MEDINFO 2023 — The Future Is Accessible
J. Bichel-Findlay et al. (Eds.)

doi:10.3233/SHTI231074

Defining the Boundaries of Psychiatric and Medical Knowledge: Applying Cartographic Principles to Self-Organising Maps

Andrew AMOS[a1], Kyungmi LEE[b], Tarun Sen GUPTA[a] and Bunmi Malau-Aduli[a]
[a]*College of Medicine and Dentistry, James Cook University, Townsville, Australia*
[b]*College of Science and Engineering, James Cook University, Cairns, Australia*
ORCiD ID: Andrew Amos https://orcid.org/0000-0002-9145-0212, Kyungmi Lee https://orcid.org/0000-0003-3304-4627, Tarun Sen Gupta https://orcid.org/0000-0001-7698-1413, Bunmi Malau-Aduli https://orcid.org/0000-0001-6054-8498

Abstract. Biases in selection, training, and continuing professional development of medical specialists arise in part from reliance upon expert judgement for the design, implementation, and management of medical education. Reducing bias in curriculum development has primarily relied upon consensus processes modelled on the Delphi technique. The application of machine learning algorithms to databases indexing peer-reviewed medical literature can extract objective evidence about the novelty, relevance, and relative importance of different areas of medical knowledge. This study reports the construction of a map of medical knowledge based on the entire corpus of the MEDLINE database indexing more than 30 million articles published in medical journals since the 19th century. Techniques used in cartography to maximise the visually intelligible differentiation between regions are applied to knowledge clusters identified by a self-organising map to show the structure of published psychiatric evidence and its relationship to non-psychiatric medical domains.

Keywords. Machine learning, medical informatics, science of science, medical education, information science

1. Introduction

Modern medical curricula are almost exclusively developed using consensus methods like the Delphi technique, which are prone to biases which have adversely affected women and other groups in various ways [1,2]. A severe lack of readily accessible and objective evidence on which to base curriculum development decisions has contributed to the problem. Visualisations of the published medical literature based on health informatics and cartographic techniques are an underutilised source of objective evidence about medical knowledge that may be capable of reducing bias.

[1]Dr Andrew Amos, andrew.amos@jcu.edu.au, College of Medicine and Dentistry, Building 39, James Cook Drive, James Cook University, Townsville, QLD 4811, AUSTRALIA +61 412 221 736

With the exponential increase in the production of scientific knowledge, a paradigm for understanding and describing universal and domain-specific mechanisms of scientific research and progress known as the Science of Science (SciSci) has recently emerged [3], which Skupin and colleagues extended using cartographic techniques to improve concision and intelligibility. Core to their approach are self-organising maps (SOMs), a method for mapping scientific knowledge that can project high dimensional datasets onto 2-D maps while retaining many of the topographical features of the original space [4-6]. Biuk-Aghai and colleagues show the versatility of Skupin's cartographic approach by visualising the distribution of author counts across multiple Wikipedia language editions [7].

One of the key problems of medical curriculum development is defining the boundaries between specialist medical domains such as internal medicine, surgery, and psychiatry; and then deciding on the relative priority to be given to each in any given training program. In order to demonstrate that it is feasible to produce a SciSci map that could address this issue, the current research aimed to develop a SOM of the structure and boundaries of psychiatric knowledge within the map of broader medical knowledge represented by the published peer reviewed medical literature. The map was designed to be an objective source of evidence suitable for guiding the construction of psychiatric and general medical curricula, capable of acting as a backdrop of well-established knowledge against which emerging topics can be contrasted and understood in their full context.

2. Methods

We downloaded complete sets of the MEDLINE and MeSH databases on 01.01.21 https://www.nlm.nih.gov/databases/download/pubmed_medline.html. The large dataset necessitated the use of sparse matrix representation of article data. We applied Kohonen's batch SOM algorithm, which is computationally less intensive than the original algorithm [5]. In the absence of standard methods, we experimented with different SOM sizes, starting with the 275 x 275 SOM used by Skupin et al [4] for their much smaller dataset, looking for a plateau of topographic error to guide the final size.

After training the SOM, we applied *neuron label clustering* in which the weights used to transform inputs into each node/neuron of the trained SOM are ranked, and the largest weight defines the term dominance for that node. Adjacent nodes with the same term dominance were clustered [4,6]. Clusters were tagged as psychiatric if they included a MeSH with a psychiatric code (all codes starting with an F; for example, the MeSH "Affective Disorders, Psychotic" is coded F03.700.150). The open source platform QGIS was used to create maps using color to differentiate psychiatric and non-psychiatric clusters.

3. Results

Our SOM included data from 33,375,863 MEDLINE articles and 29,917 distinct MeSH codes. We used a 350 x 350 node SOM after topographic error plateaued at 0.264 at this size. Figure 1 shows all term dominance clusters differentiating between those with (with a red font) and those without a psychiatric tag (black font).

Figure 1. Psychiatric (black) and non-psychiatric (red) term dominance clusters (high quality image available at: https://raw.githubusercontent.com/AndrewAmosJCU/PsychSOM/main/ColorCoded.png).

Figure 2 expands the area of the map around the term dominance cluster "Geriatric Psychiatry", showing the association between psychiatric and medical knowledge in related areas is retained in the visualisation of the SOM. The embedding of the psychiatric MeSH *Geriatric Psychiatry* within a complex of medical MeSH is consistent with the reality of geriatric psychiatry, which combines both psychiatric and medical care and is often considered closely related to consultation liaison psychiatry for that reason.

The psychiatric MeSH surrounding *Affective Symptoms* (not shown due to space constraints) include behavioural concepts such as *Sick Role*, related diagnostic groupings including *Depression* and *Panic Disorder*; related physiological states such as *Anger* and *Fear*; and predisposing factors such as *Antisocial Personality Disorder*, *Stereotyping*, and *Anxiety, Castration*. These psychiatric concepts are closely aligned with relevant medical concepts including treatment options *Benzodiazepines* and *Anticonvulsants*, with symptoms or side effects such as *Fatigue* and *Bronchoconstriction*.

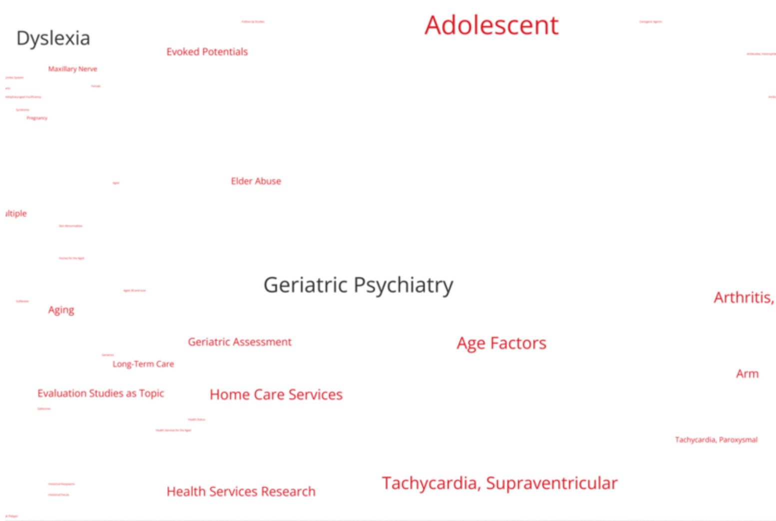

Figure 2. Focus on term dominance clusters around Geriatric Psychiatry.

While most of the contiguous MeSH can be readily conceptually linked in meaningful ways, some of the associations require more consideration. For example, the apposition of *Enuresis* to a region including *Fear*, *Depression*, and *Antisocial Personality Disorder* can be explained by the near neighbouring concept of *Anxiety, Castration*, which suggests the psychoanalytic link between castration anxiety and bed-wetting.

4. Discussion

The results show it is possible to generate a map that summarises the entire set of articles indexed by the MEDLINE database in term dominance clusters that differentiate between psychiatric and non-psychiatric literature. For the purposes of curriculum development, the global perspective of a map derived from the most objective set of data available is less likely to be affected by systematic bias than the judgments of individual experts, or even of groups of experts from within a medical specialty or sub-specialty.

Applying cartographic techniques to SOMs provides a concise visual summary of all available peer reviewed medical evidence that can help experts understand the structure and boundaries of their own domains of expertise, and their place within the broader context of all medical knowledge. In addition to the intrinsic value of an objective standard against which to measure existing curricula, this will facilitate the identification of curricular gaps and proposed changes to address gaps or incorporate new knowledge. Future research could use this comprehensive SOM as a baseline on which to project more specific data, such as emerging topics of increasing medical research activity.

While our research has limitations characteristic of exploratory research, including using trial and error to select the size of network [4]; and accepting that the convergence properties of SOMs are not well characterised [8] we have mitigated the worst associated problems by grounding our decisions on past research and relying upon relevant quality measures such as topographic error where indicated [9].

5. Conclusions

A cartographically informed approach to SOM visualisation makes objective evidence about the structure and relationships of medical knowledge available in a condensed but readily intelligible form. As an input into medical curriculum development, it has the potential to reduce historical biases that have disadvantaged women and other groups by providing a comprehensible view of the entire scope of medical knowledge.

References

[1] Thomas PA, Kern DE, Hughes MT, Tackett SA, Chen BY, editors. Curriculum development for medical education: a six-step approach. Maryland: Johns Hopkins University Press; 2015.

[2] Sheikh MH, Chaudhary AM, Khan AS, Tahir MA, Yahya HA, Naveed S, Khosa F. Influences for gender disparity in academic psychiatry in the United States. Cureus. 2018 Apr;10(4):e2514, doi: 10.7759/cureus.2514.

[3] Fortunato S, Bergstrom CT, Börner K, Evans JA, Helbing D, Milojević S, Petersen AM, Radicchi F, Sinatra R, Uzzi B, Vespignani A. Science of science. Science. 2018 Mar;359(6379):eaao0185, doi: 10.1126/science.aao0185.

[4] Skupin A, Biberstine JR, Börner K. Visualizing the topical structure of the medical sciences: a self-organizing map approach. PloS One. 2013 Mar;8(3):e58779, doi: 10.1371/journal.pone.0058779.

[5] Kohonen T. Self-organizing maps. 3rd ed. Berlin: Springer, 2001.

[6] Skupin A. A cartographic approach to visualizing conference abstracts. IEEE Comput Graph Appl. 2002 Aug;22(1):50-58, doi: 10.1109/38.974518.

[7] Biuk-Aghai RP, Pang CI, Si YW. Visualizing large-scale human collaboration in Wikipedia. Future Gener Comput Syst. 2014 Feb;31:120-33, doi: 10.1016/j.future.2013.04.001.

[8] Cheng Y. Convergence and ordering of Kohonen's batch map. Neural Comput. 1997 Nov;9(8): 1667-76, doi: 10.1162/neco.1997.9.8.1667.

[9] Pölzlbauer G. Survey and comparison of quality measures for self-organizing maps. In: Proceedings of the Fifth Workshop Data Analysis; 2004 Jan; Slovakia.