

Dear Editor,

I'd like to start by acknowledging the time taken by you and the reviewers to reevaluate this paper. The attached version includes revised and expanded sections (Introduction, Methodology, Discussion) that focused on the three improvements recommended by the editor, namely interpretation of findings, threats to validity, and research questions. I also attempted to address every actionable feedback from Reviewer 2, as detailed in the following point-by-point response to their review.

> The author does not report systematic steps followed for gathering papers' data. For instance, "....several search terms were used to assist in identifying artifacts, such as "github," "gitlab," "bitbucket," "sourceforge," and "zenodo" for repositories; variants of "available," "open source," and "download" for links; and variations of "artifact," "reproducibility," and "will release" for indirect references." Besides, it is not clear how the author checked out whether a paper received an award.

I have now detailed the complete data-collection process in a new subsection of the methodology, which should hopefully clarify how each piece of the dataset was collected.

> At the beginning of the paper, several research questions are presented without a proper justification. These detailed RQ should be addressed and contextualized in the proper section.

Upon consideration, I agree with the reviewer that these questions are presented without proper justification, and do not directly support the main research question of this article, which is, what is the quantitative relationship between artifact sharing and citations in computer systems? Instead, these questions are better described as interesting but minor incidental findings that are worth reporting (in my opinion) but not detracting from the main thrust of the paper. I've reworded the introduction accordingly.

> I understand that the authors followed a partially ad-hoc process for gathering papers' data without double-checking, instead of following a systematic one. If true, the author should explicit these relevant threats to validity. Otherwise, the author should report step-by-step how they reached each data to led the package reproducible.

> I could not find any mention to double-checking on the gathered data. It is relevant at least fo support data gathering through non-systematic steps.

There was actually significant double-checking and data reconciliation process involved, which is now detailed in the new step-by-step data-collection process subsection. Additionally, a new threats-to-validity was added as requested, and it elaborated the challenge and potential weakness of collecting and verifying artifact sharing information.

> What was the normality test applied? p-value? Yet about normality, the authros state that the data was normalized. However, after that the author argue that "...omit the 49 papers with zero citations to improve the linear fit with the predictors." Outliers? And what about those with >2000 citations?

I'm unclear on what confuses the reviewer here. I'll try to response to each question separately.

1. As detailed in the statistics subsection, t-test (and p-value) was reported when comparing means, and a signed rank-test (and p-value) when comparing medians.

2. I could not find a claim in the paper that the data is normalized. The time-since-publication for citations is normalized by measuring citations exactly 3.5 after publications.

3. As for the data being normally-distributed, it is not. The paper claims that the data appears log-normal, and therefore the linear regression uses $\log(\text{citations})$ instead of citations as the response variable. This transformation "takes care" of the outliers on the right tail (the highly-cited papers), It cannot be applied to zero values, and the $\log(p+1)$ transformation would distort the left-tail results too much, in my opinion. I therefore preferred to omit the 49 zero-values but keep the high outliers. To verify, I recomputed the linear model omitting all papers with 1000+ citations and found only negligible differences in the resulting coefficients or significance.

4. For t-tests, a log transformation is not required despite the long tail of the citations distribution because of the size of the sample and the central-limit theorem (the means of the distributions are normally distributed, permitting a t-test).

> "...It is worth noting, however, that most of the source-code repositories in these artifacts showed no development activity—commits, forks, or issues—after the publication of their paper, suggesting limited impact for the artifacts alone". I disagree about this example of poor impact. Please note that commits and forks are typically made for source code, which is not necessarily the case of research data.

The statement is specifically restricted to source-code repositories, not data (the vast majority of the repositories contained source code anyway). However, I replaced the word "impact" with "activity".

> I miss a discussion of the findings towards causality (which seems to be unfeasible) and more statistical tests (that does not help to understand the original results. Some of the complementary analysis seems to address irrelevant papers' features (at least without proper contextualization). For instance: "Incidentally, papers with released artifacts also tend to incorporate significantly more references themselves (mean: 32.31 vs. 28.71; $t = 5.25$, $p < 10^{-6}$)." I definitively could not see why this information is relevant in the context of the research. The same for "using colons" and "paper length," among others.

These factors are relevant because some studies have found them to be related or predictive of higher citations, as listed in the related work section.

> For instance, I did not see some treatment about self-citations.

The papers in our dataset are cited 85,543 times. It would be impractical to try to download all of these citations and then count self-citations, accounting properly for name disambiguation, since much of this process would be manual. So there is no specific analysis of self-citations in this study.

That said, I have no reason to suspect that the self-citation behavior would be different for papers with artifacts released than for those without. If the reviewer has any data or references to suggest otherwise, I would be glad to incorporate it. In the meantime, I added a sentence about it to the threats to validity subsection.

> "Of the 2439 papers, 292 91.8% displayed at some point an accessible link to the full text on GS. What is the impact of these 9.2% over the citations analysis found? Accessibility is definitively relevant in this context.

This is first addressed in the following text (closing paragraph of section 3.3). Slow paper accessibility is associated with fewer citations. In the linear regression, inaccessible papers were assigned "1,000" months time-to-eprint, and this factor was found to be statistically significant but of minimal effect (coefficient) in Table 4.

> "These textual relationships may not be very insightful, because of the difficulty to ascribe any causality them, but they can clue the paper's reader to the possibility of an artifact, even if one is not linked in the paper." I am afraid I could not get the point here.

The effort to identify artifacts associated with papers is both manual and error-prone, as elaborate in the paper. This tidbit is offered to other researchers attempting to collect such data, perhaps in order to reproduce these results. Knowing that certain textual properties are associated with a higher chance of locating a related artifact, even if not linked to from the paper, can focus and accelerate independent searches for such artifacts.

> "As a crude approximation, a simple search for the string "github" in the full-text of all the papers yielded 900 distinct results. Keep in mind, however, that perhaps half of those could be referring to their own artifact rather than another paper's, and that not all cited github repositories do indeed represent paper artifacts." I could not see how this analysis is relevant to support the discussions.

I've removed this analysis from the paper.

> Figure 1 is not cited. Besides, I understand that this figure could be omitted from the paper.

Fig. 1 is cited (2nd paragraph of Sec. 3.1, line 214 in the original revision and now 231). I think it is important to show the wide distribution of sharing rates by conference, and the tendency of ACM conferences to have higher sharing rates, so I left it in place.

> I suggest combining the analysis presented in Figure 2 and Figure 3 into a single one.

Both figures show related information, but on different scales. Since there is no space constraint, I prefer to leave them as is than to create a single figure that is harder to interpret.

> Table 1: please standardize the number of decimal places

Corrected.

> 2439 papers=> 2,439 papers

Corrected throughout.

> "...when skimming these paper*s*..."

Corrected.

> The zenodo package is not cited in the paper.

I've added the DOI to the existing Zenodo package citation (see Code and data availability section).

Sincerely,

Eitan Frachtenberg.