Available online at www.sciencedirect.com

ScienceDirect

Procedia Social and Behavioral Sciences 2 (2010) 6582–6586

Procedia
Social and Behavioral Sciences

COINs2009: Collaborative Innovation Networks Conference

Alzheimer's Disease Research: A COIN Study Using Co-authorship Network Analytics

Aaron A. Sorensen[a] , Andrew Seary[b], Kenneth Riopelle[c]

[a]*Collexis Holdings, Inc, 1418 Laurel St, Suite 100, Columbia, SC 29201,USA*
[b]*School of Communication, Simon Fraser University, 8888 University Drive, Burnaby B.C. V5A 1S6, Canada*
[c]*Department of Industrial & Manufacturing Engineering, 4815 Fourth St., Room #2149, Wayne State University, Detroit, MI 48202, USA*

Elsevier use only: Received date here; revised date here; accepted date here

Abstract

Using bibliometric data from 269 Alzheimer investigators and the 167,142 researchers contained in their two-step collaboration network (i.e., co-authors and co-authors of co-authors), an eigen decomposition of the 13,254 unique Medical Subject Heading (MeSH) terms associated with the 43,736 papers authored by the Alzheimer researchers was performed. A correspondence-analysis-based transformation of the data produced a bench-to-bedside translational spectrum along which each of the original 269 Alzheimer investigators were placed. The spectrum was found to naturally divide into two partitions one of which housed basic scientists while the other grouped together clinical researchers. In addition to the semantic partitions, two main coauthor subgroups were isolated, and the authors who were most central to those co-author subgroups were analyzed for their ability to bridge the "translational divide" which separated researchers grouped in the "bench" (i.e. basic science) partition from those in the "bedside" (i.e., clinical investigation) partition. If a given research community can be partitioned into bench and bedside components, then the possibility exists to use such a dataset to identify people who might be best suited to attempt to bridge the "translational divide" which often exists between those researchers who make scientific breakthroughs in the lab and those clinical investigators capable of bringing the bench discoveries to the patients in the form of clinical trials.

Keywords: bibliometrics, informatics, scientometrics, correspondence analysis, MultiNet, Negopy, translational research, research networks, alzheimer's disease, Clinical and Translational Science Award, CTSA, social network analysis

INTRODUCTION

Over the last decade, the National Institutes of Health has put an increasing amount of emphasis on translational research in an effort to make more quickly available to humans medical breakthroughs made in the laboratory[1]. Nowhere is this desire more marked than at the NIH's National Center for Research Resources (NCRR) and among the academic institutions which make up the Clinical and Translational

1877-0428 © 2010 Published by Elsevier Ltd. Open access under CC BY-NC-ND license.
doi:10.1016/j.sbspro.2010.04.068

Science Award consortium[2]. The purpose of this study was to identify a scientometric approach which combined accepted social-network-analysis approaches and bibliometric datasets to determine an initial placing of Alzheimer's Disease (AD) scientists on the translational spectrum and to subsequently identify which of those researchers were best suited to perform the role of a bridge between the worlds of basic and clinical science.

METHODS

The initial selection of 269 AD authors was taken from the pool of scientists evaluated as part of a 2009 study to rank the top 100 investigators in the AD research community[3]. A two-step, co-author network (i.e., co-authors and co-authors of co-authors) based on the papers published by the investigators in question from January 1, 1980 to August 18, 2009 was retrieved from BiomedExperts.com (BME) and converted into the standard GraphML[4] format using Wayne-State-University conversion utilities. Extracting the Medical Subject Headings (MeSH) terms associated with the underling MEDLINE papers, allowed for an author-paper-concept network to be created in MultiNet[5]. As an analytical tool, MultiNet has been long used in the field of social network analysis in general and specifically within studies aimed at understanding the development of global virtual teams[6]. The resulting MultiNet dataset consisted of a total of 167,142 unique authors in the two-step network (generated from the initial seed of the 269 selected AD researchers); 43,736 MEDLINE papers authored by the 269 AD investigators; and 13,254 unique MeSH terms associated with those 43,736 papers (see *Figure 1* below). Of note is the fact that the original raw number of MeSH terms was somewhat higher than the number reported above as concepts which occurred only once in the corpus of papers were deleted as they were extraneous to the task of determining collaboration across authors.

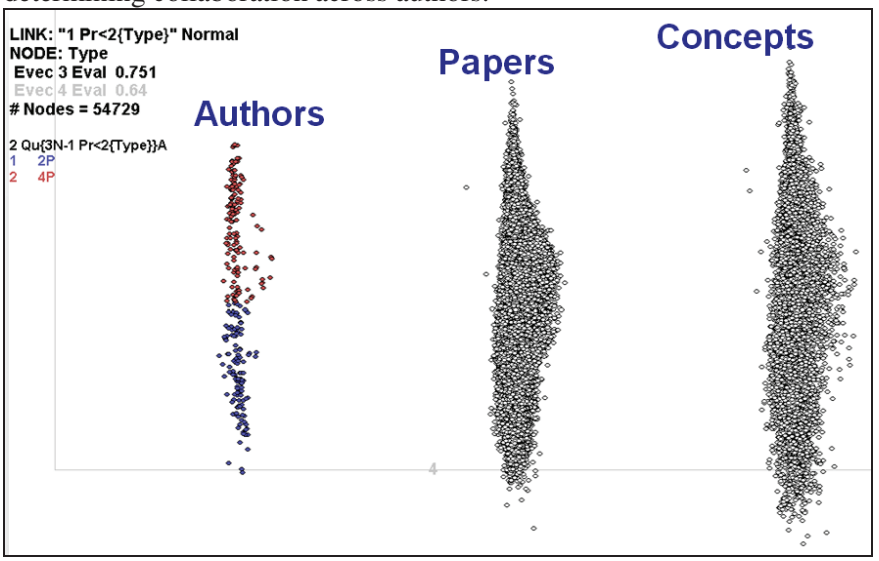

Figure 1- High-level results of the MultiNet correspondence analysis. Author partitions are indicated by the red/blue color-coding of dots.

The papers and concepts of the 269 AD authors where mapped such that those authors who had similar publications and concepts were closely aligned in Euclidean space as represented through MultiNet's visualizations of the results of the underlying correspondence analysis[7]. When looking at the frequency distribution of the authors, it was observed that the 269 AD authors fell into a bimodal distribution with the polarity of the eigenvector splitting the authors into two partitions (see *Figure 2*). The 269 authors were then assigned a new partition value based on the side of the bimodal distribution on which each fell. This partition value was then imported into the co-author network node data. Each of the 269 authors now had a partition node variable with a value of 1 or 2.

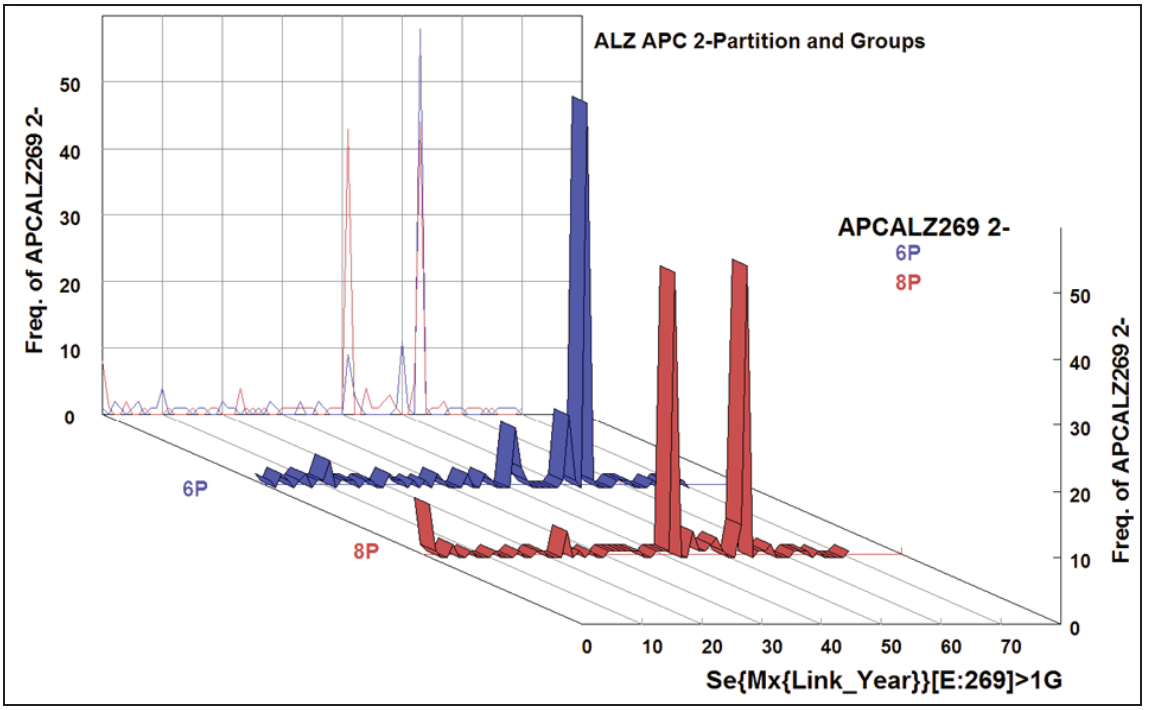

Figure 2 – the bimodal nature of the coauthor network can be seen through aggregation of the two partitions (i.e.,"6P" and "8P") pictured on the checkerboard plane (i.e., the far wall) in the above visualization.

The co-author network was next limited to authors who had collaborated on a paper with at least one of the other 269 authors. A longevity index (i.e., the count of the number of years a given pair of authors has collaborated) was calculated for all the 269 author pairs. A subsequent refining of the data limited the co-author set to those authors who had at least two years of ongoing collaboration thereby eliminating those collaborations which had only occurred once. MultNet's Negopy clustering algorithm was run to determine the unique network subgroups and roles of the 269 authors. The authors fell into two main subgroups. Within each subgroup the top 10% of authors in terms of their centrality in the co-author network of their subgroup were selected. From this pool of the most central authors in each subgroup, a crosstab by partition was done to see which of the highly-central authors contributed most to the cross-partition collaboration. In other words, the authors who were highly collaborative within the two subgroups (as defined by co-author relationships) and showed the ability to bridge the two partitions (determined by the affinity of MeSH terms used in papers) were assessed. Virginia M.-Y. Lee, Ph.D. of the University of Pennsylvania was determined to be the highly-central author who provided the most collaborative bridging between the two semantically-divergent partitions.

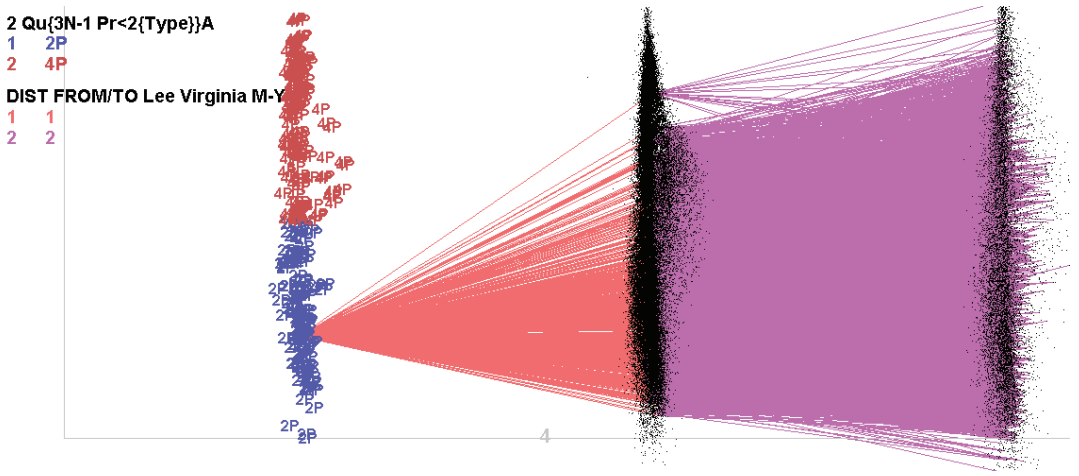

Figure 3 – Correspondence-analysis results for Virginia M.-Y. Lee showing the breadth of coauthor coverage (i.e., middle band) and of MeSH coverage (i.e., right-hand band) on the translational spectrum.

RESULTS

In analyzing the partition values of the 269 scientists, it became patently clear that the axis along which MultiNet's correspondence analysis had stratified the authors was a bench-to-bedside axis. In other words, by looking at the MeSH terms associated with the papers belonging to the authors at either end of the spectrum, it became clear that one partition (i.e. , the red partition) was made of human-subject oriented investigators who were characterized by MeSH terms such as "Computerized Medical Records," "Urban Health Services," and "Nontherapeutic Human Experimentation" while the other was comprised of basic scientists more likely to be using animals in their experimental designs and characterized by MeSH terms such as "Crotalid Venoms," "Armadillo Domain Proteins," and "Bufo marinus."

If a given research community can be partitioned into bench and bedside components, then the possibility exists to use such a dataset to identify people who might be best suited to attempt to bridge the "translational divide" [8] which often exists between those researchers who make scientific breakthroughs in the lab and those clinical investigators capable of bringing the bench discoveries to the patients in the form of clinical trials. The example given above of Dr. Virgina Lee is an example of this potential. Lee was categorized by MultiNet as being one of the most centrally connected authors in the basic scientist partition. At the same time, however, Lee was identified as being the researcher who had best shown the ability to bridge the AD translational divide by collaborating with researchers in the bedside partition. In fact, if one does a PubMed search on Dr. Lee and limits the search to AD clinical trials, one finds four papers. In the community of AD basic scientists, it is very rare to find an investigator with any papers that have been officially tagged in MEDLINE as having reported on the results of a clinical trial.

CONCLUSIONS

The approach used in this study to automatically stratify AD investigators along the bench-to-beside spectrum appears to be a reliable and reproducible approach that can be used to group investigators within any biomedical research community.

An interesting direction for future research is that of refining the ability to identify those researchers most likely to contribute to the bridging of the translational divide. If a list of such

investigators could be communicated to decision makers who were in the process of pulling together translational research teams, then empirical studies might be done to assess the ability of this approach in helping with the identification of potential linchpin members of translational teams.

REFERENCES

[1] Zerhouni E. (2001). Translational and clinical science – time for a new vision. *New England Journal of Medicine* **353**, 1621-1623.

[2] CTSA Institutions [Internet]: National Institutes of Health; c2006 [cited 2010 3/21/2010]. Available from: http://www.ctsaweb.org/index.cfm?fuseaction=home.showPartInstList.

[3] Sorensen AA (2009). Alzheimer's Disease Research: Scientific Productivity and Impact of the Top 100 Investigators in the Field. *Journal of Alzheimer's Disease* **16**, 451–465.

[4] Brandes U, Pich C (2004). GraphML transformation. *Proceedings of the 11th International Symposium on Graph Drawing* **3383**, 89–99. Springer, 2004.

[5] Seary AJ, Richards WD (1995). Partitioning Networks by Eigenvectors. Presented to European Network Conference, London. Published in Everett MG, Rennolls K (Eds). *Proceedings of the International Conference on Social Networks* **1**, 47-58.

[6] Gluesing J, Alcordo T, Baba M, Britt D, Wagner K, McKether W, Monplaisir L, Ratner H, Riopelle K (2003). The development of global virtual teams. In C. B. Gibson, S. G. Cohen (Eds.), *Virtual teams that work: Creating conditions for virtual team effectiveness:* 353–380. San Francisco: Jossey-Bass.

[7] Hill MO (1974). Correspondence analysis: a neglected multivariate method. *Applied Statistics* **23**, 340-354.

[8] Rho JM, Sankar R, Cavazos JE (2004*) Epilepsy: Scientific Foundations of Clinical Practice*. New York, NY: Marcel Dekker.