Open sourcing the economics of open source: A comprehensive literature review*

aaron-lohmann

2025-04-02

Abstract

Open source software (OSS) is powering much of our digital infrastructure. From hosting websites, to working with databases and advancement of Machine Learning models. Recently, OSS has received increased attention from the economic discipline. Some work is motivated by the curious incentive structures, the econmic value created through OSS or the available and detailed micro-data. The goal for this project is to openly and publically craft a literature review which summarises our findings about OSS relating to economic ideas. Inspired by OSS itself, this project is open to contribution from anyone. Reward in the form of co-authorship is given according to contribution (see contribution guidelines).

Contributions welcome!

Contribution guidelines:

1. Contributions to this project must be made through a pull request to the repo https://github.com/aaron-lohmann/econ-oss.
2. Position in author listing will be determind by the amount of approved pull requests. Ties between the same amount of pull requests will be broken by the earliest contribution of an author.
3. Co-authorship will be granted upon two approved pull requets. One of these contributions must introduce a new citation.

*Comments, questions, and feedback are welcome and can be submitting as an issue to the repository https://github.com/aaron-lohmann/econ-oss. Alternatively, to: aaron.lohmann@uni-bielefeld.de.

1 Introduction

Open Source Software (OSS) is key to the functioning of the modern digital infrastructure. For economists some parts of OSS propose curious questions. Why do developers contribute? Can open development outpeform closed development? On the other hand, the detailed micro data on individuals allows to study questions that are relevant to a number of knowledge intensive sectors like patenting, R & D and the development of Work from Home. OSS provides the rare oppurunity to track for a large group of workers who works when for how long at what time. Moreover, there are questions related to global integration. How do teams form across space, how much do teams across collaborate with each other? Yet, other another aspect for OSS is the network structure. One example is to study the adoption of technology. This review tries to be comprehensive. From descriptive papers to published datasets and novel and innovati ve insights. In the end, this paper addresses two groups of readers. Those that read for knowledge and look for answers and those that read for understanding, potentially seeking to enter the field of economics of OSS.

2 Theory

Theoretical contributions which directly answer questions about OSS.

3 Datasets

What datasets are out there to study OSS? What are benefits of one versus the other?

Schueller et al. (2022) provide detailed data on one particular ecosystem: Rust and its associated package manager crates.io. The data covers a time period from early 2014 all the way till September 2022. Given the relative youth of Rust, this can be seen as representative of the full history of the language. The data covers commits, issues, pull requests, importantly dependency relationships betwenn packages. Unique in this list of datasets, the authors also provide data from other code hosting platforms outside GitHub: Gitlab, Bitbucket and the likes. This dataset is well suited for more concentrated deep micro data work given the careful construction.

4 The value of OSS

Hoffmann, Nagle, and Zhou (2024) is an effort in measuring the value of OSS. The authors identify $2,000$ ($1,000$ from NPM and $1,000$ from non-NPM) key packages in the OSS infrastructure. For those they map supply side costs of creation and a demand side value based on firms and websites using these packages. Their preferred estimate for the supply side value is 4.15 billion USD and demand side value greater than 8.8 trillion USD.

5 Geographical composition and team formation

Goldbeck (2025) concentrats on software developers in the U.S. Using the self-reported locations, develoepers are matched to 179 economic areas. Running gravity-style regressions, the author finds sizable colocation effects. However, this effect is much smaller than that of traditional inventor networks as for example in patenting. For developers with joint organizational membership or those collaborting on high quality projects, measured by stars, the colocation effect is found to be weaker.

6 Team outcomes

Betti et al. (2024) study team dynamics with a special emphasis on leaders. Builiding on the data by Schueller et al. (2022) they identify leaders of projects and reveal that often the workload is highly heterogeneous. In teams of 5 to 7, the leader roughly does 80 per cent of the commits. More hetereogenous effort distribution is correlated with higher success. Finally, they study changes in leaderships and reveal that subsequent to a change in leadership the success of the project increases.

7 References

Betti, Lorenzo, Luca Gallo, Johannes Wachs, and Federico Battiston. 2024. "The Dynamics of Leadership and Success in Software Development Teams." *arXiv Preprint arXiv:2404.18833*.

Goldbeck, Moritz. 2025. "Bit by Bit: Colocation and the Death of Distance in Software Developer Networks." *Journal of Economic Geography*, lbaf002.

Hoffmann, Manuel, Frank Nagle, and Yanuo Zhou. 2024. "The Value of Open Source Software." *Harvard Business School Strategy Unit Working Paper*, no. 24-038.

Schueller, William, Johannes Wachs, Vito D. P. Servedio, Stefan Thurner, and Vittorio Loreto. 2022. "Evolving Collaboration, Dependencies, and Use in the Rust Open Source Software Ecosystem." *Scientific Data* 9 (1): 703. https://doi.org/10.1038/s41597-022-01819-z.