Scientometrics (2012) 92:711–719
DOI 10.1007/s11192-012-0662-4

Citation behavior in popular scientific papers: what is behind obscure citations? The case of ethnobotany

Marcelo Alves Ramos · Joabe Gomes Melo · Ulysses Paulino Albuquerque

Received: 22 November 2011 / Published online: 17 February 2012

Abstract Citation studies have become an important tool for understanding scientific communication processes, as they enable the identification of several characteristics of information-retrieval behavior. This study seeks to analyze citation behavior using two popular ethnobotany articles, and our analysis is guided by the following question: when an author references a work, is he pointing out the work's theoretical contribution, or is bias a factor in citing this reference? Citation analysis reveals an interesting phenomenon, as the majority of citing texts do not consider the theoretical contributions made by the articles cited. Two possible conclusions can be drawn from this scenario: (1) citing authors read the original texts that they cite only superficially, and (2) the works cited are not read by the vast majority of people who reference them. Thus, it is clear that even with sufficient access to reference texts; ethnobotanical studies highlight elements less relevant to the research and reproduce discussions in a non-reflective manner.

Keywords Citation analysis · Scientometrics · Scientific quality

Introduction

Now more than ever, scientists are evaluated by the quality and quantity of their publications, and these assessment criteria are often measured by the number of articles that they have published in high-impact journals. In principle, publishing an article in this type of journal affords greater visibility to both the research and the author (Seglen 1997), but it is important to note that some studies show no direct relationship between publishing an article in a high-impact journal and a high number of citations (Albuquerque 2010).

Citation of an article may be influenced by various biases, such as self-citation, nationality, sex, and the author's institution; in other words, this practice may be influenced by a spectrum of psychological, sociological, political and historical factors

M. A. Ramos · J. G. Melo · U. P. Albuquerque (✉)
Laboratório de Etnobotânica Aplicada, Departamento de Biologia,
Universidade Federal Rural de Pernambuco, Rua Dom Manoel de Medeiros,
s/n, Recife, PE, Brazil
e-mail: upa@db.ufrpe.br

(Albuquerque 2010; Alvarenga 1998; Vanz and Caregnato 2003). It is therefore unsurprising that the scientific community questions the use of quantitative measures based on citations, such as the *h-index*, to measure an author's performance and, by extension, the "popularity" of his or her articles.

According to Simkin and Roychowdhury (2003), many studies that investigate the popularity of scientific publications have been conducted; the majority, however, have focused on analyzing citation distribution. Simkin and Roychowdhury developed a quantitative method based on the misprint distribution in citations and concluded that only 20% of the citations analyzed were actually consulted by the citing authors. This result indicates that many citations are not read but are merely copied from other works' reference lists, potentially leading to various problems, such as overestimation of the impact of a publication, propagation of errors of interpretation, and even incorrect and/or inappropriate citations. Simkin and Roychowdhury (2005) also reported a cumulative effect that leads to a greater probability of a work being cited in the future, a phenomenon alternately known as the "Matthew effect," "cumulative advantage," or "preferential attachment."

Consequently, the following fundamental question not highlighted by mathematical models arises: when an author cites a work, even if he or she has read the original text, are these texts cited for their intrinsic value (merit), or does bias play a role in maximizing the citation? Using this question as a starting point, this study aims to evaluate aspects of scientists' citation behavior based on a qualitative and quantitative analysis of articles that cite work of great relevance and popularity while performing a first examination of information flow in scientific communication in the field of ethnobotany. In this sense, ethnobotany offers a good scenario to study "citation behavior". According to Bermudez et al. (2005): "the interdisciplinary nature of ethnobotany allows for a wide range of approaches and applications. Nevertheless, little interchange of theories and methods among related disciplines has taken place to date, resulting in the predominance of descriptive works, which are primarily limited to compilations of useful plants".

Materials and methods

In selecting citations, we considered two main inclusion criteria in this study: (1) articles must be among the most cited in *Economic Botany*, which is one of the oldest reference journals that publishes articles on ethnobotany, and (2) articles must present a degree of originality and novelty from a theoretical point of view. The two works that fulfilled these criteria were Phillips and Gentry (1993) and Bennett and Prance (2000).

Following their selection, a search was conducted in November 2011 of all available periodicals in the SCOPUS database (http://www.scopus.com) to locate all articles that cited these works since their publication. We identified 131 articles that cited Phillips and Gentry (1993) and 81 that cited Bennett and Prance (2000). The articles were analyzed individually, and for each instance that they referred to the selected authors, the identified text was classified according to the following three categories (listed in ascending order of relevance): (1) Category 1: citation of minor relevance (i.e., the central idea of the article cited was not considered); (2) Category 2: citation of intermediate relevance (i.e., replication of quantitative techniques suggested by the work); (3) Category 3: citation of great relevance (i.e., novelty on the part of the authors that highlighted its theoretical point of view, which was the intrinsic value of the research). If a work contained more than one citation, and if these citations fit into more than one category, only the most relevant citation was considered.

Peritz (1992) states that a citation motivated by a theoretical connection to the topic must be quantitatively different from a citation used to simply indicate a study's use or application, which emphasizes the importance of classifying bibliographical citations into different levels. However, it should be noted that the classification established in this study is not intended to measure the number of times a particular author is correct or incorrect in deciding to quote particular passages of the works being accessed; that is, we do not want to discount citations in Categories 1 or 2, given that each author is responsible for highlighting (i.e., citing) information that he or she considers important in a scientific work. For the intended objectives of this research, however, it was necessary to create this classification to understand both the behavior of citing popular texts in ethnobotany and the possible biases associated with their citation.

Phillips and Gentry's (1993) article, which Albuquerque (2009) considers a benchmark work in ethnobotany, was one of the first clear responses to the criticism that ethnobotany suffers from subjectivity. The work of these authors relies on the hypothetico-deductive method for testing hypotheses in ethnobotany. For this purpose, these authors proposed using the use value, which is a quantitative tool that determines the relative importance of plants. From this tool, it is possible to test hypotheses; for example, the relative importance of a resource can be explained by its availability in nature, which is an offshoot of the hypothesis of ecological apparency (Albuquerque and Lucena 2005; Lucena et al. 2007). Since the publication of this work, the term "quantitative ethnobotany" has been gradually spreading and is now being used in many different approaches (Albuquerque 2009). We assume, however, that the authors' actual proposal (i.e., intrinsic value) was to introduce hypothesis testing in ethnobotany research rather than to propose a new quantitative tool.

Bennett and Prance's (2000) research unfolds in a similar manner. Ethnobotanical literature has long pointed to the use of exotic plants in different cultures as an acculturative phenomenon with no analytical effort or more detailed interpretation. Like Phillips and Gentry (1993), the authors proposed a quantitative tool that measures the relative importance of plants based on their versatility and hypothesized that medicinal plants are introduced into a culture for other reasons, such as food nourishment and ornamental use. This idea may help explain how different cultures' pharmacopeias develop structurally over time. Interestingly, to the best of our knowledge, this "versatility hypothesis" was only formally tested and compared with alternative hypotheses 10 years later (see Alencar et al. 2010).

Data analysis

All articles that cited Phillips and Gentry (1993) and Bennett and Prance (2000) were grouped by year of publication (Fig. 1). The following two temporal categories were identified for each work: (1) for Phillips and Gentry, works that cited the reference before 2005 and those that cited it after 2005 (Fig. 1a); (2) for Bennett and Prance, works that cited the reference before 2006 and those citing it after 2006 (Fig. 1b). The period between 2005 and 2006 was identified as the moment when these articles became most cited in the scientific literature.

Using BioEstat software 5.0 (Ayres et al. 2007), a Chi-square test (χ^2) was applied to ascertain if the category distribution between the two articles presented statistical differences. Meanwhile, a G test (contingency table) was used to analyze citations of Phillips and Gentry (1993) and Bennett and Prance (2000) in the two time categories classified in this work.

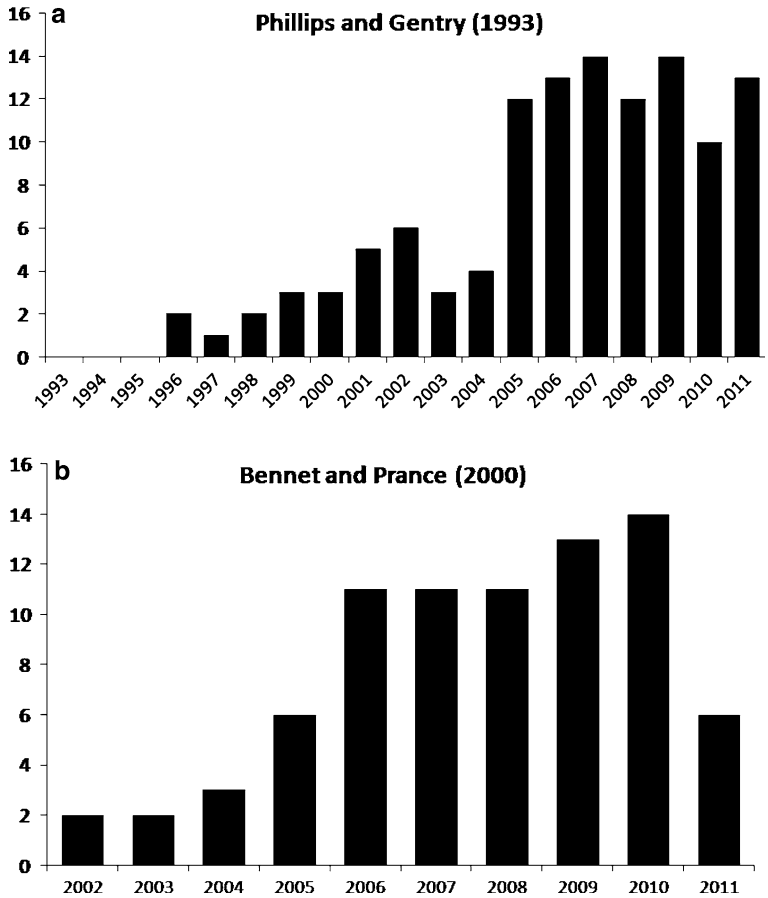

Fig. 1 Number of articles that cited Phillips and Gentry (**a**) and Bennett and Prance (**b**)

Results and discussion

Quantitative analysis

In the literature studied, the majority of articles that cited Phillips and Gentry (1993) and Bennett and Prance (2000) did not consider in their citations the primary theoretical contributions made by these authors (42.3 and 56.5% of publications, respectively) (Fig. 2); instead, these articles highlighted seemingly less relevant aspects of the referenced texts (Category 1). We also found citations made exclusively to replicate the quantitative techniques proposed by these authors (Fig. 2), which occurred in 28.7% of the works citing Phillips and Gentry (1993) and in 38.5% of those citing Bennett and Prance (2000). Works that used the indices proposed by the authors were allocated to this group without consideration for the theoretical implications involved or, in many cases, the context in which they were originally applied. Interestingly, only a small number of the works analyzed cited references that considered the articles' real theoretical contribution (i.e., intrinsic value) (Fig. 2).

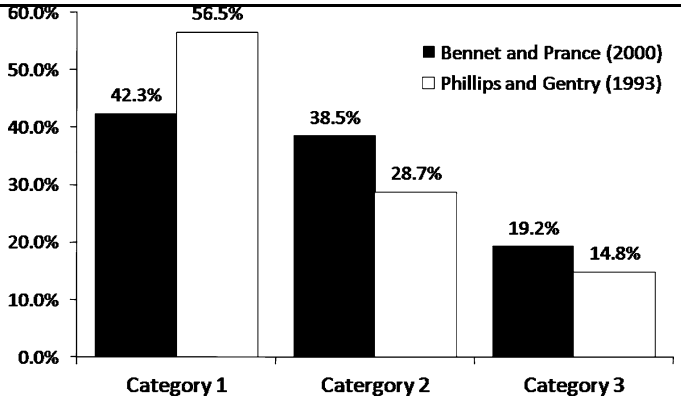

Fig. 2 Classification of citations of articles Phillips and Gentry (1993) and Bennett and Prance (2000) in the scientific literature, considering the content presented in these citations. *Category 1*: citation of minor relevance; (2) *Category 2*: citation of intermediate relevance; (3) *Category 3*: citation of great

The data presented here can be interpreted in the following two distinct ways: (1) the citing authors of these articles read the original texts superficially, and (2) these works are not read by the vast majority of people who reference them. According to Simkin and Roychowdhury (2003), the practice of replicating bibliographical citations without reading the original source occurs somewhat frequently in science; therefore, we cannot rule out this possibility for the references analyzed in this study. Similarly, Carvalho (1975) reports that we cannot expect that all authors are careful, objective, and conscientious when referencing their sources.

The above scenario was observed for the two works examined here, i.e., the distribution of citations between the different categories was the same ($\chi^2 = 3.762$, $p > 0.05$), showing consistency in how these works are cited. Despite this similarity, it is noteworthy that while the article by Phillips and Gentry (1993) was cited predominantly in the introductions of other works (44.87%) and subsequently in the methods and discussion sections (Fig. 3), references made to Bennett and Prance (2000) were normally located in the discussion section (39.8%) followed by the introduction and methods sections (Fig. 3). Regardless of whether the citations were concentrated in a work's introduction or discussion, the pattern observed in the two analyzed references was the same, i.e., in the texts that cited them, the authors do not expound on the major theoretical issues proposed by these references.

This study is not concerned with understanding the reasons that led different authors to cite the references analyzed in this work, as the citation process is subjective, and one cannot quantify the human element that generates subjectivity in the act of citation. However, as outlined in the literature, a number of reasons may lead an author to cite a specific reference, such as the significance of using a particular theory or paradigm; recognition of a renowned work to highlight or rectify the research of the individual citing it; promotion of a publication's own articles (i.e., self-citation); demonstration of the evolution of an area of research; and the citing authors' beliefs that their peers consider what is being cited to be important (Bavelas 1978; Carvalho 1975; Vanz and Caregnato, 2003). In this case, we believe that a powerful force that may have guided the use of these citations is the false idea that quality of scientific work equates to use of quantitative tools (Albuquerque 2009).

In the annual distribution of works that cite Phillips and Gentry (1993) and Bennett and Prance (2000), it was observed that these two articles were most cited in the years 2005 and

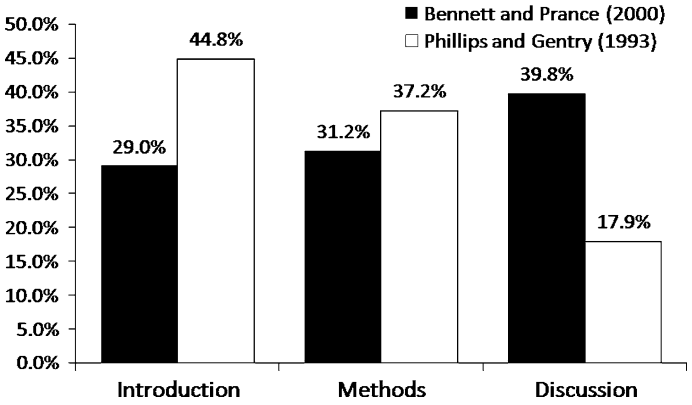

Fig. 3 Distribution of citations of Bennett and Prance's (2000) and Phillips and Gentry's (1993) articles in various parts of the publications mentioned

2006, respectively (Fig. 1). These works have since been cited frequently in ethnobotany literature, which may reflect that these texts have now become "classics" (see Simkin and Roychowdhury 2003). The dynamics of this process involve a chain of events initiated each time a particular work is cited by another; therefore, after being repeatedly cited, a scenario is created in which the possibility of the work being quoted in the future increases exponentially. However, a major concern occurs when these works are cited without consulting the original, as some of our findings suggest and as has been confirmed in the literature (Simkin and Roychowdhury 2003).

Analyzing references to Phillips and Gentry (1993) in the two time categories considered here (Fig. 1a), the same pattern of citation was observed in both periods (G test: 1.38, $P > 0.05$). Most citations were classified as Category 1 (Fig. 4) (i.e., the group of quotations of minor relevance) because they did not take into account the main idea presented by the authors. This category was closely followed by Category 2 (i.e., replication of the quantitative techniques suggested by the authors) and by Category 3 (i.e., works that discuss the main theoretical contribution of the cited work) (Fig. 4). These results show that regardless of the fact that these articles are referenced more often, the manner in which they are cited remains similar.

Turning to Bennett and Prance (2000) and taking the year 2006 as a historical milestone for its citation (Fig. 1b), statistical differences were not observed in the distribution of citation categories when comparing the two periods (G test: 1.59; $P > 0.05$). Nevertheless, we note that before 2006, citation of this study normally fell into Category 1 (see Fig. 4) and became more popular after 2006 when it began to be used for replicating the quantitative technique proposed by the authors (Category 2). Despite being considered a category of intermediate relevance, according to Garfield (1979), the citation of works for their methodological contribution is as important as those citations that highlight their theoretical contributions.

Many factors contribute to the increasing/decreasing popularity of scientific papers, and as noted here, they are not always related to a work's theoretical contribution. Wren (2005) argues that the advent of the internet has caused a paradigm shift, as the Internet has allowed access to a wide variety of sources of content and knowledge dissemination. Eysenbach (2006) reinforces this argument when he assumes that articles immediately published in Open Access format are cited more often, comparatively, than those published

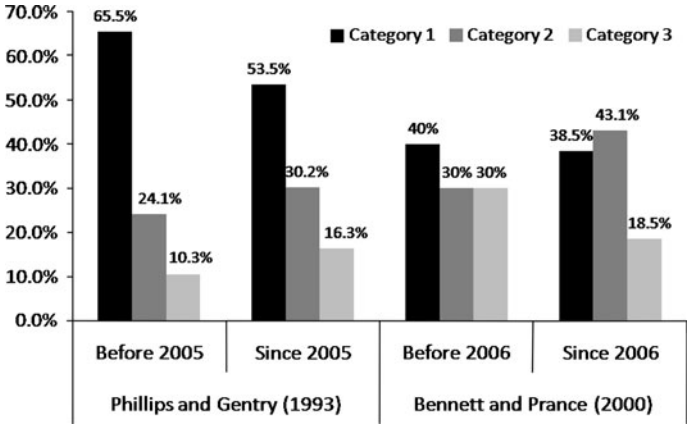

Fig. 4 Classification of citations of Phillips and Gentry (1993) and Bennett and Prance (2000) in the scientific literature, considering the content presented in these quotations (categories) and the different periods in which these quotations have become more popular. *Category 1*: citation of minor relevance; (2) *Category 2*: citation of intermediate relevance; (3) *Category 3*: citation of great

in traditional format, although we did not find support for this notion in the articles analyzed here.

Qualitative analysis

One of the most serious concerns encountered in the evaluated work was "incorrect attribution," which refers to attributing information to the wrong author. This problem occurs with widely disseminated scientific writing and may cause the dissemination of erroneous information, among other problems. We found several inconsistent pieces of information on Phillips and Gentry's (1993) and Bennett and Prance's (2000) findings. One example we identified was Phillips and Gentry's (1993) work, which was conducted in Peru, being cited to support information related to North America, as if the study had been conducted in that region. The study was also cited to support the need to identify the economic value of plant species, when in fact these authors did not deal with this issue, nor does their species' use value quantitative index consider economic issues related to plant usage.

In addition to attribution errors, the underuse of a citation also contributes to its dissemination. As mentioned earlier, in terms of the theoretical contributions of the studies analyzed, their scientific importance is unquestionable, but the factors responsible for the promotion of their popularity are curious. Bennett and Prance (2000), for example, have been widely cited to support well-known claims in ethnobotany, such as the cultural importance of the families Asteraceae, Lamiaceae, and Leguminosae or the importance of exotic plants in the pharmacopoeia. In addition to being commonly cited to support the classification of the uses of plants in utility categories (e.g., medicinal, fuel, etc.), Phillips and Gentry (1993) are also mentioned in methods sections to justify conducting one-on-one interviews with informants to avoid interference from others in the process.

Vanz and Caregnato (2003) point out that citations reflect the influence of the author cited on the citing author's work, but this hypothesis does not seem to correspond to the

findings described here, as the true importance of the evaluated studies was not highlighted in the research presented by the citing author.

Plagiarism was also found in the works evaluated in this study, which was not surprising, given that this practice is recognized as a common phenomenon in scientific circles, often due to faults in the publishing process (Long et al. 2009).

Given the reproduction of identical sentences and repeated misspellings in how an author is cited, plagiarism was not difficult to identify in this study, even without the use of commercial tools used to detect this practice. Repeated misspellings are especially alarming because a spelling error being perpetuated in more than one work may mean that the original sources were never consulted, which may then lead to the reproduction of original misinterpretations. This finding reinforces our interpretation that the works may have been cited without having been read.

The issues raised throughout this study show that the forces that maximize the citation of a reference, as noted here for Phillips and Gentry (1993) and Bennett and Prance (2000) in ethnobotany, are not always related to the works' theoretical contributions. The current pressures in academia to "publish or perish" may lead some researchers to adopt strategies to maximize their output and to include citations from a greater number of articles. Consequently, time becomes a limiting factor, which, when combined with the ease of accessing information via the Internet, contributes to plagiarism and reproduction of citations that, in most cases, have not been read by the author but merely copied from other sources. This practice must change to ensure that information is truly validated and referenced without replication of erroneous or less important data and that the true contribution of a work is highlighted, for as Vanz and Caregnato (2003) point out, the obligation of every researcher is to coherently disseminate scientific knowledge through the publication of their own investigations.

This need is intensified in ethnobotany, as this field must advance theoretically through work that seeks to test hypotheses, thereby reinforcing the need for researchers to contextualize their research from a theoretical point of view (see Albuquerque and Hanazaki 2009; Albuquerque 2011). Therefore, the situation identified in this work demonstrates that even with sufficient access to reference texts, ethnobotanical studies continue to highlight less relevant elements of the research and reproduce arguments in a non-reflective way.

Citation studies have become an important tool for better understanding the processes of scientific communication, thereby allowing the identification of a series of behavioral characteristics in the use of retrieved information. Such studies have also become indicators of scientific activity, as they help to elucidate the structure and development of science and identify the way in which it functions (Vanz and Caregnato 2003).

Acknowledgments The authors thank the Programa Nacional de Pós Doutorado (PNPD) of Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and the Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (FACEPE), by grants for postdoctoral awarded to M. A. Ramos and J. G. Melo, respectively; and also to the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for productivity grant awarded to U. P. Albuquerque.

References

Albuquerque, U. P. (2009). Quantitative ethnobotany or quantification in ethnobotany? *Ethnobotany Research and Applications, 7*, 01–03.

Albuquerque, U. P. (2010). The tyranny of the impact factor: why do we still want to be subjugated? *Rodriguesia, 61*, 353–358.

Albuquerque, U. P. (2011). The tragedy of the common reviewers: the peer review process. *Revista brasileira de. Farmacogosia, 21*(1), 1–3.
Albuquerque, U. P., & Hanazaki, N. (2009). Five problems in current ethnobotanical research and some suggestions for strengthening them. *Human Ecology (New York, N.Y.), 37*, 653–661.
Albuquerque, U. P., & Lucena, R. F. P. (2005). Can apparency affect the use of plants by local people in tropical forests? *Interciencia, 30*(8), 506–510.
Alencar, N. L., Araújo, T. A. S., Amorim, E. L. C., & Albuquerque, U. P. (2010). The inclusion and selection of medicinal plants in traditional pharmacopoeias evidence in support of the diversification hypothesis. *Economic Botany, 64*, 68–79.
Alvarenga, L. (1998). Bibliometria e Arqueologia do Saber de Michel Foucault: traços de identidade teórico-metodológica. *Ciência da Informação, 27*(3), 1–9.
Ayres, M., Ayres, M., Jr, Ayres, D. L., & Santos, A. S. (2007). *BioEstat: aplicações estatísticas nas áreas das ciências biológicas e médicas*. Belém: Sociedade Civil Mamirauá: MCT-CNPq.
Bavelas, J. B. (1978). The social psychology of citations. *Canadian Psychological Review, Calgary, 19*(2), 158–163.
Bennett, B. C., & Prance, G. T. (2000). Introduced plants in the indigenous pharmacopoeia of northern South America. *Economic Botany, 54*(1), 90–102.
Bermudez, A., Oliveira-Miranda, M. A., & Velazquez, J. (2005). La Investigación etnobotánica sobre plantas medicinales: Una revisión de sus objetivos y enfoques actuales. *Interciencia, 30*(8), 453–459.
Carvalho, M. M. (1975). Análises Bibliométricas da Literatura de Química no Brasil. *Ciência da Informação, 4*(2), 119–141.
Eysenbach, G. (2006). Citation Advantage of Open Access Articles. *PLoS Biology, 4*(5), 692–698.
Garfield, E. (1979). Is citation analysis a legitimate evaluation tool? *Scientometrics, 1*(4), 359–375.
Long, T.C., Errami, M., George, A.C., Sun, Z., & Garner, H.R. (2009). Responding to possible plagiarism. *Science, 323*(5919), 1293–1294.
Lucena, R. F. P., Araújo, E. L., & Albuquerque, U. P. (2007). Does the local availability of woody caatinga plants (Northeastern Brazil) explain their use value? *Economic Botany, 61*, 347–361.
Peritz, B. C. (1992). On the objectives of citation analysis: problems of theory and method. *Journal of the American Society for Information Science, 43*(6), 448–451.
Phillips, O., & Gentry, A. H. (1993). The useful plants of Tambopata, Peru: I. Statistical hypothesis tests with a new quantitative technique. *Economic Botany, 47*, 15–32.
Seglen, O. P. (1997). Why the impact factor of journals should not be used for evaluating research. *BMJ, 314*, 498–502.
Simkin, M. V., & Roychowdhury, V. P. (2003). Read before you cite! *Complex Systems, 14*, 269–274.
Simkin, M.V., & Roychowdhury, V. P. (2005). Copied citations create renowned papers? *Annals of Improbable Research, 11*(1), 24–27. http://arxiv.org/ftp/cond-mat/papers/0305/0305150.pdf.
Vanz, S. A. S., & Caregnato, S. E. (2003). Estudos de Citação: uma ferramenta para entender a comunicação científica. *Em Questão, 9*(2), 295–307.
Wren, J.D. (2005). Open access and openly accessible: a study of scientific publications shared via the internet. *BMJ, 330*(7500), 1128.