CrossMark

Creating a Self-Plagiarism Research Topic Typology through Bibliometric Visualisation

Peter Kokol[1] · Jernej Završnik[2] · Danica Železnik[3] ·
Helena Blažun Vošner[4]

Abstract Self-plagiarism, textual recycling and redundancy seemed to be controversial and unethical; however some questions about its definition are still open. The objective in this paper presented study was to use bibliometric analysis to synthesise and visualize the research literature production and derive a typology of self-plagiarism research. Five topics emerged: Self-plagiarism, Institutional self-plagiarism, Self-plagiarism and ICT, Self-plagiarism in academic writing, Self-plagiarism in science. The state of the art topics seem to be "social medium", "virtual world", "face book", "sociomateriality", "knowledge sharing", "open access", "institutional repository", "retraction" and "responsible conduct".

Keywords Self – plagiarism · Bibliometrics · Bibliometric mapping · Science landscapes

✉ Peter Kokol
peter.kokol@um.si

Jernej Završnik
jernej.zavrsnik@zd-mb.si

Danica Železnik
zeleznik.danica@gmail.com

Helena Blažun Vošner
helena.blazun@um.si

[1] Faculty of Electrical Engineering and Computer Science, University of Maribor, Smetanova 17, 2000 Maribor, Slovenia

[2] Dr. Adolf Drolc Healthcare Center, Ulica talcev 9, 2000 Maribor, Slovenia

[3] University College of Health Sciences Slovenj Gradec, Glavni trg 1, p.p. 144, 2380 Slovenj Gradec, Slovenia

[4] Faculty of Health Sciences, Center for International Cooperation, University of Maribor, Žitna 15, 2000 Maribor, Slovenia

⌂ Springer

Introduction

Plagiarism has occurred historically as long as humans have been creating science and research (Maurer et al. 2006). A humorous, but quite concise motto defining plagiarism is, "If you steal from one author, it is plagiarism; if you steal from many, it is research" (Cunha 1946). A special case of plagiarism is self-plagiarism. In a study of 63 recent editorials on plagiarism found through PubMed, no consensus on the definition on self-plagiarism was reached (Roig 2014). Burton (2014) claims that "textual recycling" as a part of self-plagiarism is particularly controversial and unethical. The American Psychological Association (2010) defined self-plagiarism as "…refer[ring] to the practice of presenting one's own previously published work as though it were new". Their online survey, in which 277 editors of education journals expressed their opinion on self-plagiarism, showed that a substantial majority agreed with the APA definition. However, many of these editors felt unease with the general concept of self-plagiarism; some even noted that in limited cases self-plagiarism might be acceptable (Bruton and Rachal 2015). Redundancy in science is another issue associated with self-plagiarism: whether a conference paper later published in a journal is still an original contribution or not remains a contemporary ethical question yet to be answered (de Vasconcelos and Roig 2015; Robinson 2014). Since May 2013, student self-plagiarism (referred to as"double-dipping") has been defined by the US Department of Health and Human Services as: "…when a student submits a whole paper or a substantial portion of a paper to fulfil a course requirement even though that paper had earlier been submitted to satisfy the requirements for another course taught by another professor…" (US Department of Health and Human Services 2011). In addition, several major United States universities are addressing self-plagiarism in their academic honesty policies in ways similar to those above (Halupa 2014).

   The objective of the study presented in this paper was to analyse the scientific literature production dealing with self-plagiarism research, using bibliometric analysis. The aim was to identify the most productive countries, institutions and source titles and to visualize the chronological topic evolution and the most important topics in the form of scientific landscapes. In this manner we hope to provide additional insights into the evolution of self-plagiarism research and its state of the art.

Materials and Methods

Bibliometric analysis is a method for discovering empirical evidence about research literature production dynamics, the most productive countries, organisations and source titles, and for identifying hot research topics or patterns in the evolution of a scientific field (Pritchard 1969; Garfield 2006; De Bellis 2009). Lately, classic bibliometric methods have been extended with bibliometric mapping (Heersmink et al. 2011). Bibliometric mapping generates visual representations of the relations between various bibliometric units. In our case an open license software tool called VOSviewer (Leiden University, The Netherlands) (Van Eck and Waltman 2013) was used. The VOSviewer mapping technique is closely related to multidimensional scaling (Young and Hamer 2013). However, VOSviewer constructs a term map based on a terms co-occurrence matrix obtained from the text analysed (usually paper abstracts). This matrix is then normalized according to the total number of occurrences and co-occurrences of terms, using a similarity measure based on association strength. In the next step VOSviewer constructs a two-dimensional map in which the terms are located in such a way that the

distance between any pair of terms reflects their similarity (Van Eck and Waltman 2013). VOSviewer software induces term maps, sometimes called science landscapes. A science landscape reveals various associations between terms, such as the following:

- proximity between terms reflects their similarity – terms positioned near each other on the term map are more similar than terms positioned far away;
- highly associated terms are grouped in clusters indicated by the same colour;
- importance of a term (according to the number of occurrences in analysed paper abstracts) is indicated by the size of the terms' fonts.

For our analysis, the document corpus was retrieved from the Scopus bibliographic database (Elsevier, Netherlands), on February 26th, 2016 using the search string ((*auto OR self*) *AND plagiarism*) in the title, abstract and keywords. Owing to the limitations of the software used, the search was limited to information sources written in English. For each information source, nationality and the name of the author's institutional affiliation, year of publication, source title and abstract were extracted. The most prolific countries, organisations and source titles were identified by Scopus built-in services. The literature production dynamics graph was formed in Excel (Microsoft Office, Microsoft Corporation, USA). Abstracts were exported to VOSviewer and analysed using default VOSviewer parameters. The common words related to research methodology like *study*, *significance*, *case*, *objective* and *experiment* were removed from the analysis. Country co-authorship, chronological and cluster landscapes were created. Using the cluster scientific landscape, we created the self-plagiarism research topic typology. Based on representative terms found in the clusters, we assigned a research topic to each cluster using interpretive synthesis (Carroll and Booth 2015). For each topic from the typology, we identified the most frequently cited recent publications and summarized these in a scoping review.

Results and Discussion

A total of 313 information sources written in English during the period 1946 to 2015 were extracted. They were published by 49 countries, 168 institutions in 144 source titles (journals and conference proceedings). The most productive countries were the United States of America (USA) ($n = 84$, United Kingdom ($n = 30$), India ($n = 17$), China ($n = 16$), Australia ($n = 15$), Nigeria ($n = 8$), Canada ($n = 8$), Germany ($n = 8$), Thailand ($n = 8$), Switzerland ($n = 8$) and Malaysia ($N = 8$). The most productive institutions were St John's University ($n = 8$), Joseph Ayo Babalola University, Northern Illinois University and University of Malaya, with four published information sources each, and Wiwanitkit House, Hainan Medical University, University of Limerick, University of Chile, University of Michigan Ann Arbor, University of Malaya, University of Georgia and University of South Australia with three published information sources each. The most productive source titles were the Journal of Academic Ethics ($n = 7$), the European Journal of Clinical Investigation ($n = 5$), Science and Engineering Ethics ($n = 5$), American Journal of Roentgenology ($n = 4$), Communications of the ACM ($n = 4$) and Ethics and Behaviour ($n = 4$), followed by the Journal of Cognitive and Behavioral Psychotherapies, Research in Nursing and Health, the British Journal of Educational Technology, Communication in Computer and Information Sciences, and Biochemia Medic, all with three published information sources each. The large number of different countries, institutions

and source titles shows that research is widely dispersed, with no strong research centres and dedicated source titles having yet formed.

The cooperation between countries based on co-authorship is presented in Fig. 1. The USA is the most networked country, cooperating with six countries, followed by Serbia ($n = 5$) and Nigeria ($n = 4$). All of them are among the 15 most productive countries identified above, so it seems that strong cooperation has a positive effect on productivity and vice versa.

Trends in Research Literature Production and Evolution of Research Topics

As is evident from Fig. 2, the first paper on academic plagiarism was published in the year 1946 (Cunha 1946). Production was steady until 2002, followed by linear growth till the year 2006, when the exponential growth started. The literature production peak was reached in 2012, after which productivity decreased. This decrease came quite surprisingly, taking into account the many discussions about self-plagiarism. According to the literature production dynamics shown in Fig. 1, the development of self-plagiarism research can be divided into four phases (Shen et al. 2013):

- incubation phase in the period 1946–2002, with 19 information sources,
- initiation phase in the period 2003–2006, with 27 information sources,
- rapid growth phase in the period 2007–2012, with 145 information sources, and

Fig. 1 The cooperation between countries based on co - authorships
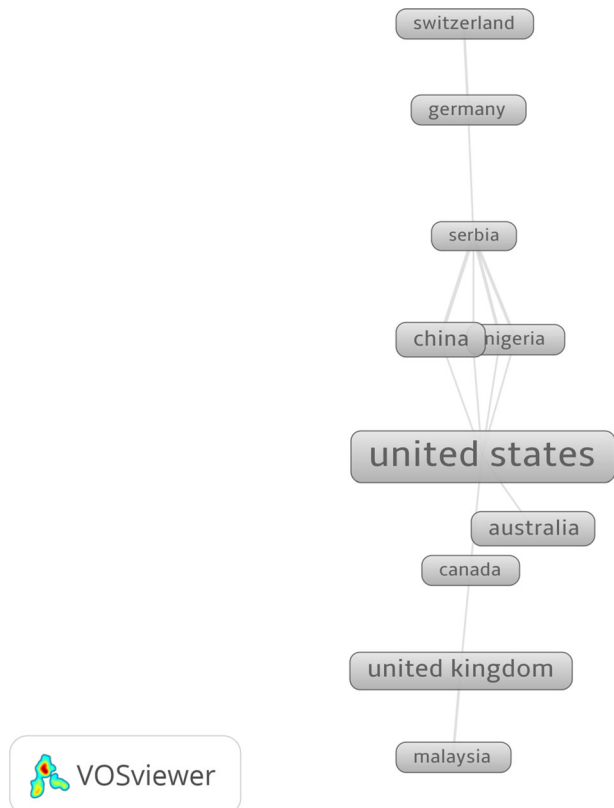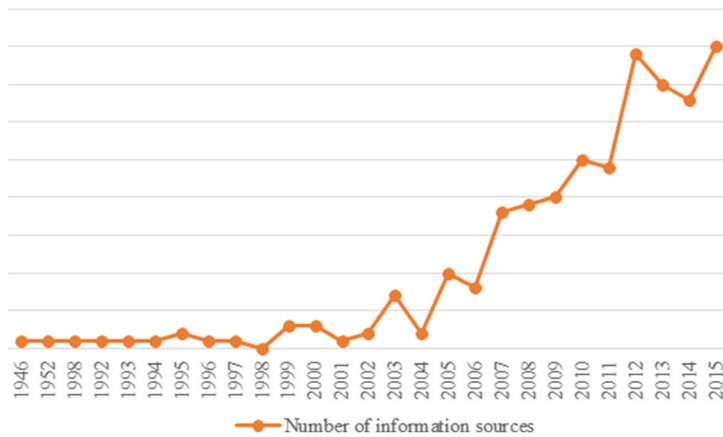
Fig. 2 The trends in literature production in self - plagiarism research

- attenuation phase between 2013 and 2015, with 108 information sources.

The evaluation of research terms is shown in Fig. 3. The most important research areas in the incubation and initiation topic were "unethical behaviour", "experience" and "cryptomnesia". The research at the beginning of rapid growth (period 2007–2009) was focused on research related to the terms "copyright", "fraud", "scientist", "judgement", "reference" and "misconduct". In the period 2009–2011 the research shifted to "self-plagiarism", "academic integrity", "adaption", "examination", "assessment", in combination with terms related to science and scientific publishing, including "science", "researcher, "author", "authorship", "publication", "fabrication, "breach" and "scientific misconduct". The research was also concerned with plagiarism detection approaches like "self-organising maps". Recent research (attenuation phase) is related to information communication technology (ICT) like "social medium", "virtual world", "Facebook", "sociomateriality", "knowledge sharing", "open access", "dissertation", "institutional repository", "retraction", "research ethics", "responsible conduct" and "AMS (American Meteorology Association) policy". The research is also concerned with plagiarism detection approaches like "self-organising maps".
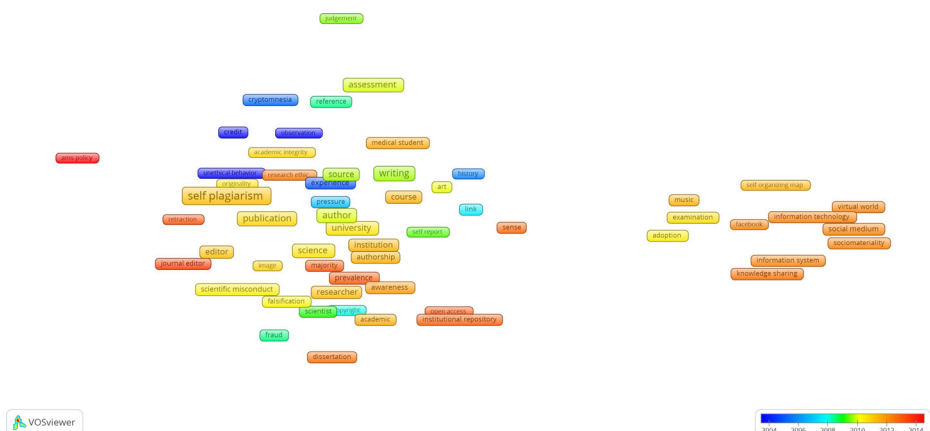

Fig. 3 Evolution of self - plagiarism research terms

Research Topic Typology

The science map presented in Fig. 4 consists of five clusters. Based on the cluster terms, we first merged some of the clusters and then labelled them with an appropriate research topic. Based on the topic label and representative words, we identified the most frequently cited recent paper(s) for each topic and summarised these in the form of a scoping review. The resulting typology is shown in Table 1.

In the next step of the analysis we used the derived typology to assign popular representative terms to the periods identified above (Table 2). We can observe the shift from a more vague ethical definition of self-plagiarism to stricter policies concerning it, and the shift from copy right to open access in institutional self-plagiarism. Additionally, we can observe that research in self-plagiarism in academic writing is diminishing in intensity in the attenuation phase. Additionally, the retraction of articles has become a popular term in the attenuation phase under the topic of Self-plagiarism in science.

In common with other bibliometric studies, our study also has some limitations. The corpus is relatively small, so removing common words may have affected the emergence of clusters. Using a method other than the default VOSViewer might have resulted in a different appearance of the scientific landscapes. The bibliometric analysis was limited to publications written in English and to the Scopus database, so some interesting publications may have been omitted.

Conclusion

Using the bibliometric mapping approach, we created three science landscapes dealing with self-plagiarism research. Based on the clusters emerging in the science landscape, we devised an empirically grounded typology consisting of five topics: Self-plagiarism, Institutional self-plagiarism, Self-plagiarism and ICT, Self-plagiarism in academic writing, and Self-plagiarism in science. The literature related to the typology topics shows that researchers are reaching the agreement that self-plagiarism is unethical and dishonest. Consequently, firmer policies and guidelines have been established. However, some reservations about self-plagiarism still exist:

Fig. 4   The scientific landscape of self -plagiarism research

Table 1 The typology o self -plagiarism research topics

Research topic	Representative terms	Scoping review of most cited papers from the attenuation phase (2013–2016)
Self-plagiarism (light blue colour)	*self-plagiarism, AMS policy, research ethics*	Berquist in an editorial writes that many authors when confronted by self-plagiarism claim that it was an "honest mistake" (Berquist 2013). Textual recycling in self plagiarism is a controversial concept not well defined in research ethics literature (Burton 2014). 30 different paraphrasing strategies were identified. Substitution and copying verbatim were most often used strategies (Sun and Yang 2015). "Self-plagiarism occurs when substantial amounts of text previously published by the same author are used without citation and without quotation" is the definition of self-plagiarism provided by American Meteorology Society (Schultz et al. 2015).
Institutional self-plagiarism (yellow colour)	*university, institution, academic, awareness, copyright, dissertation, institutional repository, open access*	Attempts to increase journal's 'impact factor' through 'coercive citations,' and growing trend to quantify research performance through various indicators may result in problems of research integrity including self – plagiarism not only by authors, but also by journal editors (Martin 2013). Zero tolerance to any form of plagiarism is advocated by some authors (Andreescu 2013). In addition to support the accessing of knowledge open access also contributes to minimizing negative aspects of plagiarism/self-plagiarism due to the fact that documents are exposed to peers in real time (Rodrigues and Rodrigues 2012).
Self-plagiarism and ICT (dark blue colour)	*information technology, information system, self-organising map, social medium, sociomateriality, adaptation, knowledge sharing, virtual world*	As on the one side ICT accelerated the phenomenon of plagiarism enabling simple text adaptations (Ozmercan 2015), it on the other side also enabled to detect it in any form (Sun 2013). Detecting software is mainly based on text mining and self-organising maps (Oberreuter and Velásquez 2013). On the other hand Chew et al. (2015) suggest that plagiarism detection software in their case Turnitin should not be used as a plagiarism detection tool but rather self – assessment and self - service aid. Additionally, it has been shown that the appearance of Internet is not statistically significantly changed the prevalence of plagiarism in doctoral dissertations (Ison 2015).

Table 1 (continued)

Research topic	Representative terms	Scoping review of most cited papers from the attenuation phase (2013–2016)
Self.plagiarism in academic writing (green and violet colours)	*writing, reference, course, assessment, essay, academic integrity, cryptomnesia,*	Halupa and Bolliger (2015) reported that students didn't received information about self – plagiarism and that students believed that they own their previously published works and in that manner self-plagiarism should not be considered as an academic dishonesty. Burton (2014) asserts that self-plagiarism is unethical because it is deceptive and dishonest. Robinson (2014) agrees and claims that self-plagiarism is an academic misconduct in every case. Contrary, Callahan (2014) is more critical and asks if it is unethical to reuse for example background statement for a new paper which presents original insights or it is on the other hand unethical on the part of those who gain (financially or otherwise) from requiring exclusivity? An interesting case of inadvertent self – plagiarism is cryptomnesia, especially by novices who by increased cognitive load my show an inability to recall the source of an idea (Dow 2015)
Self-plagiarism in science (red colour):	*originality, authorship, misconduct, scientific misconduct, fabrication, falsification, fraud, breach, publication, researcher, scientist, journal editor,* and *authorship.*	The self - plagiarism is a threat to the research integrity and is thus unacceptable (Juyal et al. 2015). Hoverer, an interesting view of self-plagiarism is given by Robinson (2014), namely that questions about self-plagiarism should be replaced by the notation about "fortunate or unfortunate scientific publications".

for example, whether it is unethical to reuse a background statement for a new paper presenting original insights, or when extending a conference paper to a full journal paper, or whether the financial interest of those who gain from requiring exclusivity is even more unethical (sociomateriality). The state of the art topics seem to be "social medium", "virtual world", "Facebook", sociomateriality", "knowledge sharing", "open access", "institutional repository", "retraction" and "responsible conduct".

Our study can help concerned outsiders to learn more about the evolution of self-plagiarism research and its main topics; it can also guide them to appropriate literature in each topic and enable all interested parties to identify guidelines for further research. Furthermore, the derived typology can serve as a starting point for performing systematic reviews of the production of self-plagiarism literature.

Table 2 The distribution of representative terms according to the self-plagiarism research topics and chronological phases

Self-plagiarism research topic	Initiation phase (2003–2006)	Rapid growth phase (2007–2012)	Attenuation phase (2013–2015)
Self-plagiarism	*N/A*	*self-plagiarism, research ethics*	*AMS policy*
Institutional self-plagiarism	*copy right*	*institution, university, awareness, academic*	*dissertation, institutional repository, open access*
Self-plagiarism and ICT	*N/A*	*adaption, self-organising maps, music*	*Social medium, information, technology, socimateriality, knowledge sharing*
Self-plagiarism in academic writing	*Cryptomnesia, unethical behaviour,*	*reference, judgment, assessment, authorship*	*N/A*
Self-plagiarism in science	*N/A*	*originality, authorship, misconduct, scientific misconduct, fabrication, falsification, fraud, breach, publication, authorship*	*journal editor, retraction, responsible conduct*

References

American Psychological Association. (2010). *The Publication Manual of the American Psychological Association* (6th ed.). Washington, DC: American Psychological Association.

Andreescu, L. (2013). Self-plagiarism in academic publishing: the anatomy of a misnomer. *Science and Engineering Ethics, 19*(3), 775–797.

Berquist, T. H. (2013). Self-plagiarism: a growing problem in biomedical publication. *American Journal of Roentgenology, 200*(2), 357.

Bruton, S. V., & Rachal, J. R. (2015). Education journal editors' perspectives on self-plagiarism. *Journal of Academic Ethics, 13*(1), 13–25.

Burton, S. V. (2014). Self-plagiarism and textual recycling: legitimate forms of research misconduct. *Accountability in Research, 21*(3), 176–179.

Callahan, J. L. (2014). Creation of a moral panic? Self-plagiarism in the academy. *Human Resource Development Review, 13*(1), 3–10.

Carroll, C., & Booth, A. (2015). Quality assessment of qualitative evidence for systematic review and synthesis: Is it meaningful, and if so, how should it be performed? *Research and Synthesis Methods, 6*(2), 149–154.

Chew, E., Ding, S. L., & Rowell, G. (2015). Changing attitudes in learning and assessment: cast-off 'plagiarism detection' and cast-on self-service assessment for learning. *Innovations in Education and Teaching International, 52*(5), 454–463.

Cunha, H. (1946). Gastrostomy. Its inception and evolution. *The American Journal of Surgery, 72*(4), 610–634.

De Bellis, N. (2009). *Bibliometrics and citation analysis*. Lanham: The Scarecrow Press, Inc.

de Vasconcelos, S. M., & Roig, M. (2015). Prior publication and redundancy in contemporary science: are authors and editors at the crossroads? *Science and Engineering Ethics, 21*(5), 1367–1378. doi:10.1007/s11948-014-9599-8.

Dow, G. T. (2015). Do cheaters never prosper? The impact of examples, expertise, and cognitive load on cryptomnesia and inadvertent self-plagiarism of creative tasks. *Creativity Research Journal, 27*(1), 47–57.

Garfield, E. (2006). The history and meaning of the journal impact factor. *Journal of American Medical Association, 295*(1), 90–93.

Halupa, C. M. (2014). Exploring student self-plagiarism. *International Journal of Higher Education, 3*(1), 121–126.

Halupa, C., & Bolliger, D. U. (2015). Student perceptions of self plagiarism: a multi-university exploratory study. *Journal of Academic Ethics, 13*(1), 91–105.

Heersmink, R., van den Hoven, J., van Eck, N. J., & van den Berg, J. (2011). Bibliometric mapping of computer and information ethics. *Ethics in Information Technology, 13*, 241–249.

Ison, D. C. (2015). The influence of the internet on plagiarism among doctoral dissertations: an empirical study. *Journal of Academic Ethics, 13*(2), 151–166.

Juyal, D., Thawani, V., & Thaledi, S. (2015). Plagiarism: an egregious form of misconduct. *North American Journal of Medical Sciences, 7*(2), 77–80.

Martin, B. R. (2013). Whither research integrity? Plagiarism, self-plagiarism and coercive citation in an age of research assessment. *Research Policy, 42*(5), 1005–1014.

Maurer, H., Kappe, F., & Zaka, B. (2006). Plagiarism - a survey. *Journal of Universal Computer Science, 12*(8), 1050–1084.

Oberreuter, G., & Velásquez, J. D. (2013). Text mining applied to plagiarism detection: the use of words for detecting deviations in the writing style. *Expert Systems with Applications, 40*(9), 3756–3763.

Ozmercan, E. E. (2015). Determining the tendencies of academic dishonesty and senses of self-efficacy with discriminant analysis. *Anthropologist, 20*(1–2), 353–359.

Pritchard, A. (1969). Statistical bibliography; an interim bibliography. *SABS, 5*, 184–244.

Robinson, S. R. (2014). Self-plagiarism and unfortunate publication: an essay on academic values. *Studies in Higher Education, 39*(2), 265–277.

Rodrigues, M. E., & Rodrigues, A. M. (2012). Analyzing the performance of an institutional scientific repository – A case study. *LIBER Quarterly, 22*(2), 98–117.

Roig, M. (2014). Journal editorials on plagiarism: what is the message? *European Science Editing, 40*(3), 58–59.

Schultz, D. M., Rauber, R. M., & Heideman, K. F. (2015). AMS policy on plagiarism and self-plagiarism. *Journal of Climate, 28*(3), 909–910.

Shen, J., Yao, L., Li, Y., Clarke, M., Wang, L., & Li, D. (2013). Visualizing the history of evidence-based medicine: a bibliometric analysis. *Journal of the American Society for Information Science and Technology, 64*(10), 2157–2172.

Sun, Y.-C. (2013). Do journal authors plagiarize? Using plagiarism detection software to uncover matching text across disciplines. *Journal of English for Academic Purposes, 12*(4), 264–272.

Sun, Y.-C., & Yang, F.-Y. (2015). Uncovering published authors' text-borrowing practices: paraphrasing strategies, sources, and self-plagiarism. *Journal of English for Academic Purposes, 20*, 224–236.

US Department of Health and Human Services (2011) *Academic self-plagiarism (Double dipping)*. Retrieved October 26, 2015, from http://ori.hhs.gov/plagiarism-15

Van Eck, N. J., & Waltman, L. (2013). *VOSviewer Manual*. Leiden: Univeristeit Leiden.

Young, F. W., & Hamer, R. M. (2013). *Multidimensional scaling: history, theory, and applications*. Florence: Psychology Press.