Michael Gunn
LIS 9411
Professor Liwen Vaughan
August 8th, 2014

Does the Co-citation Map Predict Disciplinary Clusters in Neuroplasticity Theory?

By the turn of the 21st century, a new trend was rapidly growing in the scientific community. Research projects, out of necessity, were becoming less specialized and more holistic in theme. While many scientists continued to work in the same communities of specialized experts, others were looking for opportunity to do research in unfamiliar disciplines. The long-standing tradition of specialization in science has run its course and in order to find creative solutions in the now corporate-driven academic system, the traditional disciplines have mixed to produce a variety of new fields, including: biotechnology, nanotechnology, human-computer interaction, information science, education, business administration, philosophy of science, and neuroscience (Weingart, 2010).

Perhaps the most popular and well-established interdisciplinary field is neuroscience. Historically, neuroscience begins in the late nineteenth century in the labs of German physiologists studying the leopard frog. They observed that living muscles would contract when touched by a charged electrode. These muscles would also contract if just the nerve that innervated the muscle was touched by the electrode (Waters, 1885). Experiments such as these demonstrated that the spirit[1] which passes through the brain and nerves is a form of electricity. Shortly after this time, histologists were staining, isolating, and drawing neurons; discovering that they take many different forms with respect to axonal and dendritic branching (Nicholls *et al.*, 2001, p. 5). The intracellular fluid of squid giant axons was the first to be physically and chemically analyzed because their axons can conveniently grow to be 1mm in diameter; it is from this line of research that the Nernst Equation was derived (Nicholls *et al.*, pp. 35-36). The

[1] "I shall not at present meddle with the Physical Consideration of the Mind; or trouble myself to examine, wherein its Essence consists, or by what Motions our Spirits, or Alterations of our Bodies, we come to have any Sensation by our Organs, or any *Ideas* in our Understandings; and whether those *Ideas* do in their Formation, any, or all of them, depend on Matter, or no." Locke, 1690, p. 13.

advent of biochemistry in the mid-twentieth century revolutionized neuroscience by introducing drug-receptor interactions to research (Sengupta, 1989). This led to major advancements in cellular biology, genetics, biophysics, physiology, and pharmacology (Nicholls *et al.*, 2001, pp. 19-21). Psychologists have developed cognitive science as a means to relate physiology with behaviour (Kolb, & Whishaw, 2008). Developments in neuroscience inspired President George Bush Sr. to proclaim that the 1990s would be the “Decade of the Brain” (Bush, 1990). Today, computer scientists attempt to model neuronal networks, and there is an international project to simulate an entire human brain with one supercomputer (Pearn, 2014). A few curious philosophers are producing commentary on theories in neuroscience, relating them to age-old inquiries into the nature of consciousness, perception, experience, and our understanding of the external reality (Askenasy, & Lehmann, 2013). Neuroscience is truly the most interesting and popular research field, and it is imperative that information scientists study its progress and help to improve its infrastructure. The goal of the present study is to assess the viability of author co-citation analysis (ACA) to study interdisciplinarity in neuroscience.

The Study of Interdisciplinarity

It is important to begin any study with the clarification of terms. Definitions for relevant terminology from the library and information science (LIS) perspective follow:

> Disciplinarity – A highly honed approach with focused objectives and specific methodological and technical characteristics. Specialized nomenclature and consensus-driven protocols and procedures are maintained.
> Multidisciplinarity – Several disciplines involved, providing their unique perspectives without actually melding. Disciplines come together to explore phenomena and work on stated objectives, while retaining their singular characteristics.
> Interdisciplinarity – Two or more disciplines actively engaged, synthesizing their efforts within a given range of objectives and conditions. Techniques and methodologies mesh and meld in order to accomplish objectives.
> Hérubel, 2010, p. 27.

Hérubel's definitions emphasize methodology and protocol as the hallmarks of a discipline. They ignore the contributions that institutional organization and publication outlets have on the formation and maintenance of disciplines, but from a bibliometric perspective, the former consideration is a nonissue. Publication outlets (i.e. journals) are known to play a major role in the development of research fronts (Bruer, 2010; Lee, 2005). The distinction between multidisciplinarity and interdisciplinarity is especially important to analyze. According to Klein, "'Multidisciplinarity' signifies the juxtaposition of disciplines. It is essentially *additive*, not *integrative*. Even in a common environment, educators, researchers, and practitioners still behave as disciplinarians with different perspectives." (Klein, 1990, p. 56.). The purpose of the present study is not to measure how disciplines collaborate *per se*, since collaboration does not necessarily involve communication and the sharing of methodologies and ideas. Many interdisciplinary research fields have been studied using citation analysis, including astrobiology (Gowanlock, & Gazan, 2013), climate change research (Bjurström, & Polk, 2011), agriculture (Morooka, Ramos, & Nathaniel, 2014), business administration (Kushkowski, & Shrader, 2013), life sciences (Baumwol *et al.*, 2011), LIS (Huang, & Chang, 2011; Lu, & Wolfram, 2012; White, & McCain, 1998), the social sciences (Levitt, Thelwall, & Oppenheim, 2011), and neuroscience (Braun, Glänzel, & Schubert, 2001; Bruer, 2010; Schwechheimer, & Winterhager, 2001; Sengupta, 1989). Citation analysis is an appropriate tool to measure interdisciplinarity because a citation is evidence that the cited work was read and digested by the author, and that its impact on the author's own work was significant enough to justify its inclusion in the citing document.

There are many disciplines that contribute to the field of neuroscience, amounting to tens of thousands of researchers worldwide (Society for Neuroscience, 2014). But it does not necessarily follow that these contributions are appreciated by researchers beyond the contributor's discipline. Palmer argues that interdisciplinary researchers require support from librarians to build new indices and classification schemes that recognize the interdisciplinarity of subjects (Palmer, 2005; 2010). Neuroscientists are also known to have difficulty reviewing literature that is outside of their speciality, generally due to a lack of knowledge on classical works and prominent authors in other fields; their literature searches have less direction and their research goals are unstructured (Palmer, Cragin, & Hogan, 2007). Researchers tend to share information on a need-to-know basis when collaborating across disciplines, which leads to communication barriers that inhibit progress (Haythornthwaite, 2006). There is also debate about the degree to which neuroscientists should share their raw data, a practice that has greatly benefited the fields of genetics (National Human Genome Research Institute, 2014) computer science (GitHub, 2014), anthropology, economics, and demography (Jacoby, 2010). But neuroscientists have been reluctant to share their unpublished CT scans, electron micrographs, electrophysiological recordings, questionnaires, interviews, and other forms of raw data (Koslow, 2000; 2002). Thankfully, federal granting agencies like the National Science Foundation and Wellcome Trust are working to change this trend by encouraging applicants to share their raw data (Jacoby, 2010, pp. 79-80).

Despite the challenges of interdisciplinarity, there is evidence that scientific infrastructure can make it easier for disciplines to communicate. For example, Bruer (2010) used ACA to study the history of the journal *Mind and Brain* over 25 years. At its inception in 1980, only a few authors were facilitating dialogue between neurobiologists and psychologists on the journal's topic: the neuroscience of attention. As time passes, the number of co-citations between neurobiologists and psychologists increases, resulting in the development of cognitive neuroscience of attention by 1990.

Co-citation Analysis

Citation analysis has been around since the beginning of bibliometrics (Garfield, & Sher, 1963; Price, 1965). Co-occurrence analysis is a technique used to study a network of entities by quantitatively measuring and analyzing their relationships. Small (1973) was the first bibliometrician to propose the use of co-citation analysis to measure document relationships. The first co-citation analysis was done at the article level by Small and Griffith, (1974). The technique was eventually adapted to the author level by White and Griffith (1981). Many other authors have used this technique to measure other aspects of the bibliographic universe.

To accompany their article citation analysis, Schwechheimer and Winterhager (2001) interviewed an expert in the field of retrograde amnesia about the MDS map they generated from their analysis of ISI *Neuroscience Citation Index*. The expert was familiar with almost all of the highly cited articles used in the study, and was able to interpret the co-citation map as it pertains to the history of research on retrograde amnesia. The authors concluded that co-citation cluster analysis is a viable instrument for identifying research fronts. However, the present study analyzes co-citations at the author level, rather than the article level. Some authors have stated that going beyond the article level compromises some of the accuracy of the data (Andrés, 2009). The sacrifice of accuracy is made in order to enhance the scope of the study from single articles to entire bodies of writings (White, & Griffith, 1981). More articles factor into the analysis, making ACA a more economical method of data collection.

Co-citation analyses require the gathering and processing of large amounts of data. These data are typically mined by algorithms and manually checked for errors. In order to make such large amounts of data readable, multidimensional scaling (MDS) is a favourable statistical test because it allows the data to be visualized in a way that humans can easily read and interpret

(Kruskal, & Wish, 1978). MDS is a statistical test that plots a co-occurrence matrix onto a multidimensional plot (usually a 2-dimensional plot in bibliometrics) where individual entities are represented by points on the plot, and the relative locations of the points express their similarity in terms of the measured value. In the case of ACA, a similarity matrix is constructed from author co-citation frequencies. These frequencies have to be normalized to bring the data into relative terms. The present study's methodology draws from Leydesdorff and Vaughan (2006), particularly with respect to the method of data normalization and to the use of SPSS to produce the MDS map. A study by Kreuzman (2001) was also useful because being a philosopher, he needed to learn ACA and MDS as he went along, so his explanation of the procedure is elementary and makes no assumptions about the reader's knowledge.

The Present Study

The purpose of this study was to measure the level of cross-disciplinary co-citations in the field of neuroscience. This is a preliminary study to test the feasibility of author co-citation analysis (ACA) for measuring interdisciplinarity. Four MDS maps were constructed from three co-citation matrices to make three comparisons. The hypotheses of these three comparisons are stated below:

1. *First Initial vs. All Initials*: Querying author names by surname and first initial only will produce a more accurate map than querying by surname, first initial, and all middle initials. Querying middle initials will reduce noise for authors with middle initials, but articles that don't print their middle initials will not be collected.
2. *First Initial vs. First Initial With Limits*: Maps will have significantly less noise if queries are limited by the subject categories relevant to the authors in the sample.

3. *First Initial vs. First Initial No Normalization*: Data normalization by Jaccard's Index does not significantly change the MDS map.

Materials and Methods

Kreuzman (2001), reports taking five steps to perform an ACA: 1) specify authors and time period; 2) collect raw co-citation data to produce a co-citation matrix, 3) normalize raw data into relative coefficients; 4) analyze with MDS; and 5) interpret the resulting map. The present study uses a comparable methodology, which is explained in detail below.

Software

There are three major citation indices in the market of academic metadata: ISI Web of Science, Elsevier Scopus, and Google Scholar. There are pros and cons to the use of each of these databases. WoS is the original academic citation index (Garfield & Sher, 1963) and has the longest temporal coverage. Scopus is much newer, and only has coverage from 1995 onward, but it was developed with the mistakes of WoS in mind (Andrés, 2009). Google Scholar is unique because it is free to use and covers a wider range of academic works than WoS or Scopus. The inclusion of books and non-scholarly periodicals is a significant advantage for Google Scholar because many researchers, especially in the humanities rely on book citations to prove their productivity (Currie & Monroe-Gulick, 2013). However, because Google Scholar takes so many sources into account, it tends to report a greater number of citations than either WoS or Scopus (Andrés, 2009). WoS was used in the present study because a) it is still the most commonly used citation index and has the widest temporal converage, b) I have boycotted Elsevier (Neylon, 2014), and c) Google search algorithms are trade secrets, and querying with an unknowable search algorithm is not appropriate for academic research.

All spreadsheets were made with Microsoft Office Excel 2013. All MDS maps were generated with IBM SPSS 17, and clusters were added with Adobe Photoshop CS6. The report was written with Microsoft Office Word 2013. Web browsing was done with Microsoft Firefox 26.

Selection of Authors

The number of neuroscientists worldwide is in the tens of thousands (Society for Neuroscience, 2014). To scale the study to a reasonable level, the special topic of neuroplasticity[2] was used as a sampling focus. The sampling method was to mine authors' names from scholarly reviews on neuroplasticity theory, and then to select the most highly cited authors from those reviews based on how many articles are in the reference section. Bruer (2010) mined two review articles on the subject of attention, one from the neurobiological perspective and the other from the neuropsychological perspective, to develop a sample of 'core authors.' His definition of a core author is, "an author cited five or more times in one of the two reviews." In the present study, the most prolific authors were mined from reviews, but my threshold for sampling is different (see below).

Many previous studies only mined the first author of each study, but this limits the exposure of many authors and assumes that the principle investigator is always the first author. For cases where only the first six authors were listed in the review (a rule of the American Psychological Association's citation style) the rest of the names were found using the Western Libraries' *Summon* portal.

---

[2] Neuroplasticity refers to the inherent changeableness of the nervous system as a response to physical, chemical, and behavioural stimuli (Kolb, & Whishaw, 2009, pp. 656-657; Nicholls *et al.*, 2001, pp. 228-229).

I am aware of the historical relevance of three disciplines in the research on neuroplasticity theory. In broadly construed terms, these disciplines are: biology, psychology, and medicine. For the sake of compatibility with the WoS citation index, these three disciplines were operationally defined by WoS subject categories. WoS subject categories have been used as representation of discipline in previous studies (Boyack, Klavans, & Börner, 2005; Levitt, Thelwall, & Oppenheim, 2011; Leydesdorff & Rafols, 2011), they are assigned to journals based on the scope reported by each journal, as compared to ISI's own scope notes on subject categories (Institute for Scientific Information, 2014). Articles are categorized by the journal that published them. Levitt and colleagues (2011) considered WoS categories to represent the disciplinarity of an article, but noted the limitations a) that WoS categories often overlap, and b) that disciplines are assigned to journals, which creates a problem of precision at the article level. This is also problematic because it gives the ISI authority to decide what discipline every journal covers, when historically, the academic disciplines formed as a result of the social networks of university lecturers in the eighteenth and nineteenth centuries (Weingart, 2010). However, because each journal selects articles for publication based on its own disciplinary criteria, this is likely not a problem for the present study.

Reviews were found in WoS Advanced Search by querying both 'neuroplasticity' as the topic and one of the WoS subject categories, and combining these two basic searches with the AND operator. This is essentially a topic search limited by subject category. Topic search finds the presence of the queried words in the title, abstract, and author-assigned keywords of each document. Results were limited to English reviews. I qualitatively analyzed the title, abstract, and keywords of each review to narrow down the most appropriate one from which to mine authors. The selected reviews and their respective journals' WoS subject categories are:

- Mattson (2007) [Biochemistry & Molecular Biology; Neurosciences]
- Rabipour & Raz (2012) [Psychology; Neurosciences]
- Vinogradov, Fisher, & de Villers-Sidani (2012) [Psychiatry; Pharmacology & Pharmacy; Neurosciences]

These three reviews provide recent analysis of the key authors in neuroplasticity from three important disciplines of neuroscience. Mattson (2007) describes the research on the mitochondria's role in neuroplasticity, which plays a critical rule in its biochemical mechanism. Rabipur and Raz (2012) review the evidence of efficacy in cognitive training programs, which collectively demonstrate that the brain can be altered through behavioural intervention. Vinogradov, Fisher, and de Villers-Sidani (2012) describe the application of cognitive training techniques to treating neuropsychiatric illness, tying behavioural prognoses to physiological markers.

From the reference section of each review, author's names were mined by hand and entered into spreadsheets. It was originally planned to select the ten highest cited authors in each review, but there is no mathematically fair way to break ties for the last places, so lines were drawn based on the number of citations received by authors. The sample includes nine authors from the biology review, eleven from the psychology review, and thirteen from the medicine review for a total of 33 authors ($n_{\mathrm{BIO}} = 9$, $n_{\mathrm{PSY}} = 13$, $n_{\mathrm{MED}} = 11$, $N = 33$). Table 1 lists all of the authors in the sample. There are some limitations with this sampling method that were not problematic, but they should be mentioned here. Assuming that neuroscience is a highly interdisciplinary field, it is possible that an author could be highly cited in more than one of these reviews. This was not a problem of the present study, but some measure would have to be taken to accommodate such an author. Su and colleagues (2009) have developed a form of ACA called the “complete author pair” algorithm that allegedly allows authors in the sample to be categorized into multiple disciplines. The technique is specifically designed to identify authors with multiple expertise. The present study is not modelled after their algorithm. Another issue is that of the six authors of these three reviews, five of them were cited highly enough to be included in my sample. This is because authors tend to cite their own work liberally as a means (whether purposefully done or not) to inflate their citation counts (Andrés, 2009). However, because it takes a lot of knowledge on a topic to write a publishable scholarly review, this issue was not considered a problem of the study.

Table 1.
Surames and initials of sampled authors by WoS subject category.

Biochemistry & Molecular Biology	**Psychology**	**Psychiatry; Pharmacology & Pharmacy**
Mattson MP	Posner MI	Merzenich MM
Lu C	Rothbart MK	Vinogradov S
Begley JG	Klingberg T	Mahncke HW
Chan SL	Bialystok E	Fisher M
Cheng B	Davidson RJ	de Villers-Sidani E
Fu W	Raz A	Gazzaley A
Hollenbeck PJ	Westerberg H	Bao S
Mark RJ	Gaser C	Javitt DC
Sweatt JD	Rueda MR	Jenkins WM
	Slagter HA	Nagarajan SS
	Tennstedt SL	Ungerleider LG
	Walkup JT	
	Swanson J	

A more important limitation is author disambiguation. This is perhaps the most pressing concern of ACA because it is so common and very costly to analyze and fix. Some bibliometricians have developed algorithms that can test for the presence of author disambiguation in a citation index (D’Angelo, Giuffrida, & Abramo, 2011; Tan, Kan, & Lee 2006) but these algorithms report many false negatives and false positives. The main problem with author disambiguation is that in WoS, there is no reliable way to ensure that a queried name is not shared by multiple authors. WoS does not keep authority records on authors, which makes it very difficult and time-consuming to know if two or more authors in the citation index use the exact same name. To further complicate the matter, sometimes authors will legally change their name, for example after marriage. But without knowledge of the author's personal life, nobody would think to check for previously used names. Limiting by institution is not a reliable solution because some authors change their institutional affiliation during their publishing career. Even worse, sometimes authors will publish by different variants of their name (e.g. only using middle initials some of the time), which further complicates the issue. My first hypothesis is related to author disambiguation: while querying by surname and first initial will include citations from all authors who share this surname and first initial (which could be several authors), this is more favourable than querying by surname and all initials because records by the queried author that do not include his or her middle initials will not be retrieved. In essence, it is better to have noisy data than missing data. My second hypothesis is also related to this problem: it is relatively unlikely that two authors with the same surname and first initial will study the same discipline, so limiting queries by WoS subject category was expected to reduce some or all of the noise associated with author disambiguation.

Collection of Data

To test the viability of ACA for measuring interdisciplinarity, four MDS maps were generated from three data sets to make three comparisons. All co-citation data was collected by querying authors in WoS Cited Reference Search and entered into spreadsheets. For the first matrix, authors were queried by their surname and first initial only. For the second matrix, authors were queried by their surname, first initial, and any known middle initials. For the third matrix, authors were queried by their surname and first initial only (as with the first matrix), and the results were limited to the five WoS subject categories to which the selected review articles were categorized (Neurosciences, Biochemistry & Molecular Biology, Psychology, Psychiatry, and Pharmacology & Pharmacy).

Once every author was queried to retrieve all citations in WoS, the search history page was used to combine each possible pair of queries with the AND operator. The resulting co-citation counts were recorded in a data matrix of 528 co-citation values and 33 individual citation values down the axis of symmetry. This was all done manually with the WoS graphical user interface (GUI), and every manual step was double checked to ensure that human error would not affect the data.

Normalization of Data

The raw co-citation matrix is a record of absolute co-citation values. However, these values are meaningless unless they are converted into relative terms (Kreuzman, 2001; Leydesdorff & Vaughan, 2006; White, 2003). The Pearson's correlation coefficient is often used for this purpose (Kreuzman, 2001; Su *et al.*, 2009), but White (2003) contends that Person's *r* creates instability in matrices with large blocks of zero-values. His contention caused a debate in the bibliometrics community (White, 2004). A safe alternative is to use Jaccard's Index (Leydesdorff & Vaughan, 2006). For this study, all matrices (aside from the non-normalized condition) were normalized using Jaccard's Index. The mathematical formula for calculating this coefficient is:

$$J = C / (A + B - C)$$

Where for any given pair of authors, *A* represents the citations received by author A, *B* represents the citations received by author B, and *C* represents the times that authors A and B are co-cited by the same article. This can also be expressed through Boolean algebra:

$$J = (A \text{ AND } B) \;/\; (A \text{ XOR } B)$$

Where the numerator represents the times when authors A and B are co-cited by an article and the denominator represents the times when authors A and B are cited but not by the same article. There is a limitation associated with Jaccard's Index that is caused by author disambiguation. If author citations are overestimated by the citation index but raw co-citations are more accurate because it is unlikely that two authors with the same name would be co-cited with any given other author (White, 1990), then Jaccard’s values underestimate co-citations because their denominators are inflated. This is another purpose for my second hypothesis: if limiting queries by WoS subject category reduces the noise caused by author disambiguation, then the normalization by Jaccard’s Index will not underestimate co-citations.

Generation of MDS Map

The co-citation matrices will be imported into SPSS and subjected to the MDS test by selecting from the drop-down menu:

```
Analyze > Scale > Multidimensional Scaling (PROXICAL)
```

The map was edited in the SPSS output file using the built-in Chart Editor.

Interpretation of MDS Map

The data was analyzed using a multidimensional scaling (MDS) statistical test, which converts a symmetrical similarity matrix into a multidimensional plot (Kruskal, & Wish, 1978). MDS maps do not require axis scales because distances between plotted points are relative. The test is subject to a margin of error in terms of the goodness of fit between all of the points. The MDS map is a tool used to visualize data and as such is meant to be interpreted by a human

(Kruskal, & Wish, 1978). Therefore, clusters were drawn on the maps to aid in the visualization of disciplinarity and interdisciplinarity. Clusters represent the disciplines of authors in the map and on their relative locations.

Results

Data Matrices

Table 2 shows the average percent decreases between two pairs of co-citation matrices. Querying by surname, first initial, and all middle initials decreased the citation counts reported by WoS by 27% on average, but co-citation counts apparently decreased by -85%. This was an unexpected result that was either caused by erroneous collection of data or by a problem of data transmission from WoS to my computer terminal. Limiting queries to relevant subject categories decreased citations by 64%, and co-citations by 20% on average. This result shows that WoS subject categories greatly reduce noise in citation counts, which should improve the reliability of normalization by Jaccard's Index. Co-citation reductions are less likely to be reductions in noise (White, 1990) but probably suggest that these authors publish in journals that fall outside the limiting WoS subject categories.

Table 2.
Average % decreases between compared matrices.

Comparison	**Citations**	**Co-Citations**
Initials	27%	-85%
Limits	64%	20%

MDS Maps

The four MDS maps were constructed by inputting the matrices into SPSS and analyzing by multidimensional scaling (PROXICAL), which was set to treat the matrices as symmetrical, and the co-citation values as similarities. Goodness of fit values (called 'Stress-I' in the SPSS output file) are included in the figure descriptions.

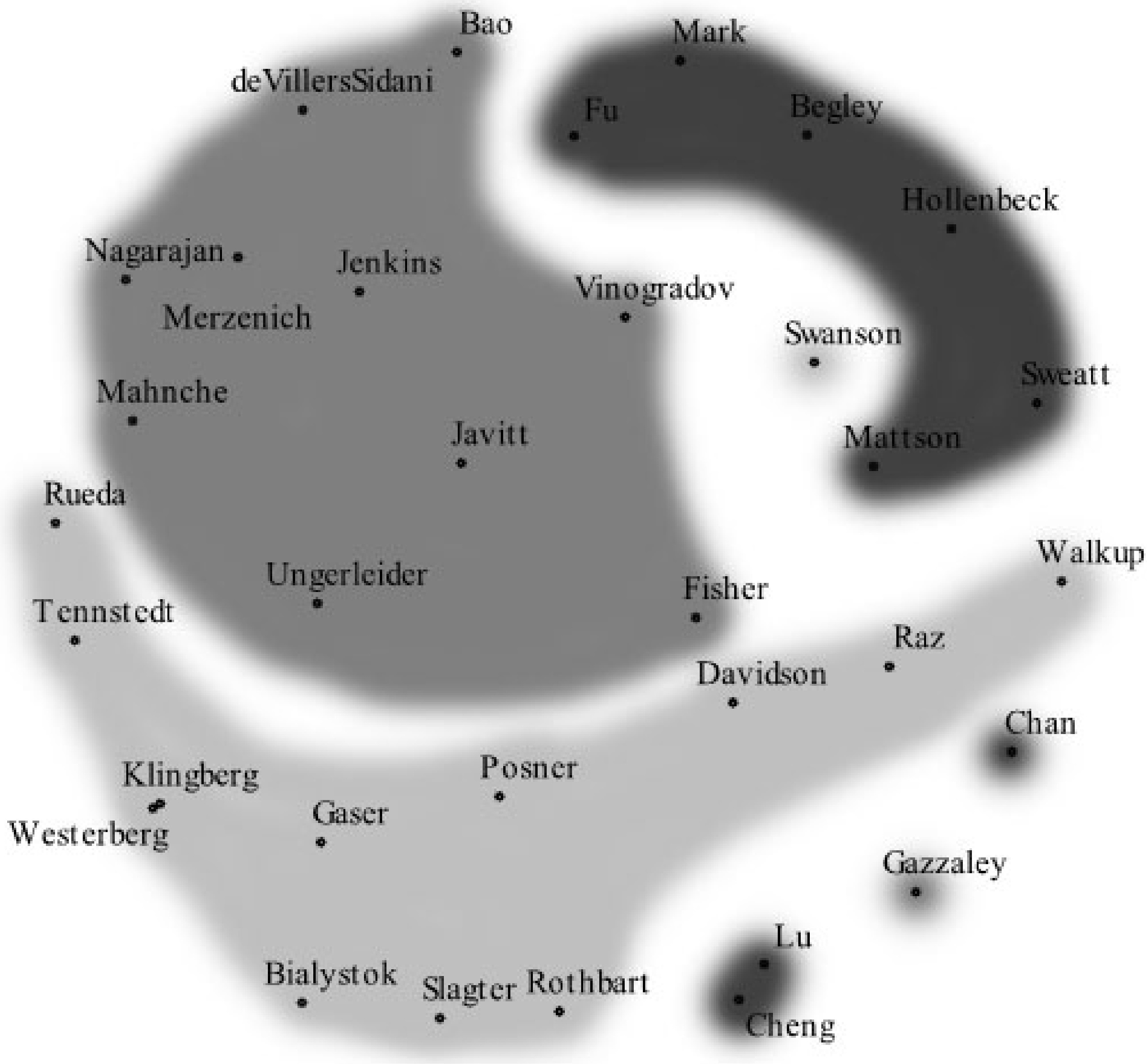


Figure 1: The authors were queried in WoS Cited Reference Search by surname and first initial only. No subject category limits. Similarity values were normalized with Jaccard's Index. Disciplinary clusters are very large. Stress-I = 0.38.

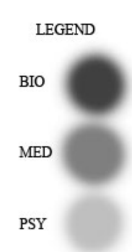


Figure 2. The authors were queried in WoS Cited Reference Search by surname, first initial, and all known initials. No subject category limits. Similarity values were normalized with Jaccard's Index. Stress-I = 0.39.

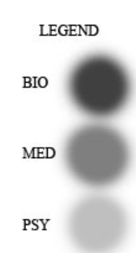


Figure 3.The authors were queried in WoS Cited Reference Search by surname and first initial only. No subject category limits. Similarity values were not normalized. Stress-I = 0.38.

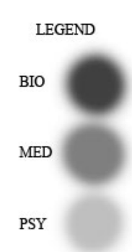
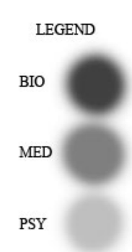


Figure 4. The authors were queried in WoS Cited Reference Search by surname and first initial only. Results were limited to the five relevant subject categories. Similarity values were normalized with Jaccard's Index. Stress-I = 0.39.



Cluster Map Comparisons

Clusters were identified based on the following criteria: a) the authors in any given cluster are in the same discipline (i.e. mined from the same review); b) a line can be drawn between the two authors without coming close to an author from a different discipline; and c) the authors are relatively close to each other on the map. Table 3 presents the number of solo and paired disciplinary clusters identified in the MDS maps. A greater proportion of authors alone or in pairs would suggest that their co-citations are more interdisciplinary than if the authors were found in large disciplinary clusters.

Table 3.
Number of lone or paired authors in cluster map.

Map	Solo	Paired	Total	% of Sample
First Initial	3	2	5	15
All Initials	2	2	4	12
No Normalization	3	2	5	15
With Limits	6	4	10	30

Discussion

First I will discuss the comparisons between the MDS maps. Then I will discuss the limitations of the study and how they affected the method and results of the present study. Finally, I will discuss some issues that I encountered while using WoS Advanced Search to find reviews for sampling and Cited Reference Search to build the co-citation matrices.

First Initial vs. All Initials

The map of authors queried by surname and first initial (Figure 1) is very different from the map of authors queried by surname, first initial, and all middle initials (Figure 2). While some of the noise was reduced, it is unclear exactly how much because some relevant citations are also removed (those from articles where the authors with middle names were published without their middle names). There are 19 authors in the sample who have a middle name: Mattson, Begley, Chan, Hollenbeck, Mark, Sweatt, Posner, Rothbart, Davidson, Rueda, Slagter, Tennstedt, Walkup, Merzenich, Mahncke, Javitt, Jenkins, Nagarajan, and Ungerleider (see Table 1); the rest have no middle name. In Figure 2, the big cluster of medical researchers is broken by Tennstedt and Raz, two psychologists who are far from one another in Figure 1. The biology cluster of Cheng and Lu from Figure 1 includes Sweatt in Figure 2, who unlike the two of them has a middle name. Hollenbeck, a biologist with a middle name is now found behind the large cluster of psychologists in Figure 2 instead of in the middle of the big biology cluster in Figure 1. Fisher, who has no middle name, was at the lowest part of the big medical cluster in Figure 1, but is found in the upper medicine cluster in Figure 2.

Authors who don't have middle initials should, theoretically, not be greatly affected, because their citation counts are the same in both matrices. This was not always the case though. Raz has no middle name and yet moved next to Tennstedt, but this is perhaps because the strength of their co-citations was increased when middle initials were queried. However, authors who use middle names in their publications might not have used their middle names in every article that they have published, so their citation counts may not include all of their articles. This complicates the relationships and suggests that Figure 2 is not a reliable source of data. Including middle initials in author queries will reduce noise from alternate names, i.e. authors with the same surname and first initial, of which there could be several (White, 1990). Therefore, there is a trade-off that while querying by all initials should reduce noise when the queried authors have middle initials, it is possible that some articles would be missed of those journals do not print the middle initials of the authors. Therefore, author name queries including middle initials cannot be trusted to yield representative results.

First Initial vs. First Initial No Normalization

The difference between normalized (Figure 1) and prenormalized data (Figure 3) is very small. The larger disciplinary clusters take almost the same shape and have the same orientation with respect to the others. In Figure 1, I identified Swanson, Chan, and Gazzaley as solo while Lu and Cheng were the only pair. In Figure 1, Chan is a solo biologist found behind the psychology cluster and Fu is in the big biology cluster, but in Figure 3, Chan is in the big biology cluster and Fu is found alone, but close to the medicine cluster. Also, Figure 1, the points representing Klingberg and Westerberg are almost touching, but they are further away in the Figure 3. The relative positions of Tennstedt and Rueda in the large psychology cluster are switched, as are those of Gaser and Klingberg. In terms of similarities, Gazalley, Lu, and Cheng are in the lower right quadrant of both Figures 1 and 3, behind the psychology cluster. The psychology cluster is unbroken in both figures, but in Figure 3, Davidson, Raz, and Walkup are far enough from the other psychologists that they could be given their own cluster.

Generally speaking, these two maps are almost the same, but several minor differences are noticeable and are probably attributed to the underestimation of co-citation values after normalization with Jaccard’s Index. This problem would only affect authors whom share their surname and first initial with many other highly cited authors because it is overestimation of individual citation counts that causes Jaccard’s Index to underestimate co-citation values.

First Initial vs. First Initial With Limits

There is significantly less noise when limiting queries by WoS subject categories (Figure 4) compared to querying without limits (Figure 1). Limiting only to the relevant subject categories decreases the size of disciplinary clusters and increases the number of them: I identified six clusters in Figure 1, and twelve clusters in Figure 4. Almost one third of the authors are either alone or in paired clusters, which is at least twice as many as the other figures show (see Table 3). In Figure 4, the only large cluster on the map contains nine psychologists, while Figure 1 has twelve psychologists in one cluster. Swanson was the only lone psychologist in Figure 1, wedged between the big biology and medicine clusters. Davidson and Raz are very close to Walkup in Figure 1, but in Figure 4 Walkup is found close to Gaser. The large medical cluster includes Gazzaley in Figure 4, but Fisher, Nagarajan, Vinogradov, Ungerleider, and Bao have separated from it to form four more clusters. The large biology cluster is reduced to five authors in Figure 4 (ironically including Lu and Cheng whom are in a paired cluster in Figure 1), and a second cluster emerges on the opposite side of the map.

Limiting by relevant WoS subject categories should greatly reduce the noise caused by author disambiguation (assuming that all homonyms are of authors from different disciplines, which cannot be guaranteed), but it could also eliminate citations from papers that fall outside of the relevant subject categories. The comparison of Figures 1 and 4 is the most interesting because disciplinary clusters become smaller and more oblong, and the mixture of disciplines becomes more homogeneous. The citations lost from limiting by subject categories are likely not relevant to this map of neuroplasticity theory, so their absence should not be taken as any significant loss. Of the three methodological comparisons I made in the present study, limiting by subject category shows the most promise as a measure of interdisciplinarity. I encourage future research into this method of querying for co-citations.

Limitations

Author Name Disambiguation

The problem of how to query authors' names is difficult to solve because it is based on centuries of scholarly journals publishing without any standardization of author names or mechanism of authority control (D'Angelo, Giuffrida, & Abramo, 2011). Despite the decades of hard work made by ISI on the WoS architecture, the problem is, at this point, too large to fix. The Author Search feature lists name variants, and has built in options to limit by institution or by research area. It would be ideal if there was an internationally recognized identifier for authors, like an ID number, but this would be impossible to implement because every academic institution in the world would have to participate, and not every published author is affiliated with a traditional academic institution. And what to do about independent scholars? It would especially be difficult to identify deceased authors, and to know at what point in history to draw the line. There are some citation indices that keep authority records on authors in order to prevent name disambiguation, however these databases are often limited by nation (D'Angelo, Giuffrida, & Abramo, 2011) or by field of study (Morooka, Ramos, & Nathaniel, 2014; White, Wellman, & Nazer, 2004), making them much easier to maintain and use than WoS, Scopus, or Google Scholar.

There are many other problems with author names. There is no way to account for changes in name, for example, by marriage. Errors in data entry can make it impossible to retrieve data on some records; these are referred to as orphan records. Non-English name conventions are especially complicated. Names with non-standard characters like accentuated vowels are not always entered correctly. The Chinese convention is to put the family name before given names, so Westerners may have trouble knowing which name is the surname, and there is no standard convention for Chinese people with multiple characters in their given name so the second character could be confused with a middle name. Sometimes two given names are connected by a hyphen, and other times there is no space between them. The problem of author disambiguation does not have a clear solution that will eliminate all possibility of erroneous data collection.

I found that limiting by WoS subject category reduced the noise of citation counts by a considerable amount, suggesting that not limiting is likely to produce noisy citation counts. This creates a problem when normalizing co-citation counts with Jaccard's Index because those values will be underestimates of the true value, and each of those values will be underestimated differently based on how citation counts were overestimated by noise.

Data Transfer and Processing Rate

Data was collected manually using the GUI in the current Web version of WoS. When WoS was first released electronically via the Internet, data was retrieved through dialog (Bruer, 2010; White, & McCain, 1998). The GUI of Cited Reference Search is typical of any contemporary search engine or Web directory. WoS saves search queries during the user's session, which made the retrieval of co-citation data much easier. While this subjects the collection to human error, I was careful to double-check all manual steps. It was imperative that every dynamic Web page was given the chance to load completely because requesting new Web pages faster than WoS could process requests led to errors in retrieval. This problem was discovered during data collection, but was corrected when gathering the data presented above.

Whether data is collected by algorithm or by hand, having a human check the work will always help to reduce errors, as algorithms can make mistakes too. Mining data by hand means that less data can be collected in any given amount of time, however this is offset by the fact that data does not need to be cleaned manually later because it can be cleaned by the miner as it is gathered.

Goodness of Fit

The goodness of fit values of all MDS maps were comparable, and much higher than the maximum threshold of 0.05. This presents a problem with the reliability of the results. Obviously, it is unlikely that a perfect map of distances could be drawn, so some margin of error must be tolerated, but a value approaching 0.4 is perhaps beyond tolerance. It is possible that there is no way to improve the goodness of fit for this test, but unfortunately, bibliometricians often do not report their goodness of fit values, so it is impossible to compare my results to those of other ACA studies. A future analysis could focus entirely on the minimization of the error associated with goodness of fit.

Defining Clusters

This was a challenge that I could not adequately meet due to time constraints. Because distances between points are relative, it is difficult to decide at what distance to draw the line. In the end, I took a liberal approach to clustering. Previous researchers would define clusters based on their familiarity with the authors in their sample (Kreuzman, 2001; Lu, & Wolfram, 2012; White, & Griffith, 1981), and if not for time constraints, I would qualitatively analyze the faculty Web pages of the authors in my sample to determine what topic-based clusters they would fall under and compare that information to the clusters generated by the MDS maps.

Complications with Sampling Philosophers

I originally planned to have a forth group of authors in the sample to represent the discipline of philosophy. Unfortunately, it was impossible to find a good article on neuroplasticity under the WoS subject category of ‘Philosophy.’ Perhaps philosophers do not contribute significantly to neuroplasticity theory. What is more likely is that their work on this topic is found in books and essays, which are not indexed in WoS. It is well known in LIS that not all disciplines are alike in citation patterns (Andrés, 2009; Currie & Monroe-Gulick, 2014; Hunt, 2009; Klein, 1990), so perhaps scholarly reviews are not a suitable place to mine humanities authors. Kreuzman (2001) chose to mine philosopher names from books, so a future study might find a way to incorporate books into the author selection process for the sake of including humanities researchers in the analysis.

Web of Science Search Features

Advanced Search

In the interest of saving time, the sample of authors was selected from the citations of scholarly reviews on the topic 'neuroplasticity.' This topic was selected because it is well known to have relevance in biology (Nichols *et al*., 2001), psychology (Kolb, & Whishaw, 2008), and medicine (Julien *et al.*, 2011). To select review articles, it is permissible to search the content of an article to identify its topic because the title, abstract, and keyword portions of an article are written by the authors in the interest of presenting the most information in the smallest number of words. There is no controlled vocabulary for the keywords (which are selected by the authors anyway), so they cannot be the only source for identification of an article's topic.

WoS subject categories were used to limit article searches to discipline. This method does not work as accurately for an author co-citation analysis as it would for an article co-citation analysis because articles are clearly divided into categories based on the associated journal, but authors can publish in journals on any given topic, making it difficult to categorize authors by the same method (Levitt, Thelwall, & Oppenheim, 2011). However, the highest-cited authors were selected from reviews, and these authors are the most likely to have a significant interest and specialization on the topic of neuroplasticity.

Human error associated with my methodology was addressed through manual filtering and cleaning of the referenced author lists and co-citation matrices. All manual steps were double-checked for correctness, and all names were recorded in the same format (e.g. Posner M or Posner MI).

Cited Reference Search

Because an author query by surname and all initials is narrower than a query by surname and first initial only, it is impossible that co-citations would be decreased by -85%. In hindsight, there were problems with my first few trials of data collection. The GUI in WoS is expressed by a series of dynamic Web pages. After repeating certain trials multiple times, I noticed that sometimes the exact same query would yield different results. I cannot be certain if any of the results were the cause of the WoS architecture, but at least part of the problem was with the amount of time that is required to process the large amounts of data that I was requesting. After several practice trials, I realized that rushing through Web pages interrupted my queries and caused them to retrieve only a fraction of the citations that I requested. During my final data collection trials, I made sure to let every Web page load completely before calling the next Web page. To explain the negative average decrease in co-citation counts from Table 2, it is possible that the co-citation matrix of queries by surname and all initials was not collected correctly by not allowing WoS Web pages to load completely, but due to time limitations I was not able to collect the data again and compare it to what is presented in Figure 2.

Conclusion

This introduction to bibliometrics, interdisciplinarity, and multidimensional scaling was very useful to me as an educational exercise. The potential for multidimensional scaling analyses to visualize co-occurrence matrices is very powerful, albeit a little overwhelming.

Bibliometricians who study interdisciplinarity have already been using these techniques for decades, but it is still challenging to take full advantage of their capability. In my humble opinion, the traditional way of enclosing clusters within shapes started by White and Griffith (1981) could use an update. I chose to shade my MDS maps rather than draw shapes around the points to make it easier to visualize the different disciplines and how their authors are interrelated by the map. In a future study, these two clustering techniques could be combined to add another layer of complexity to the analysis without overwhelming the viewer. The shading of disciplinary clusters could be combined with enclosing topic clusters within shapes to show two levels of clustering and how they interact to form multidisciplinary and interdisciplinary relationships between authors. It would be interesting to incorporate philosophy and computer science into the sample, and to include the former discipline I would consider using a book or an essay to mine authors. If I were to do this study again, that is how I would attempt to build upon the present body of research in this field.

References

Andrés, A. (2009). *Measuring Academic Research: How to Conduct a Bibliometric Study*. Oxford: Chandos.

Askensay, J., & Lehmann, J. (2013). Consciousness, brain, neuroplasticity. *Frontiers in Psychology, 4*, 412.

Baumwol, K., Mortimer, S. T., Huerta, T. R., Norman, C. D., & Buchan, A. M. J. (2011). Promoting interdisciplinarity in the life sciences: a case study. *Research Evaluation, 20*(4), 283-292.

Bjurström, A., & Polk, M. (2011). Climate change and interdisciplinarity: A co-citation analysis of IPCC Third Assessment Report. *Scientometrics, 87*, 525-550.

Boyack, K. W., Klavans, R., & Börner, K. (2005). Mapping the backbone of science. *Scientometrics, 64*(3), 351-374.

Braun, T., Glänzel, W., & Schubert, A. (2001). Publication and cooperation patterns of the authors of neuroscience journals. *Scientometrics, 51*(3), 499-510.

Bruer, J. T. (2010). Can we talk? How the cognitive neuroscience of attention emerged from neurobiology and psychology, 1980-2005. *Scientometrics, 83*(3), 751-764.

Bush, G. W. Sr. (1990). *Decade of the Brain: Presidential Proclamation 6158 (Library of Congress)*. Retrieved from http://www.loc.gov/loc/brain/proclaim.html

Currie, L., & Monroe-Gulick, A. (2013). What do our faculty use? An interdisciplinary citation analysis study. *The Journal of Academic Librarianship, 39*, 471-480.

D'Angelo, C. A., Giuffrida, C., & Abramo, G. (2011). A heuristic approach to author name disambiguation in bibliometrics databases for large-scale research assessments. *Journal of the American Society for Information Science and Technology, 62*(2), 257-269.

Garfield, E., & Sher, I. H. (1963). New factors in the evaluation of scientific literature through citation indexing. *American Documentation, 14*(3), 195-201.

GitHub. (2014). About: GitHub. Retrieved from https://github.com/about/

Gowanlock, M., & Gazan, R. (2013). Assessing researcher interdisciplinarity: A case study of the University of Hawaii NASA Astrobiology Institute. *Scientometrics, 94*(1), 133-161.

Haythornthwaite, C. (2006). Learning and knowledge networks in interdisciplinary collaborations. *Journal of the American Society for Information Science and Technology, 57*(8), 1079-1092.

Hérubel, J.-P. V. M. (2010). Being *undisciplined*; or traversing disciplinary configurations in social science and humanities databases: conceptual considerations for interdisciplinartiy and multidisciplinarity. In *Social Science Libraries: Interdisciplinary Collections, Services, Networks,* Witt, S. W., & Rudasill, L. M. (Eds.). pp. 25-39. International Federation of Library Associations Publications 144: The Hague.

Huang, M.-H., & Chang, Y.-W. (2011). A study of interdisciplinarity in information science: using direct citation and co-authorship analysis. *Journal of Information Science, 37*(4), 369-378.

Hunt, G. E. (2011). Making sense of bibliometrics. *Acta Neuropsychiatrica, 23*(2), 80-81.

Institute for Scientific Information. (2014). *Science Citation Index Expanded: IP & Science: Thompson Reuters*. Retrieved from http://ip-science.thomsonreuters.com/mjl/scope/scope_scie/

Jacoby, J.A. (2010). Share and share alike? Data-sharing practices in different disciplinary domains. In *Social Science Libraries: Interdisciplinary Collections, Services, Networks,* Witt, S. W., & Rudasill, L. M. (Eds.). pp. 53-61. International Federation of Library Associations Publications 144: The Hague.

Julien, R. M., Advokat, C. D., Comaty, J. E. (2011). *A Primer of Drug Action* (12th ed.). Worth Publishers: New York.

Klein, J. T. (1990). *Interdisciplinarity: History, Theory, & Practice*. Wayne State University Press: Detroit.

Kolb, B., & Whishaw, I. Q. (2008). *Fundamentals of Human Neuropsychology* (6th ed.). Worth Publishers: New York.

Koslow, S. H. (2000). Should the neuroscience community make a paradigm shift to sharing primary data? *Nature Neuroscience, 3*(9), 863-865.

Koslow, S. H. (2002). Sharing primary data: A threat or asset to discovery? *Nature Reviews Neuroscience, 3*(4), 311-313.

Kreuzman, H. (2001). A co-citation analysis of representative authors in philosophy: Examining the relationship between epistemologists and philosophers of science.

Kruskal, J. B., & Wish, M (1978). *Multidimensional Scaling*. Sage Publications: Beverly Hills.

Kushkowski, J. D., & Shrader, C. B. (2013). Developing a core list of journals in an interdisciplinary area: Implications for academic faculty and librarians demonstrated in the business literature. *Library Resources & Technical Services, 51*(1), 51-65.

Lee, J. D. (2005). Bibliometric analysis of *Human Factors* (1970-2000): A qualitative description of scientific impact. *Human Factors, 47*(4), 753-766.

Levitt, J. M., Thelwall, M., & Oppenheim, C. (2011). Variations between subjects in the extent to which the social sciences have become more interdisciplinary. *Journal for the American Society of Information Science and Technology, 62*(6), 1118-1129.

Leydesdorff, L., & Rafols, I. (2011). Indicators of the interdisciplinarity of journals: Diversity, centrality, and citations. *Journal of Infometrics, 5*(1), 87-100.

Leydesdorff, L., & Vaughan, L. (2006). Co-occurrence matrices and their applications in information science: Extending ACA to the web environment. *Journal of the American Society for Information Science and Technology, 57*(12), 1616-1628.

Locke, J. (1690). *An Essay Concerning Humane Understanding*. Oxford University Press: Oxford.

Lu, K., & Wolfram, D. (2012). Measuring author research relatedness: A comparison of word-based, topic-based, and author cocitation approaches. *Journal of the American Society for Information Science and Technology, 63*(10), 1973-1986.

Mattson, M. P. (2007). Mitochondrial regulation of neuronal plasticity. *Neurochemistry Reviews, 32*(4), 707-715.

Morooka, K., Ramos, M. M., & Nathaniel, F. N. (2014). A bibliometric approach to interdisciplinarity in Japanese rice research and technology development. *Scientometrics, 98*(1), 73-98.

National Human Genome Research Institute. (NHGRI). (2014). *All About the Human Genome Project (HGP)*. Retrieved from http://www.genome.gov/10001772

Neylon, T. (2014). *The Cost of Knowledge*. Retrieved from http://thecostofknowledge.com/

Nicholls, J. G., Martin, A. R., Wallace, B. G., & Fuchs, P. A. (2001). *From Neuron to Brain* (4th ed.). Sinauer Associates: Sunderland.

Palmer, C. L. (2010). Information research on interdisciplinarity. In Frodeman, R., Klein, J. T., & Mitcham, C. (Eds.). *The Oxford Handbook of Interdisciplinarity* (174-188). New York: Oxford University Press.

Palmer, C. L., Cragin, M. H., & Hogan, T. P. (2007). Weak information work in scientific discovery. *Information Processing & Management, 43*(3), 808-820.

Pearn, J. (2014). *Blue Brain Project*. Retrieved from http://www.artificialbrains.com/blue-brain-project

Price, D. J. S. (1965). Networks of scientific papers: The pattern of bibliographic references indicates the nature of the scientific research front. *Science, 149*(3683), 510-515.

Rabipour, S., &Raz, A. (2012). Training the brain: Fact and fad in cognitive and behavioral remediation. *Brain and Cognition, 79*(2), 159-179.

Schwechheimer, H., & Winterhager, M. (2001). Mapping interdisciplinary research fronts in neuroscience: A bibliometric view to retrograde amnesia. *Scientometrics, 51*(1), 311-318.

Sengupta, I. N. (1989). The growth of knowledge and literature in neuroscience. *Scientometrics, 17*(4), 253-288.

Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science, 24*(4), 265-269.

Small, H., & Griffith, B. C. (1974). The structure of scientific literatures I: Identifying and graphing specialties. *Science Studies, 4*, 17-40.

Society for Neuroscience. (2014). *Neuroscience 2014*. Retrieved from http://www.sfn.org/Annual-Meeting/Neuroscience-2014

Su, Y.-M., Yang, S.-C., Hsu, P.-Y., & Shiau, W.-L. (2009). Extending co-citation analysis to discover authors with multiple expertise. *Expert Systems with Applications, 36*(3), 4287-4295.

Tan, Y. F., Kan, M.-Y., & Lee, D. (2006). Search engine driven author disambiguation. *Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries, 2006*, 314-315.

Vinogradov, S., Fisher, M., & de Villers-Sidani, E. (2012). Cognitive training for impaired neuroal systems in neuropsychiatric illness.

Waters, W. H. (1885). Some vaso-motor functions of the spinal nerves in the frog. *Journal of Physiology*, *6*(6) 460-463.

Weingart, P. (2010). A short history of knowledge formations. In Frodeman, R., Klein, J. T., & Mitcham, C. (Eds.). *The Oxford Handbook of Interdisciplinarity* (3-14). New York: Oxford University Press.

White, H. D. (1990). Author cocitation analysis: Overview and defense. In *Bibliometrics and Scholarly Communication*, Borgman, C. (Ed.). pp. 84-106. Sage: Newbury Park.

White, H. D. (2003). Author cocitation analysis and Pearson's *r*. *Journal of the American Society for Information Science and Technology, 54*(13), 1250-1259.

White, H. D. (2004). Author cocitation analysis and Pearson's *r*: Replies and a correction. *Journal of the American Society for Information Science and Technology, 55*(9), 843-844.

White, H. D., & Griffith, B. C. (1981). Author co-citation: A literature measure of intellectual structure. *Journal of the American Society for Information Science, 32*(3), 163-171.

White, H. D., & McCain, K. W. (1998). Visualizing a discipline: An author co-citation analysis of information science, 1972-1995. *Journal of the American Society for Information Science and Technology, 49*(4), 327-355.

White, H. D., Wellman, B., & Nazer, N. (2003). Does citation reflect social structure? Longitudinal evidence from the "Globenet" interdisciplinary research group. *Journal of the American Society for Information Science and Technology*, *55*(2), 111-126.