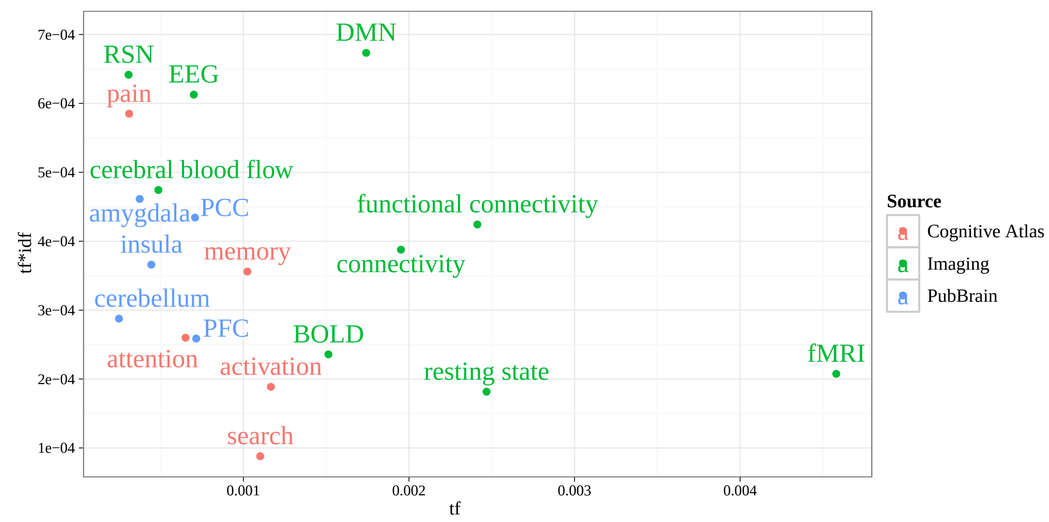

Figure 3 : tf and tf*idf top terms

- The most frequent imaging modality is fMRI, which is the corpus domain.
- The most investigated cognitive domains are memory and activation.
- The PFC, PCC, and insula are the most discussed brain regions.

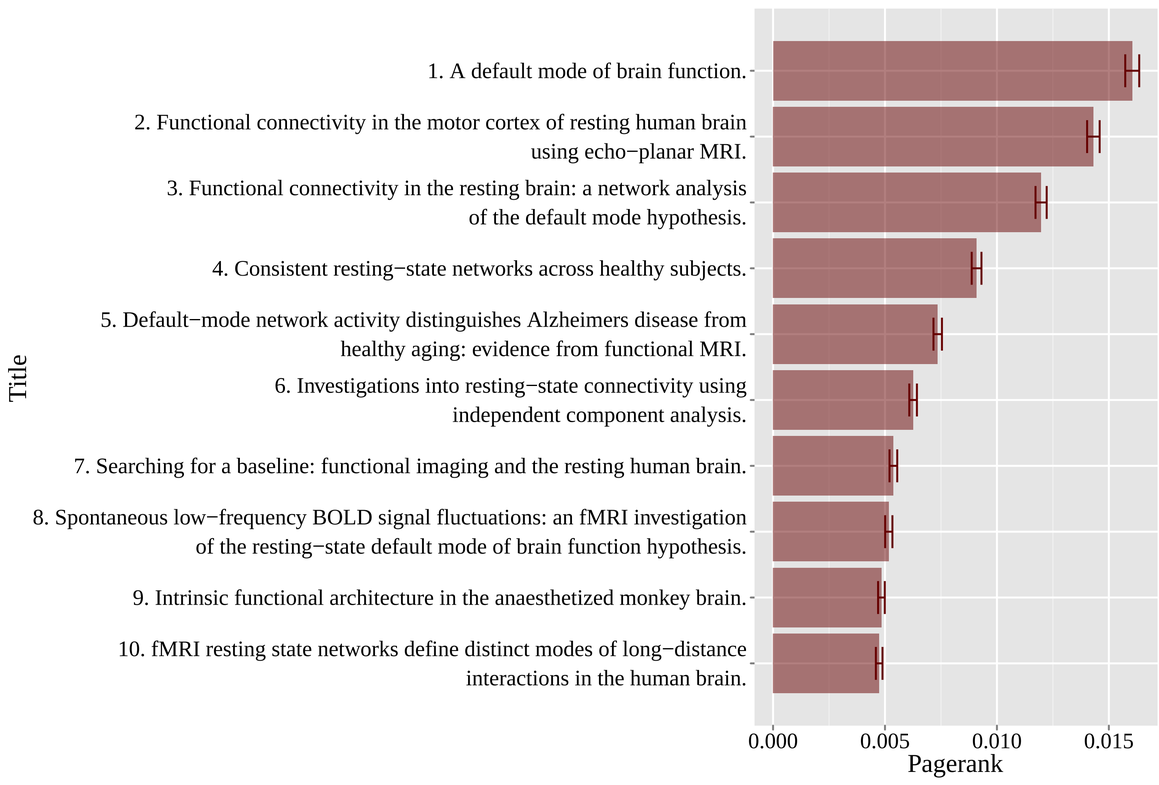

Figure 4 : Top resting state publications by pagerank

- The first few publications established the groundwork of the default and resting state networks.
- The top 10 publications are cited by 66% of the corpus.
- The top 1% of publications account for 10% of the total pagerank.
- After these publications, page rank falls off more slowly, with the next 10% of publications accounting for 40% of the total pagerank.
- Small world statistics:
  - Average clustering coefficient: 0.094
  - Average path length: 4.4

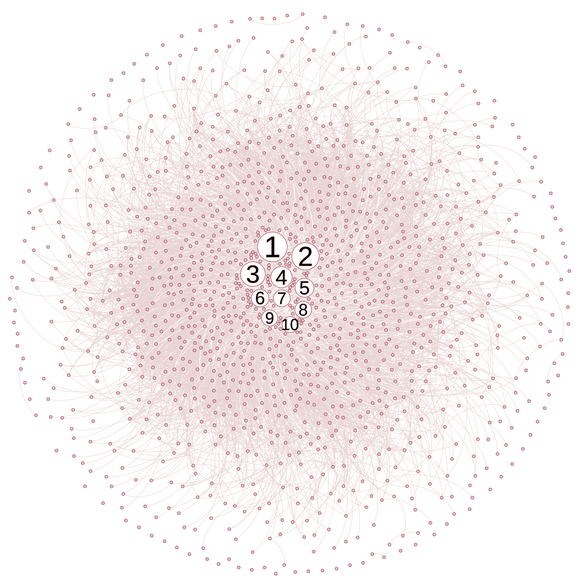

Figure 5 : Top pageranked publications in complete graph

Conclusions

- Bibliometric analysis lends valuable insight into the current state of the field, demonstrating its strength, areas of focus, and future potential.
- The growth of RS literature is currently faster than fMRI, the most common imaging modality in RS research.
- Analysis of open access showed that it is not universal in resting state or fMRI, but has a strong foothold.
- A few prolific working groups together cover 17.5% of RS publications, yet individually have little overlap.
- The analysis identified a focus on PFC, involved with executive function, as well as the PCC, which is a central node of the DMN. Activation was the most discussed cognitive domain, reflecting a current research trend.
- A handful of seminal publications provide much of the groundwork for the field. Relatively consistent pagerank among the remaining publications suggests that the field continues to grow in new and interesting directions.

Acknowledgements

This research was supported by the Child Mind Institute Endeavor Scientist program.

References: [1]Child Mind Institute. http://www.mendeley.com/profiles/cmi-librarian/ 2012, [2]National Center for Biotechnology Information, U.S. National Library of Medicine, http://www.ncbi.nlm.nih.gov/pubmed/ 2012, [3]Poldrack, R.A. http://www.russpoldrack.org/2011/07/my-analysis-of-ohbm-2011-abstracts.html 2011, [4]Poldrack, R.A., et. al. *Front. Neuroinform.* 5(17) 2011, [5]Kalar, D. http://www.pubbrain.org/ 2012, [6]Hagberg, A.A., et. al. *Proc. SciPy2008* 2008.
Top pageranked publications: [1]Raichle, M.E., et. al. *Proc. Natl. Aca.d Sci.* 98(2) 2001, [2]Biswal, B., et. al. *Magn. Reson. Med.* 34(4) 1995, [3]Greicius, M.D., et. al. *Proc. Natl. Acad. Sci.* 100(1) 2003, [4]Damoiseaux, J.S., et. al. *Proc. Natl. Acad. Sci.* 103(37) 2006, [5]Greicius, M.D., et. al. *Proc. Natl. Acad. Sci.* 101(13) 2004, [6]Beckmann, C.F., et. al. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 360(1457) 2005, [7]Gusnard, D.A., et. al. *Neuroscience* 2(10) 2001, [8]Biswal, B., et. al. *Hum. Brain Mapp.* 26(1) 2005, [9]Vincent, J.L., et. al. *Nature* 447(7140) 2007, [10]De Luca, M., et. al. *Neuroimage* 29(4) 2005.

mdoherty@childmind.org