Diagnosing Git Repository Health

Tyler Brown
NUID: 001684955

1 Introduction

Understanding the health of open source projects is important to industry because it helps them assess risk associated with their technology stack. Forecasting health is important to investors because this information can help them make profitable investments in open source. Academia is interested to know if there are links between theory, such as programming language design [1], and the health of open source software. A previous metric used to assess open source software health is the Truck Factor [2]. Truck Factor is the smallest subset size of developers who contributed 50% of the code in an open source project. The underlying intuition is that open source projects with a lower Truck Factor are more susceptible to project disruption in the event of adverse circumstances.

By borrowing from Physics, this paper contributes to the advancement of understanding the health of open source projects with the following:

- Introducing a health measure that can be assessed at time $t_{i \pm k}$ where k is a multiple of i.
- New health measure can be used in forecasting as well as description
- Health measure is rooted in Physics so we can derive related measures using preexisting theory.

2 Related Work

Related to truck factor

3 Methods

ok

4 Experiment

5 Discussion

yup

References

[1] Baishakhi Ray, Daryl Posnett, Vladimir Filkov, and Premkumar Devanbu. A large scale study of programming languages and code quality in github. In *Proceedings of the 22nd ACM SIGSOFT International Symposium on Foundations of Software Engineering*, pages 155–165. ACM, 2014.

[2] Guilherme Avelino, Marco Tulio Valente, and Andre Hora. What is the truck factor of popular github applications? a first assessment. Technical report, PeerJ PrePrints, 2015.