DSI DATA SCIENCE AND INFORMETRICS
VOL.3, NO.2, MAY. 2023

RESEARCH ARTICLE

# Dissemination network of referencing errors: A comparison between the two most cited papers in the world

Liming Liang[a]*, Zhen Zhong[b]

a. Institute for Science, Technology and Society, Henan Normal University, Xinxiang, China
b. School of Management, Henan University of Technology, Zhengzhou, China

**ABSTRACT**

Lowry's paper (1951, *Journal of Biological Chemistry*) is the most highly cited paper in the world up to 26th, December 2022. Laemmli's paper (1970, *Nature*) ranks second. These two papers are selected as our sample to compare respective dissemination networks of referencing errors. There are several types of referencing errors, and we focus on the so-called "double error", that is, the co-occurrence of a wrong volume number of a journal and a wrong beginning page number of a paper. Since the probability of a double error is extremely small, so that whenever double error occurs more than once, we can deduce definitely that there are some authors who copy references from the reference list of other publications. Based on the dissemination networks of the double errors of the above-mentioned two papers, three dissemination routes are confirmed. Route 1: copying references from another paper and citing that paper. Route 2: copying references from another paper without citing that paper. Route 3: Sharing at least one author with another paper which carries the same referencing error. Some suggestions are offered to reduce or eliminate the misconduct of copying references without any checking.

**KEYWORDS**

Referencing errors; Dissemination network; Citing-cited relation; Reference copying without check

## 1 Introduction

We launched an interesting study in 2005 on the dissemination network of referencing errors. The sample used is a famous article by Laemmli, U.K., published in *Nature* in 1970 (Laemmli, 1970), which is the second most cited article in the world with about 190,000 citations (Liang & Zhong, 2007). This study is based on the citing-cited relations. A few years later, in 2012 we found that Laemmli' s paper had received many new citations since 2007, reaching 224,000 citations.

It is worth noting that its dissemination network of referencing errors continued to expand. We wondered how it occurred. Then we improved our methodology and relaunched this study. In the relaunched study we considered not only the citing-cited relation between the

* Corresponding Author: liangliming1949@sina.com

papers carrying the same error, but the collaboration among the error disseminators. Thus, some new discoveries were obtained (Liang et al., 2014). Analyzing the dissemination networks of the referencing errors, three dissemination routes were inferred.

Route 1. Copying and citing -copying references from another paper and citing that paper.

Route 2. Copying without citing - copying references from another paper without citing that paper.

Route 3. Sharing author(s) - sharing at least one author with another paper which carries the same referencing error. This points to copying a reference from an earlier paper by the same author. A case following Route 3 is of course also a case following Route 1 or Route 2.

Our first paper (Liang & Zhong, 2007) was published in a domestic Chinese journal and attracted some attention from Chinese colleagues (Yang, 2011; Fand, 2015; Zhang et al., 2018). The second paper (Liang et al., 2014) was published in *Scientometrics*, and its citations come from international scholars (Ainsworth & Russell, 2018; Olensky et al., 2016; Serenko et al., 2021) .

In our *Scientometrics* paper published in 2014, the citations to Laemmli' s paper were retrieved from Thomson Reuter' s Web of Science on December 16-20, 2012. Now it is December 2022, and a full ten years has passed. However, the study performed ten years ago still provoked us somehow. Is the error dissemination mode found through the citations of Laemmli' s paper typical and representative? It seems that we should relaunch a similar study on another highly cited paper (in fact, this would be the most highly cited paper) and make a comparison with the former study.

## 2 Data and Methods

In our previous study, Laemmli' s paper published in 1970 in *Nature* was selected as a sample. It is the second most highly cited paper in the world. Why did we select the second, not the most highly cited paper as our sample? The reason is that at that time we were performing a study on the "yield sequences" of *Nature* and *Science* (Liang & Rousseau, 2008). In the process of data collection, by accident we found a very highly cited paper in *Nature,* that is Laemmli' s paper. Though, very soon, we knew that Laemmli' s paper is just the second most highly cited paper in the world, not the most cited one. However, we still decided to select it as our sample, because it was published in *Nature*, one of the most influential multi-disciplinary journals in the world.

Now, in this study, the new sample we selected is really the most cited paper in the world. This paper was published in 1951 in *Journal of Biological Chemistry* by Lowry, OH; Rosebrough, NJ; Farr, AL and Randall, RJ (Lowry et al., 1951). Hereafter we refer to it as "Lowry' s paper" for short. Since its publication, Lowry' s paper has been cited 356,862 times. The data are retrieved from Clarivate' s SCI-Expanded on 26th, December 2022.

How can we find the referencing errors of Lowry' s paper? Our key method includes three steps. First, using "search" function in the WoS to find the correct information of Lowry' s paper. Second, using function "Cited Reference Search" to find all the citations to Lowry' s paper. Third, comparing the two searching results, and the difference between the two references is the referencing errors. Ignoring the spelling error of author name, the mistake in article title, and the wrong publication year, what we concerned are the wrong volume numbers and the wrong beginning page numbers of Lowry' s paper.

In concrete operation, first, we perform the "search" function with the following query in Web of Science Index: Science Citation Index Expanded (SCI-Expanded).

LOWRY OH (Author) and JOURNAL OF BIOLOGICAL CHEMISTRY (Publication Titles) and 1951 (Year Published)

The following result was obtained:

PROTEIN MEASUREMENT WITH THE FOLIN PHENOL REAGENT
By: LOWRY, OH (LOWRY, OH) ; ROSEBROUGH, NJ (ROSEBROUGH, NJ) ; FARR, AL (FARR, AL) ; RANDALL, RJ (RANDALL, RJ)
Volume 193 Issue 1 Page 265-275 Published1951 Indexed1951-01-01 Document Type Article

From this search, 345,868 citations are retrieved. We also verified the fact that for Lowry's paper the correct volume number is 193, and the correct beginning page is 265. We confirmed that Lowry published only one paper in *Journal of Biological Chemistry* in 1951.

Then, the function "Cited Reference Search" and the following search query was applied:

Cited Author=(Lowry, OH) AND Cited Work=(J BIOL CHEM) AND Cited Year=(1951)
Timespan=All Years. Databases=SCI-EXPANDED.

This search retrieved 347,618 citation records, 1750 records more than the previous result obtained by the "Search" function. Among the 1750 wrong citation records, 1678 carry the wrong volume number (e.g. vol. 277, vol. 224) or the wrong beginning page (e.g. page 180, 248), or with errors both in the volume number and the beginning page (e.g. vol.256 and page 495, vol.283 and page 249).

We focus on the volume-page double errors because of its very low probability of occurrence. A correct volume-page combination is 193-265, containing six digits. The probability of having one wrong digit in the volume number is 1 / (9+9+9) = 1/27, the same is true for the beginning page. Thus, the probability of a volume-page error with one wrong digit in the volume number and one wrong digit in the page number is only 1 / (27*27). Furthermore, if the wrong digits are more than one in both volume number and page number, it must be an extremely small probability event. The conclusion is that if paper A and paper B carry the same volume-page error, A or B must have copied the reference from the other without any check.

Table 1 shows the number of different error types and their frequencies. Table 2 lists the volume-page double errors (frequency≥ 4). All the data are from the citations to Lowry's paper.

**Table 1** Error type and frequency (Lowry 's paper)

| double errors or single errors | number of different types | frequency |
|---|---|---|
| correct volume, wrong beginning page | 249 | 901 |
| wrong volume, correct beginning page | 30 | 112 |
| volume–page double error | 370 | 665 |
| total | 649 | 1678 |

**Table 2** Volume-page double errors (frequency≥ 4) (Lowry 's paper)

| volume | beginning page | frequency | volume | beginning page | frequency | volume | beginning page | frequency |
|---|---|---|---|---|---|---|---|---|
| 243 | 1331 | 17 | 20 | 150 | 6 | 265 | 10118 | 5 |
| 127 | 182 | 13 | 87 | 206 | 6 | 139 | 190 | 4 |
| 226 | 497 | 13 | 98 | 719 | 6 | 183 | 501 | 4 |

| volume | beginning page | frequency | volume | beginning page | frequency | volume | beginning page | frequency |
|---|---|---|---|---|---|---|---|---|
| 261 | 6300 | 13 | 209 | 23 | 6 | 196 | 336 | 4 |
| 62 | 315 | 12 | 226 | 9453 | 6 | 235 | 3322 | 4 |
| 177 | 751 | 9 | 239 | 2370 | 6 | 237 | 3233 | 4 |
| 14 | 425 | 8 | 254 | 729 | 6 | 242 | 173 | 4 |
| 19 | 3 | 8 | 257 | 5333 | 6 | 242 | 1918 | 4 |
| 236 | 1372 | 7 | 258 | 2285 | 6 | 247 | 1106 | 4 |
| 250 | 8283 | 7 | 176 | 367 | 5 | 249 | 5153 | 4 |

# 3 Analysis and Results

We compare the error statistics and dissemination network of the two papers.

Table 1 and Table 2 are the statistical results of Lowry' s paper, while Table 3 and 4 are the corresponding results for Laemmli' s paper.

**Table 3** Error type and frequency (Laemmli' s paper)

| double errors or single errors | number of different types | frequency |
|---|---|---|
| correct volume, wrong beginning page | 49 | 8,764 |
| wrong volume, correct beginning page | 285 | 7,395 |
| volume–page double error | 225 | 463 |
| total | 559 | 16,622 |

**Table 4** Volume-page double errors (frequency≥ 4) (Laemmli' s paper)

| volume | beginning page | frequency | volume | beginning page | frequency | volume | beginning page | frequency |
|---|---|---|---|---|---|---|---|---|
| 256 | 495 | 18 | 251 | 614 | 7 | 307 | 478 | 4 |
| 283 | 249 | 15 | 277 | 580 | 6 | 277 | 174 | 4 |
| 97 | 620 | 13 | 224 | 149 | 6 | 277 | 608 | 4 |
| 348 | 699 | 12 | 69 | 646 | 6 | 277 | 689 | 4 |
| 226 | 112 | 12 | 302 | 76 | 5 | 277 | 6010 | 4 |
| 201 | 1130 | 11 | 48 | 617 | 5 | 270 | 57 | 4 |
| 27 | 580 | 10 | 318 | 78 | 4 | 680 | 685 | 4 |
| 263 | 789 | 7 | 314 | 472 | 4 | | | |
| 235 | 383 | 7 | 309 | 116 | 4 | | | |

Taking a closer look at Table 1 and Table 3, we found that the total referencing errors of Lowry' s paper account for only one-tenth (1678/16622) of Laemmli' s paper. The single errors of Lowry' s paper, no matter the pure volume errors or the pure beginning page errors, are much less than those of Laemmli' s paper. There exists a tremendous difference between the two papers. However, the scales of double errors of the two papers are in the same order of magnitude: 665 (Lowry) versus 463 (Laemmli). Figure 1 gives the distribution of their double errors according to year. In the first few years after the papers' publication there were no double errors. Most double errors occurred in 1980s and 1990s. These errors

lasted into recent years.

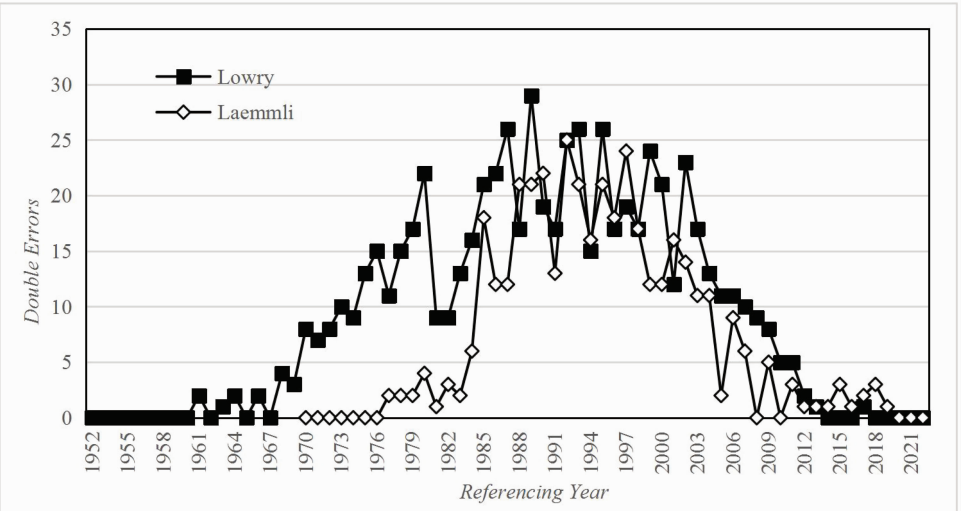


**Figure 1** Volume-page double errors of the two papers

Table 2 and Table 4 list different types of the volume-page double errors (frequency≥ 4) of the two papers. We see that both the error types and the error frequency are almost at the same level. It might be coincidental that the frequency of the top ten types of double errors of the two papers are almost the same. For Lowry' s paper the sequence is 17, 13, 13, 13, 12, 9, 8, 8, 7, 7, totally 107 errors. While for Laemmli' s paper, it is 18, 15, 13, 12, 12, 11, 10, 7, 7, 7, totally 112 errors. That means that their dissemination networks of double errors (frequency ≥ 7) have a similar scale. Thus, it is convenient for us to compare the error dissemination networks of the two papers.

In our previous paper (Liang et al., 2014) we designed the dissemination network of referencing errors of Laemmli' s paper. Figure 2 is an example.

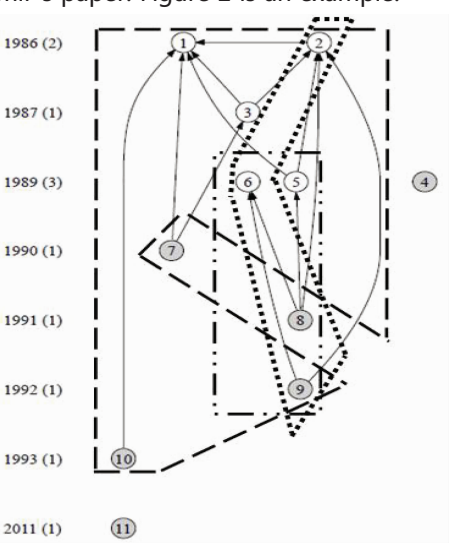


**Figure 2** Dissemination network of the 201-1130 errors (frequency: 11) (Laemmli' s paper)

In the figure, the publication year of the papers is ranked from older (top) to recent (down) and marked on the left side of the figure. Brackets after the publication year contain the number of papers published in that year. A paper's order number is located in a circle. A white circle shows that this paper has been cited by at least one paper in the same figure, thus it might transfer the referencing error. A dark circle indicates that, within the paper group in the same figure, this paper has never been cited. There are different kinds of lines in the figure. A solid line with an arrowhead indicates a citing-cited relationship between two papers. The arrowhead points to the cited paper. If two papers are connected by a dotted line, this means that the two papers have at least one common author. Some regions in the figures are circumscribed by broken lines. Papers within the same region have at least one common author.

Figure 3 to Figure 12 are the dissemination networks of the top ten groups of volume-page double errors (frequency ≥ 7) in Table 2. They are all the error dissemination networks of Lowry's paper. Our former paper (Liang et al., 2014) offered 14 figures on error dissemination networks of Laemmli's paper.

Based on the ten figures we determined that among the papers with double error (frequency ≥ 7) the earliest one appeared in 1971 with the error 226-497, see Figure 5. The latest one was published in 2017 with the error 14-425, see Figure 9. The dissemination of the double errors of Lowry's paper has spanned nearly half a century.

Observing and analyzing the ten figures (Figure 3 to Figure 12), we found that apart from Figure 3 and Figure 6, in the other eight figures, there exist solid lines with arrowheads indicating the citing-cited relation between two papers. That shows that citing and copying other papers' references is still the main route of the error dissemination. That is Route 1, which was put forward in the former study on the referencing error dissemination network of Laemmli's paper.

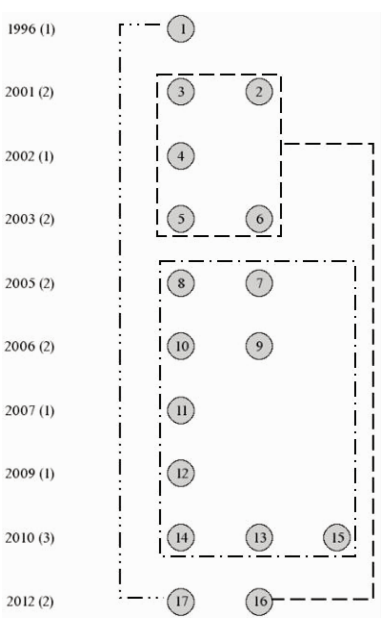


**Figure 3** Dissemination network of the 243-1331 errors (frequency: 17)

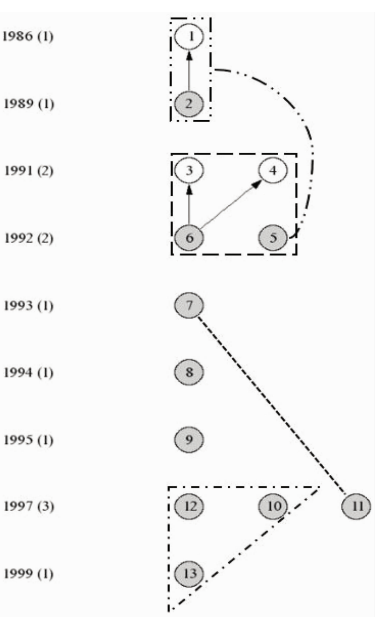


**Figure 4** Dissemination network of the 127-182 errors (frequency: 13)

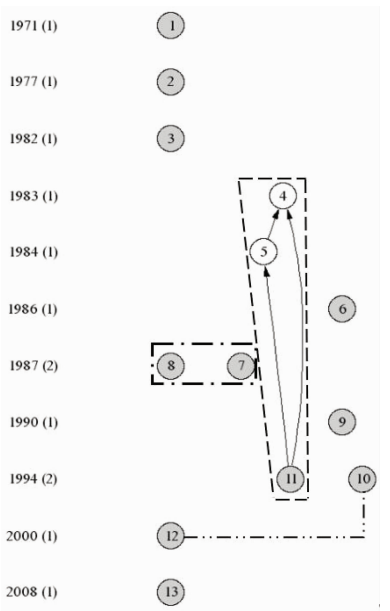


**Figure 5** Dissemination network of the 226-497 errors (frequency: 13)

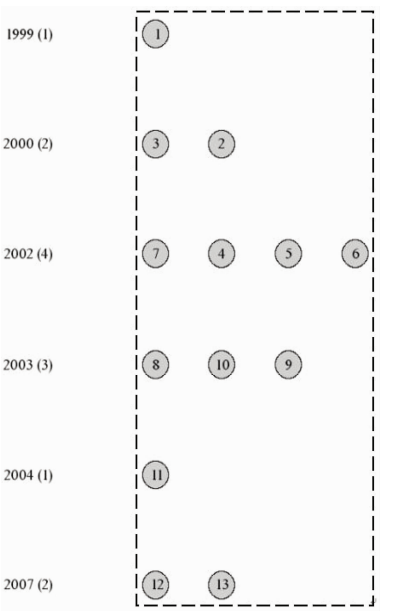


**Figure 6** Dissemination network of the 261-6300 errors (frequency: 13)

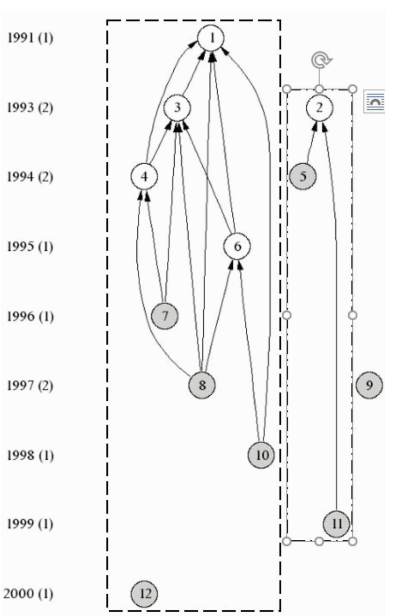


**Figure 7** Dissemination network of the 62-315 errors (frequency: 12)

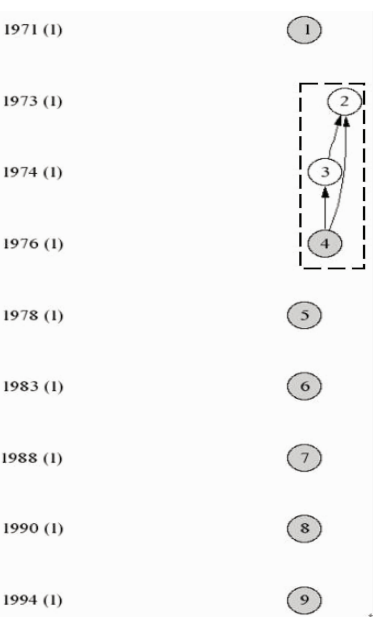


**Figure 8** Dissemination network of the 177-751 errors (frequency: 9)

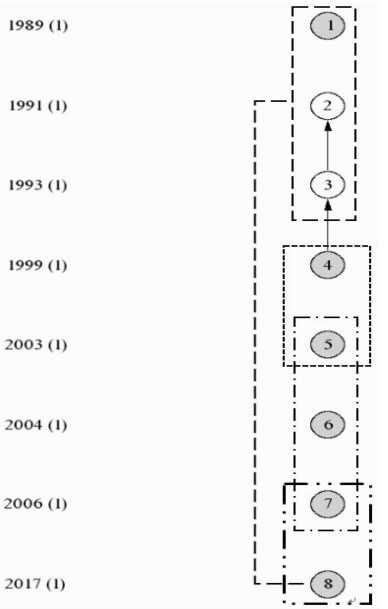


**Figure 9** Dissemination network of the 14-425 errors (frequency: 8)

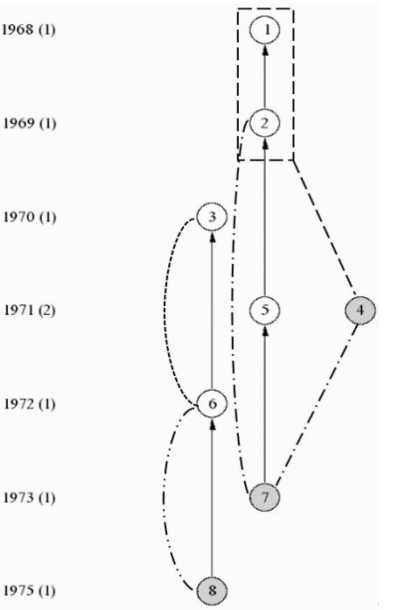


**Figure 10** Dissemination network of the 19-3 errors (frequency: 8)

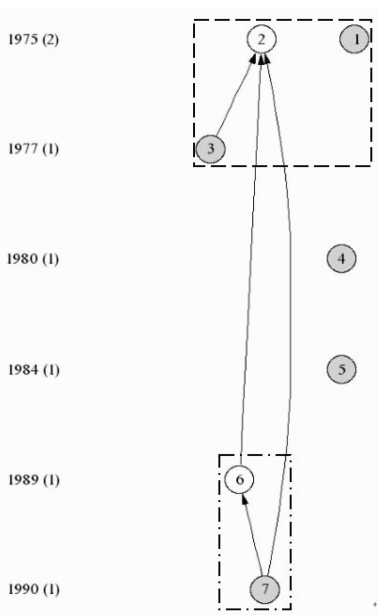


**Figure 11** Dissemination network of the 236-1372 errors (frequency: 7)

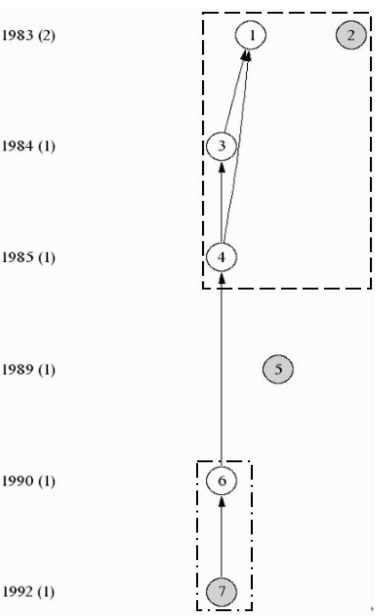


**Figure 12** Dissemination network of the 250-8283 errors (frequency: 7)

In these figures, different kinds of broken lines are striking. They indicate that two papers have at least one common author, or a few papers have at least one common author. In this situation, the author(s) copied references from her or his(their) own earlier paper. That is Route 3, which is summarized from the study on Laemmli' s paper.

In Figure 3 there is no solid line, showing that there is no citing-cited relation between any two papers. Neither citing-cited relation, nor common author, how can paper 2 (or paper 3, published in the same year), and paper 7 (or paper 8) carry the same error as paper 1, which is the initiator of the error 243-1331? Our educated guess is that the author(s) of paper 2 (or paper 3) copied the wrong reference from paper 1, but didn' t put paper 1 into their reference list. So did the author (s) of paper 7 (or paper 8). This is the typical "Route 2. Copying without citing: copying references from other paper without citing that paper." Route 2 was already put forward in our former paper (Liang et al., 2017).

The most puzzling thing occurs in Figure 6. 13 papers carry the same referencing error 261-6300. However, there are no solid lines between any papers. The 13 papers have at least one common author. In 1999 the first paper in this group made the wrong volume-page error 261-6300, then from 1999 to 2007 the other papers copy this referencing errors again and again. This is an extreme case. There is no such case in the dissemination networks of Laemmli' s paper.

## 4 Conclusion and Discussion

Lowry' s paper (1951, *Journal of Biological Chemistry*) is the most highly cited paper in the world. Laemmli' s paper (1970, *Nature*) ranks second. These two papers are selected as our sample to compare their dissemination networks of referencing errors. It seems strange that though the citations to Lowry' s paper are much more than that of Laemmli' s paper, the total referencing errors to Lowry' s paper accounts for only one-tenth of Laemmli' s paper. Both the single volume errors and the beginning page errors of Lowry' s paper are much less than those of Laemmli' s paper. There exists a tremendous difference between the two papers. However, the scales of volume-page double errors of the two papers are in the same order of magnitude: 665 (Lowry) versus 463 (Laemmli). That leads us to construct and compare their dissemination networks of referencing errors. The conclusion is that the three routes of error dissemination of Lowry' s paper are the same as that of Laemmli' s paper. Route 1: copying references from other paper and citing that paper. Route 2: copying references from other paper without citing that paper. Route 3: having at least one common author with another paper that carries the same referencing error. This points to an author copying a reference from one of her or his own earlier papers.

After all, copying references from other papers without checking, especially copying without citing, is a referencing misbehavior. Scientists should fully recognize its widespread presence and try their best to avoid such misbehavior. Anyway, self-management is the duty of science community.

Copying references from the author's previous paper is one of the main routes. Figure 6 is the most convincing illustration. If the authors kept a good habit of checking the references and other necessary things carefully before submitting the manuscript to a journal, there would not be the cases in Figure 6 at all.

We have identified the three disseminating routes of referencing errors. Nevertheless, one may still quite wonder where the initial error came from. Our educated guess is that the author(s) of paper 1, by accident, wrote the wrong volume number and wrong beginning page

number. We have given an example in our former paper (Liang et al., 2014). We warn all authors to take special care in preparing their references list to maintain the correctness of their references. Remember: The devil is in the detail.

## Acknowledgments

The authors would like to thank Professor Wu Yishan for his valuable suggestions on manuscript revision. The work presented in this paper is sponsored by the National Natural Science Foundation of China (Grant Number 71603073).

## References

Ainsworth, S., & Russell, J. M.(2018). Has hosting on science direct improved the visibility of Latin American scholarly journals? A preliminary analysis of data quality. *Scientometrics, 115* (3), 1463–1484. doi: 10.1007/s11192–018–2725–7

Laemmli, U. K.(1970). Cleavage of structural proteins during assembly of head of bacteriophage–t4. *Nature, 227* (5259), 680–685. doi: 10.1038/227680a0

Liang, L. M., & Rousseau, R. (2008). Yield sequences as journal attractivity indicators: "payback times" for Science and Nature. *Journal of Documentation, 64* (2), 229–245. doi: 10.1108/00220410810858038

Liang, L. M., & Zhong, Z. (2007). Scientists´ citation misbehavior reflected by the wrong citation records. *Studies in Dialectics of Nature, 23* (6), 62–65. doi: 10.19484/j.cnki.1000–8934.2007.06.016

Liang, L. M., Zhong, Z., & Rousseau, R. (2014). Scientists´ referencing (mis)behavior revealed by the dissemination network of referencing errors. *Scientometrics, 101* (3), 1973–1986. doi: 10.1007/s11192–014–1275–x

Lowry, O. H., Rosebrough, N. J., Farr, A. L., & Randall, R. J. (1951). Protein measurement with the folin phenol reagent. *Journal of Biological Chemistry, 193* (1), 265–275. doi: 10.1016/S0021–9258 (19)52451–6

Olensky, M., Schmidt, M., & van Eck, N. J. (2016). Evaluation of the citation matching algorithms of CWTS and iFQ in comparison to the Web of Science. *Journal of the Association for Information Science and Technology, 67* (10), 2550–2564. doi: 10.1002/asi.23590

Serenko, A., Dumay, J., Hsiao, P. C. K., & Choo, C. W. (2021). Do they practice what they preach? The presence of problematic citations in business ethics research. *Journal of Documentation, 77* (6), 1304–1320. doi: 10.1108/jd–01–2021–0018

Yang, S.(2011). The Analysis on drawback of citation analysis and its causes. *Journal of Library Science in China, 37* (3), 108–117. doi: 10.13530/j.cnki.jlis.2011.03.013

Fand, Z. Z. (2015). Research on the formation path of patent citations: From the perspective of behavior logic. *Library and Information Service, 59* (19), 15–21+46. doi: 10.13266/j.issn.0252–3116.2015.19.002

Zhang, M. Q., Liu, F., Yao, L., & Cui, J. H. (2018). Analysis on the key steps of quality control in citation search services: Practice and evaluation of citation error judgement. *Journal of Academic Libraries, 36* (5), 93–100. doi: 10.16603/j.issn1002–1027.2018.05.015