



Peer Comparison of XSEDE Publication Data

Gregor von Laszewki, Fugang Wang, Geoffrey C. Fox - Indiana University
Thomas R. Furlani, Robert L. DeLeon, Steven M. Gallo - University at Buffalo, SUNY



INTRODUCTION

We present a framework that compares the publication impact based on a comprehensive peer analysis of papers produced by scientists using XSEDE resources. The analysis is introducing a percentile ranking of citations of the XSEDE papers compared to peer publications in the same journal that do not use these resources. This analysis is unique in that it compares reported XSEDE publications to their peers from within the same journal issue. From this analysis, we can see that papers produced by utilizing XSEDE resources are cited statistically significantly more often. Hence we find that reported publications indicate that XSEDE resources exert a strong positive impact on scientific research.

Figure 1 shows the high level objectives of our work.

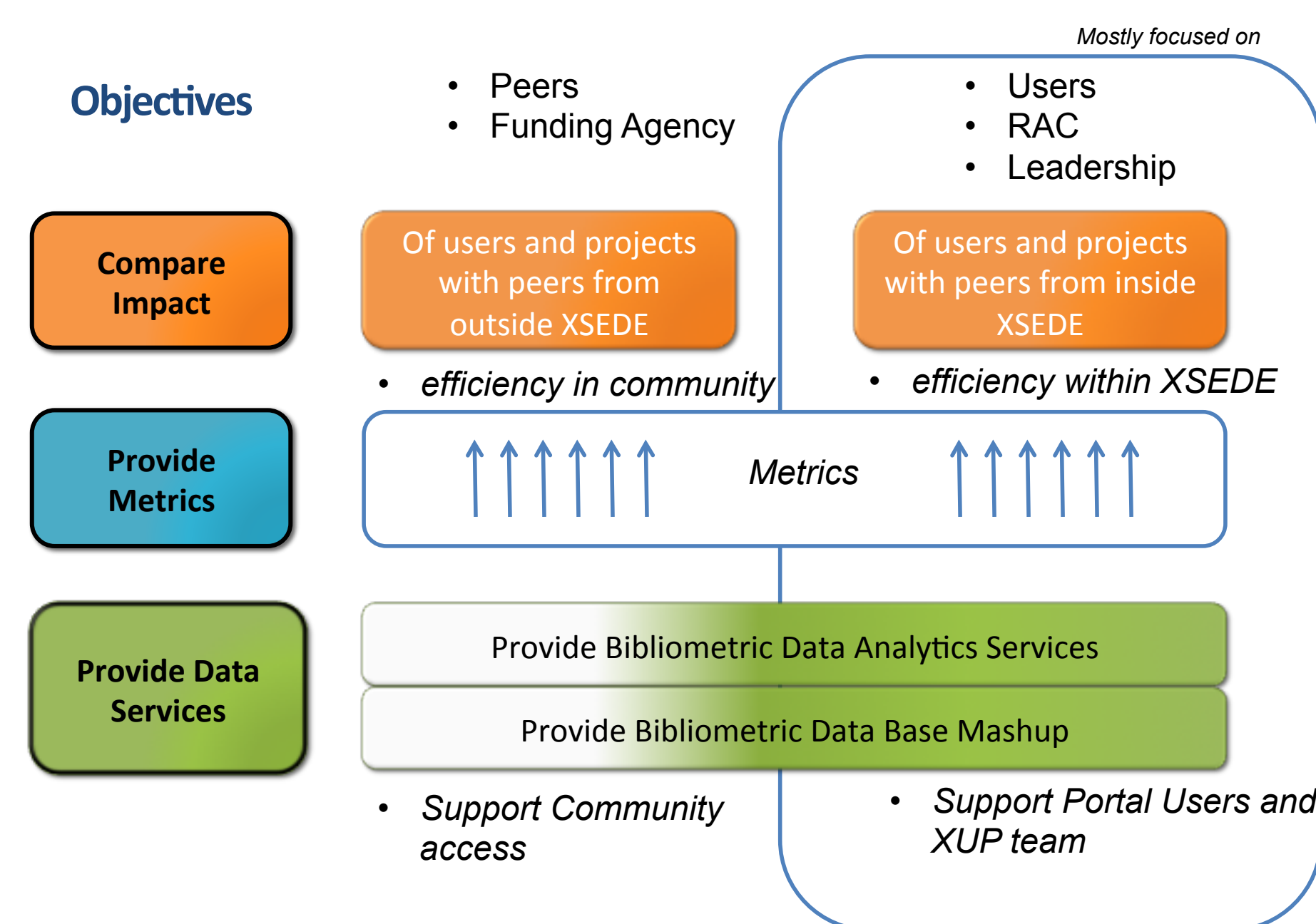


Figure 1. High level objectives

METHOD

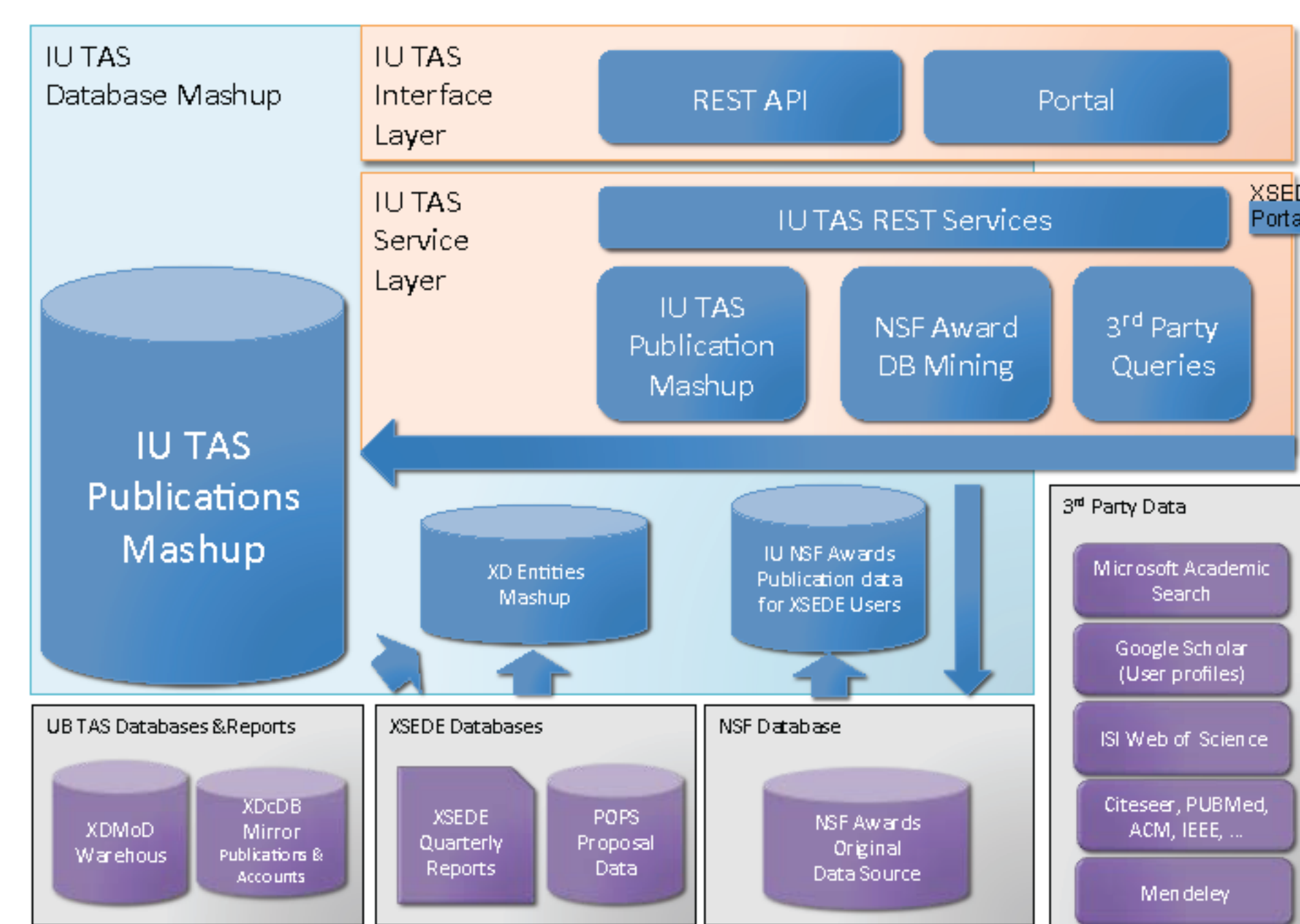


Figure 2. System architecture

We have developed a system to conduct the data gathering, processing, and metrics calculation and evaluation (Figure 2). It is a layered architecture involving various sources and interacting with multiple other services and sites. The method is applied to XSEDE to retrieve relevant data, generate metrics, as well as impact evaluation [1].

We compared XSEDE publications with their peers by looking at the citation count and relative ranking. The process is as the following:

1. Identify publication venues of the reported XSEDE publications;
 2. Retrieve all publication data from those venues during the same time period;
 3. Identify the peers comparison groups (All publications in same journal issues)
 4. Calculate ranking scores of XSEDE publications and their peers;
 5. Use the raw ranking scores to derive comparison metrics at different aggregation level (e.g. journal, Field of Study).
- During this process we have retrieved more than one million publication records to facilitate the peers comparison.

RESULTS

	# Publications	Rank - Average	Rank - Median	# Citation - Average	# Citation - Median
XSEDE	2349	61	65	26	11
Peers	168422	49	48	13	5

Table 1. Comparison data of XSEDE and peers

T-test for ranking (Welch Two sample t-test)

T=21.4134, df=2412.99, p-value<2.2e-16

95% confidence interval: [10.80, 12.98]

T-test for citation count:

T=7.057, df=2358.929, p-value=2.228e-12

95% confidence interval: [9.40, 16.63]

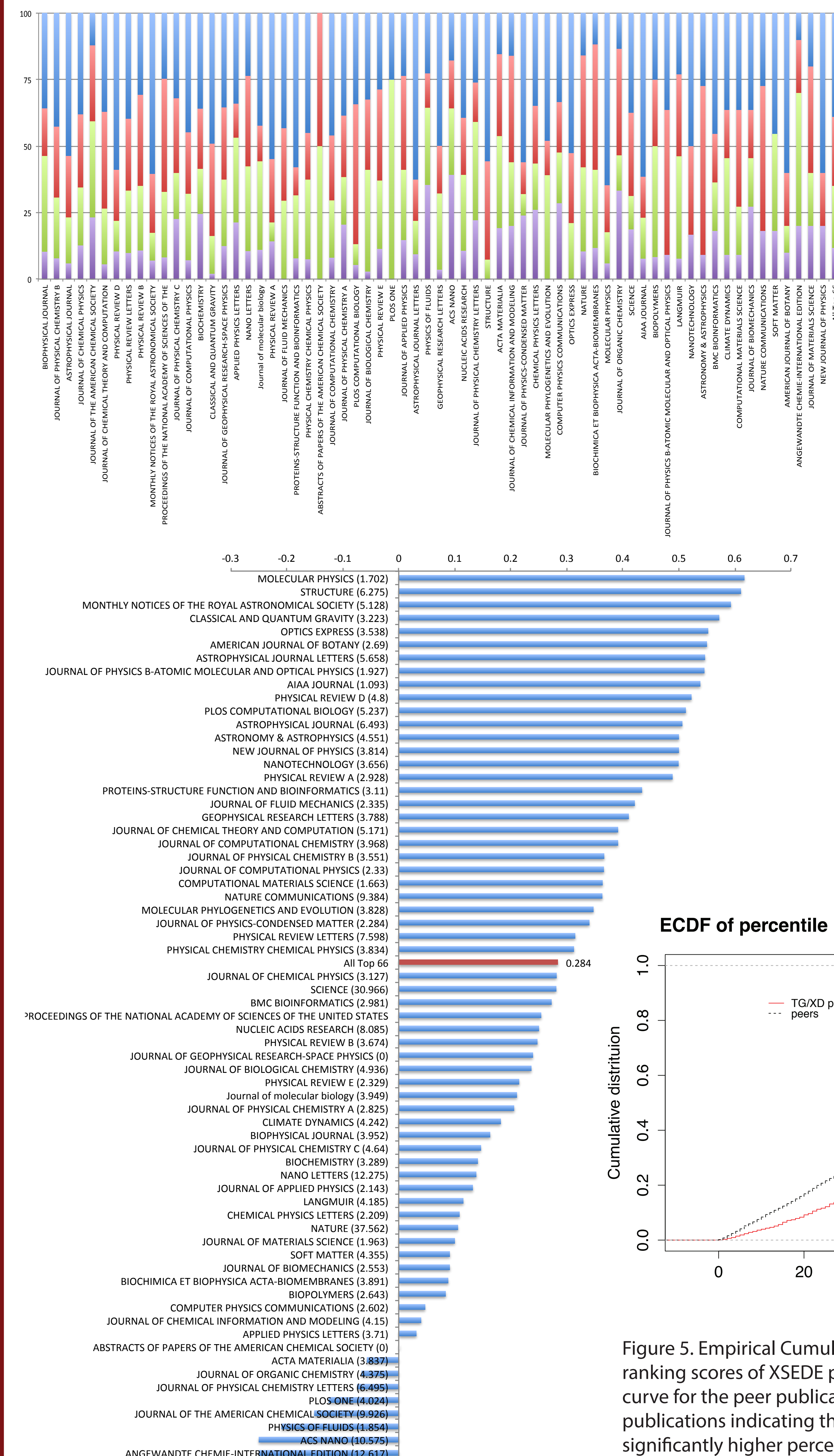


Table 1. lists the numeric results of compared XSEDE publications and the peers. The T-test shows that the result is statistically significant for both the ranking and citation count (average or median). See also Figure 5 (at the bottom).

Figure 3. (on the left) shows the comparison results when grouped by journals. For each journal, the ranking quarter distribution (based on individual percentile rankings) is shown. E.g., BLUE indicates the portion of papers falling into the top quarter by percentile ranking among peers; RED indicates those falling into the 2nd quarter. Altogether they have those falling in the better half. The chart shows XSEDE publications tend to receive more citations in most journals comparing to the peers within the same journal issues.

Figure 4. Publication venues ordered by a merit score defined as:

$$S = 1 * PQ_1 + 0.5 * PQ_2 + (-0.5) * PQ_3 + (-1) * PQ_4$$

where PQ_i is the percentage of pubs falling into the top i -th quarter. This metric enables a quantitative comparison by combining all the percentile ranking information.

ECDF of percentile ranking score for XSEDE pubs and peers

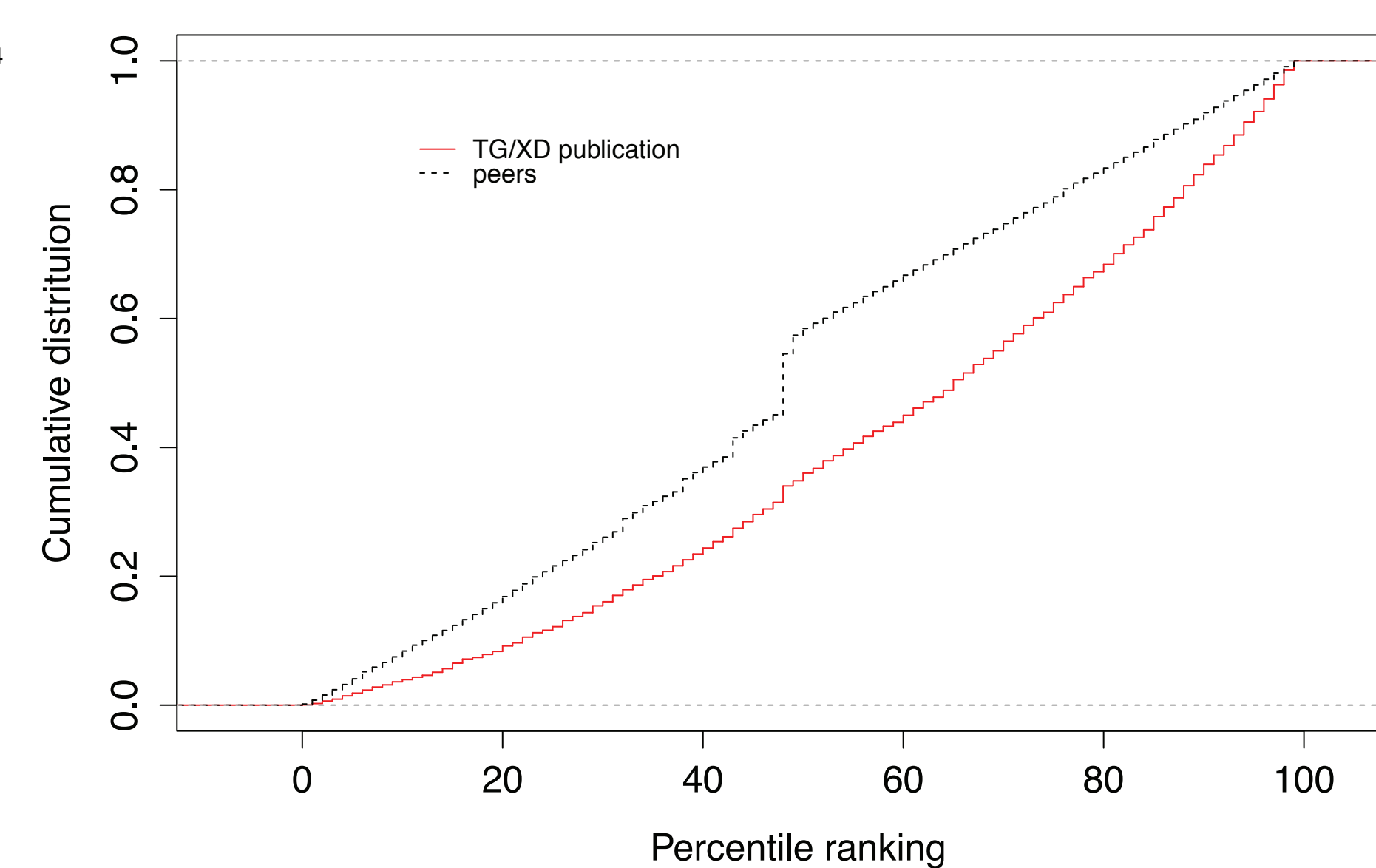


Figure 5. Empirical Cumulative Density Function (ECDF) of percentile ranking scores of XSEDE publications (Red) and peers (Black). The ECDF curve for the peer publications lies well above that of the XSEDE publications indicating that as a group the XSEDE publications have a significantly higher percentile ranking.

When aggregating the results within specific Field of Science (FOS) categories (Figure 6 and 7), and plotting the average/median percentile ranking and merit scores as defined previously, the result shows that for the majority of the FOS the XSEDE publications are receiving more citations and thus having a higher impact comparing to their peers (percentile ranking greater than 50 and positive merit score).

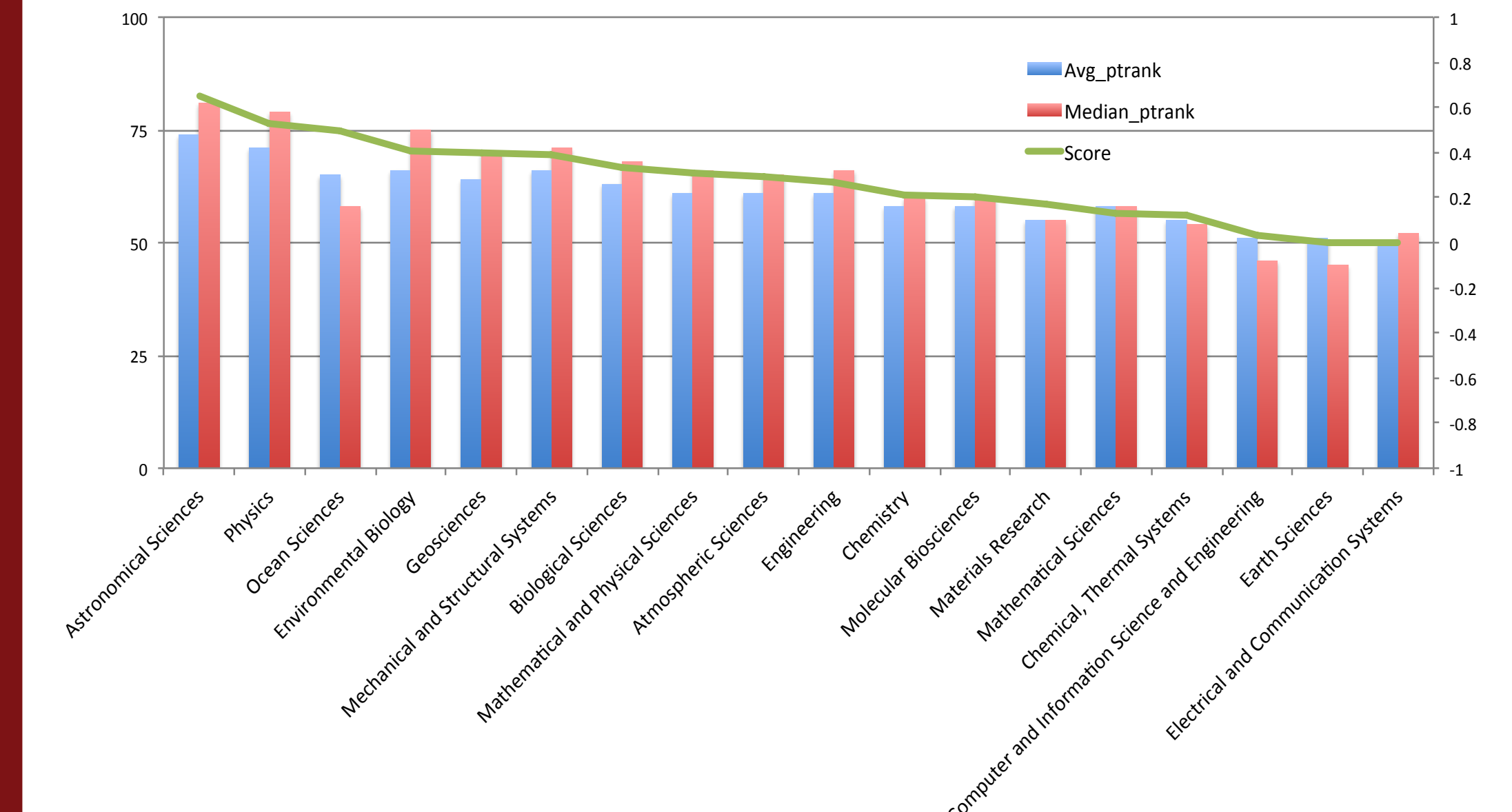


Figure 6. Further aggregation of the Field of Science (FOS) categories shows how publications from different FOS perform regarding the citation count percentile ranking among peers.

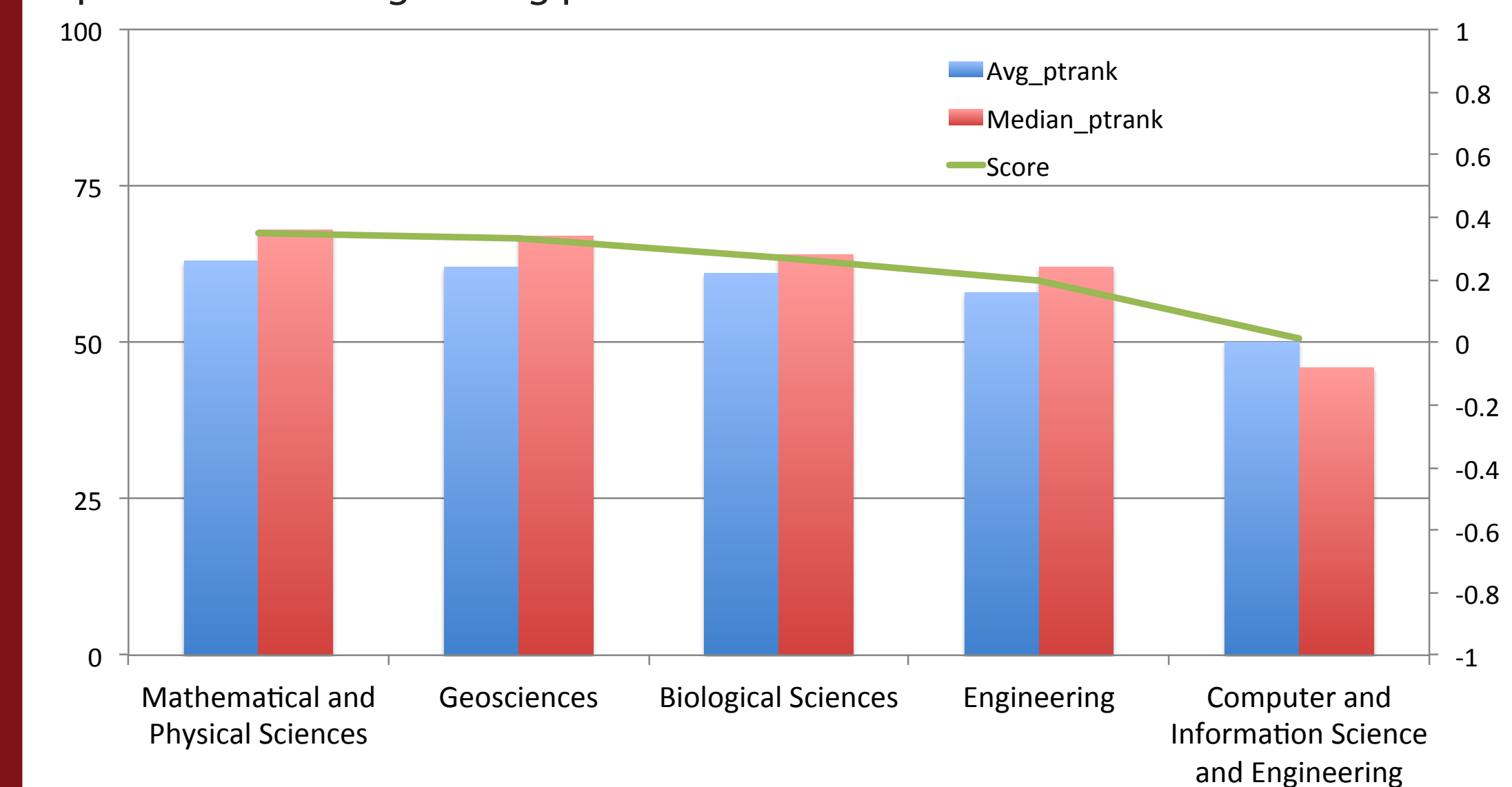


Figure 7. Aggregated at top level FOS.

CONCLUSIONS

We extended our previous work [1] to conduct peers comparison of XSEDE publications. In the new analysis we obtained publication and citation data for XSEDE and peers publications, formed comparison groups based on journal and issue information, and calculated and compared percentile rankings of the involved publications. The results show that XSEDE publications tend to receive more citations compared to their peers, and the result is statistically significant. Aggregations on journal level as well as on Field of Science level suggested the same for the majority of journals and FOS, while certain FOS benefit most (e.g., astronomy and physics).

ACKNOWLEDGEMENTS

This work is part of the Technology Auditing Service (TAS) project sponsored by NSF under grant number OCI-1025159.

REFERENCES

1. Wang, F., von Laszewski, G., Fox, G. C., Furlani, T. R., DeLeon, R. L., & Gallo, S. M. Towards a Scientific Impact Measuring Framework for Large Computing Facilities-a Case Study on XSEDE. In Proceedings of the 2014 Annual Conference on Extreme Science and Engineering Discovery Environment.
2. von Laszewski, G., Wang, F., Fox, G. C., Hart, D. L., Furlani, T. R., DeLeon, R. L., & Gallo, S. M. Peer comparison of XSEDE and NCAR publication data. To be presented in IEEE Cluster'15.