Gaining Popularity

Yegor Bugayenko

Lecture #8 out of 8
80 minutes

The slidedeck was presented by the author in this YouTube Video

A quick recap of the previous lectures:

L1: Be nice, say “Please,” “Thanks,” and “Sorry”

L2: Expect and enjoy bug reports

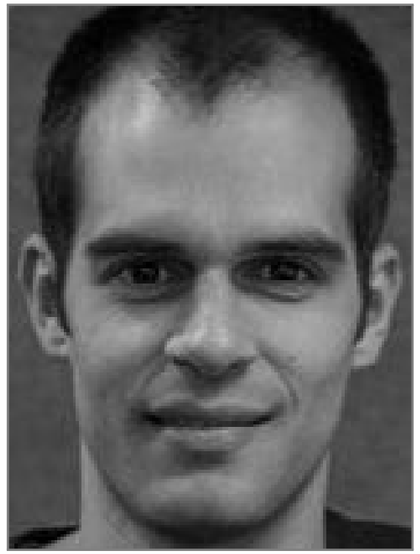
L3: Make pull requests to others — boost your profile

L4: Prevent chaos by reviewing pull request carefully

L5: Make a nice `README` and use MIT license

L6: Setup many GitHub Actions jobs

L7: Make frequent SemVer releases



Andre Hora

"We found that general models, i.e., models produced using the top GitHub repositories, start to provide accurate predictions when they are trained with data from six months and used to predict the number of stars six months ahead."

— Hudson Borges, Andre Hora, and Marco Tulio Valente. Predicting the Popularity of GitHub Repositories. In *Proceedings of the 12th International Conference on Predictive Models and Data Analytics in Software Engineering*, pages 1–10, 2016b. doi:10.1145/2972958.2972966

Repository	Stars	Predicted	% Diff
JQUERY/JQUERY	6,160	5,369	-12.84
ROBBYRUSSELL/OH-MY-ZSH	13,536	11,829	-12.61
AIRBNB/JAVASCRIPT	17,026	14,882	-12.59
H5BP/HTML5-BOILERPLATE	4,896	4,691	-4.19
METEOR/METEOR	9,919	10,082	+1.64
TORVALDS/LINUX	10,566	9,682	-8.37
DANEDEN/ANIMATE.CSS	10,492	9,452	-9.91
FACEBOOK/REACT-NATIVE	18,443	19,373	+5.04
RAILS/RAILS	5,701	5,128	-10.05
DOCKER/DOCKER	10,268	9,721	-5.33

Source: Hudson Borges, Andre Hora, and Marco Tulio Valente. Predicting the Popularity of GitHub Repositories. In *Proceedings of the 12th International Conference on Predictive Models and Data Analytics in Software Engineering*, pages 1–10, 2016b. doi:10.1145/2972958.2972966

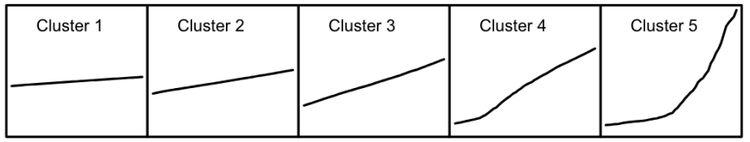


Figure 5: Five growth trends (clusters) identified for the repositories in our dataset

Table 2: Popularity Trends Description

Cluster	# Repositories	% Growth
C1	2,087 (49.1%)	19.9
C2	1,456 (34.2%)	61.3
C3	521 (12.2%)	175.1
C4	131 (3.0%)	883.2
C5	53 (1.2%)	1,659.1

Marco Tulio Valente

"We report that three out of four developers consider the number of stars before using or contributing to a GitHub project."

— Hudson Borges and Marco Tulio Valente. What's in a GitHub Star? Understanding Repository Starring Practices in a Social Coding Platform. *Journal of Systems and Software*, 146(1):112–129, 2018. doi:10.1016/j.jss.2018.09.016

Felipe Fronchetti

"We found that popularity of the project (in terms of stars), time to review pull requests, project age, and programming languages are the factors that best explain the newcomers' growth patterns."

— Felipe Fronchetti, Igor Wiese, Gustavo Pinto, and Igor Steinmacher. What Attracts Newcomers to Onboard on OSS Projects? TL;DR: Popularity. In *Proceedings of the 15th IFIP WG of the International Conference on Open Source Software*, pages 91–103. Springer, 2019. doi:10.1007/978-3-030-20883-7_9

Table 3: Ranking of the most important factors

Ranking	Factor	Score
1	# of Stars	0.1753
2	Time to merge	0.1535
3	# of Languages	0.1278
4	Age	0.1027
5	# of Integrators	0.0995
6	Main Language	0.0946
7	Domain	0.0708
8	Has contributing	0.0396
9	Has wiki	0.0308
10	Has issues template	0.0260
11	Owner type	0.0252
12	Has license	0.0236
13	Has PR template	0.0183
14	Has code of conduct	0.0118

Source: Felipe Fronchetti, Igor Wiese, Gustavo Pinto, and Igor Steinmacher. What Attracts Newcomers to Onboard on OSS Projects? TL;DR: Popularity. In *Proceedings of the 15th IFIP WG of the International Conference on Open Source Software*, pages 91–103. Springer, 2019. doi:10.1007/978-3-030-20883-7_9

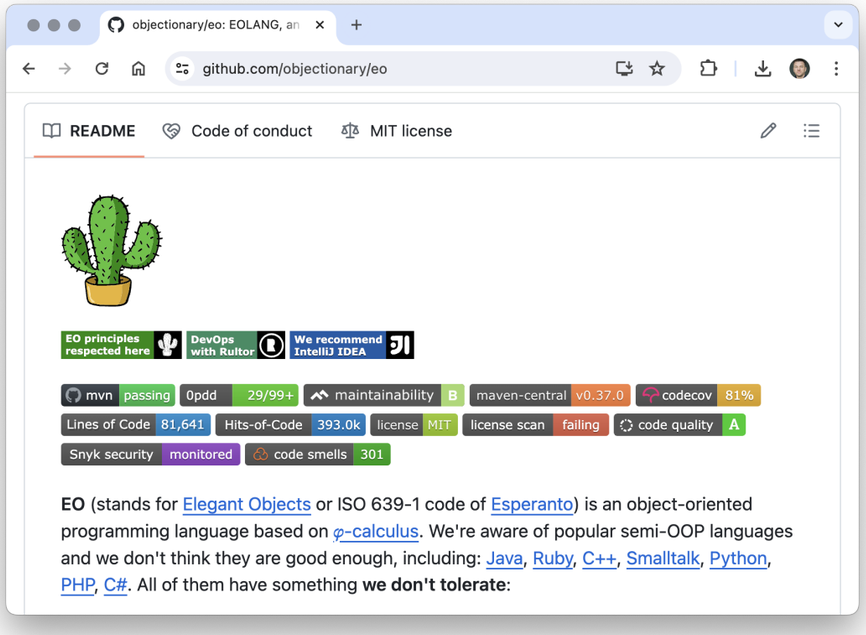
"Popularity of the project (in terms of stars), time to review pull requests, and project characteristics like age and programming languages are the factors that best explain the newcomers' growth patterns. In addition, GitHub recommended community standards have a lower influence on the observed growth patterns."

Burak Rurhan

"We find that unit tested projects have positive correlation with the open-source project metrics and have a higher acceptance rate of pull requests."

— Han Wang, Sijia Yu, Chunyang Chen, Burak Turhan, and Xiaodong Zhu. Beyond Accuracy: An Empirical Study on Unit Testing in Open-Source Deep Learning Projects. *ACM Transactions on Software Engineering and Methodology*, 33(4):1–22, 2024. doi:10.1145/3638245

1. Put some badges.

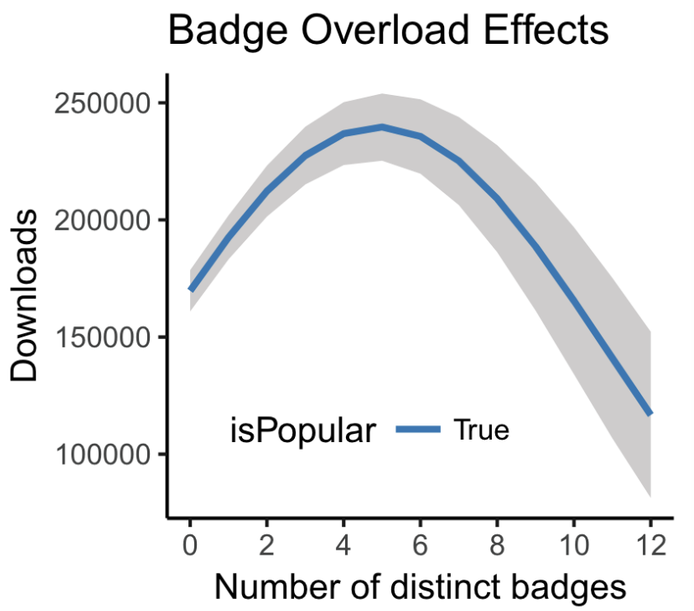


https://github.com/objectionary/eo

Asher Trockman

"A vast majority (88%) agree with the statement 'I consider the presence of badges in general to be an indicator of project quality.'"

— Asher Trockman, Shurui Zhou, Christian Kästner, and Bogdan Vasilescu. Adding Sparkle to Social Coding: An Empirical Study of Repository Badges in the *npm* Ecosystem. In *Proceedings of the 40th International Conference on Software Engineering*, pages 511–522, 2018. doi:10.1145/3180155.3180209



"Packages with many badges tend to have fewer downloads. The effect for less popular packages is negligible."

Source: Asher Trockman, Shurui Zhou, Christian Kästner, and Bogdan Vasilescu. Adding Sparkle to Social Coding: An Empirical Study of Repository Badges in the *npm* Ecosystem. In *Proceedings of the 40th International Conference on Software Engineering*, pages 511–522, 2018. doi:10.1145/3180155.3180209

2. Keep the momentum.

Hudson Borges

"We concluded that repositories have a tendency to receive more stars right after their first public release. After this period, for half of the repositories the growth rate tends to stabilize."

— Hudson Borges, Andre Hora, and Marco Tulio Valente. Understanding the Factors That Impact the Popularity of GitHub Repositories. In *Proceedings of the International Conference on Software Maintenance and Evolution (ICSME)*, pages 334–344, 2016a. doi:10.1109/ICSME.2016.31

Fang Hongbo

"We note a sizeable group of people who follow others on GitHub and tweet about these people's work, but do not otherwise contribute to those open-source projects."

— Hongbo Fang, Daniel Klug, Hemank Lamba, James Herbsleb, and Bogdan Vasilescu. Need for Tweet: How Open Source Developers Talk About Their GitHub Work on Twitter. In *Proceedings of the 17th International Conference on Mining Software Repositories*, pages 322–326, 2020. doi:10.1145/3379597.3387466

3. Keep it up.

Tony Ammeter

"It appears that vitality has a significant effect on popularity over time, indicating that the more active a project is in terms of posting new releases and making announcements, the more attention it receives from the community."

— Katherine Stewart and Tony Ammeter. An Exploratory Study of Factors Influencing the Level of Vitality and Popularity of Open Source Projects, 2002

4. Make some noise.

Some Places to Announce:

- Reddit
- HackerNews
- DZone
- StackOverflow
- Telegram Groups
- Twitter
- Your blog

Marco Tulio Valente

"We reveal that Twitter, user meetings, and blogs are the most common promotion channels used by the studied projects."

— Hudson Silva Borges and Marco Tulio Valente. How Do Developers Promote Open Source Projects? *Computer*, 52(8):27–33, 2019. doi:0.1109/MC.2018.2888770

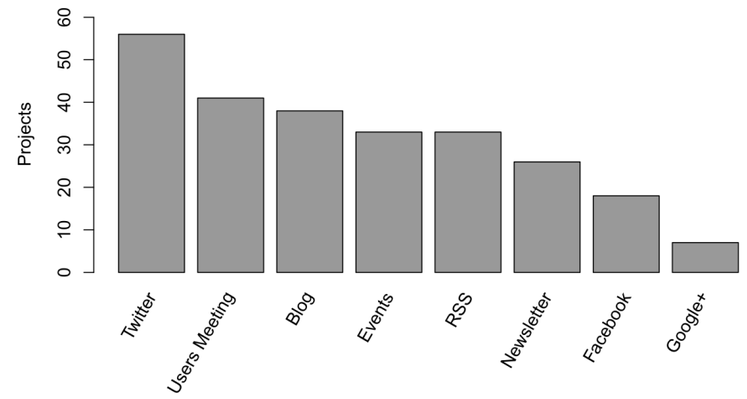


Figure 2: Most common promotion channels

Source: Hudson Silva Borges and Marco Tulio Valente. How Do Developers Promote Open Source Projects? *Computer*, 52(8):27–33, 2019. doi:0.1109/MC.2018.2888770

The Figure presents the most common promotion channels used by the top-100 projects on GitHub. The most common channel is Twitter, which is used by 56 projects. The second one is Users Meetings (41 projects), followed by Blogs (38 projects), Events (33 projects), and RSS feeds (33 projects).

Hemank Lamba

"We find that tweets have a statistically significant and practically sizable effect on obtaining new stars and a small average effect on attracting new contributors."

— Hongbo Fang, Hemank Lamba, James Herbsleb, and Bogdan Vasilescu. "This Is Damn Slick!" Estimating the Impact of Tweets on Open Source Project Popularity and New Contributors. In *Proceedings of the 44th International Conference on Software Engineering*, pages 2116–2129, 2022. doi:10.1145/3510003.3510121

Bibliography

Hudson Borges and Marco Tulio Valente. What's in a GitHub Star? Understanding Repository Starring Practices in a Social Coding Platform. *Journal of Systems and Software*, 146(1):112–129, 2018. doi:10.1016/j.jss.2018.09.016.

Hudson Borges, Andre Hora, and Marco Tulio Valente. Understanding the Factors That Impact the Popularity of GitHub Repositories. In *Proceedings of the International Conference on Software Maintenance and Evolution (ICSME)*, pages 334–344, 2016a. doi:10.1109/ICSME.2016.31.

Hudson Borges, Andre Hora, and Marco Tulio Valente. Predicting the Popularity of GitHub Repositories. In *Proceedings of the 12th International Conference on Predictive Models and Data Analytics in Software Engineering*, pages 1–10, 2016b. doi:10.1145/2972958.2972966.

Hudson Silva Borges and Marco Tulio Valente. How Do Developers Promote Open Source Projects? *Computer*, 52(8):27–33, 2019. doi:0.1109/MC.2018.2888770.

Hongbo Fang, Daniel Klug, Hemank Lamba, James Herbsleb, and Bogdan Vasilescu. Need for Tweet: How Open Source Developers Talk About Their GitHub Work on Twitter. In *Proceedings of the 17th International Conference on Mining Software Repositories*, pages 322–326, 2020. doi:10.1145/3379597.3387466.

Hongbo Fang, Hemank Lamba, James Herbsleb, and Bogdan Vasilescu. "This Is Damn Slick!" Estimating the Impact of Tweets on Open Source Project Popularity and New Contributors. In *Proceedings of the 44th International Conference on Software Engineering*, pages 2116–2129, 2022. doi:10.1145/3510003.3510121.

Felipe Fronchetti, Igor Wiese, Gustavo Pinto, and Igor Steinmacher. What Attracts Newcomers to Onboard on OSS Projects? TL;DR: Popularity. In *Proceedings of the 15th IFIP WG of the International Conference on Open Source Software*, pages 91–103. Springer, 2019. doi:10.1007/978-3-030-20883-7_9.

Katherine Stewart and Tony Ammeter. An Exploratory Study of Factors Influencing the Level of Vitality and Popularity of Open Source Projects, 2002.

Asher Trockman, Shurui Zhou, Christian Kästner, and Bogdan Vasilescu. Adding Sparkle to Social Coding: An Empirical Study of Repository Badges in the *npm* Ecosystem. In *Proceedings of the 40th International Conference on Software Engineering*, pages 511–522, 2018. doi:10.1145/3180155.3180209.

Han Wang, Sijia Yu, Chunyang Chen, Burak Turhan, and Xiaodong Zhu. Beyond Accuracy: An Empirical Study on Unit Testing in Open-Source Deep Learning Projects. *ACM Transactions on Software Engineering and Methodology*, 33(4):1–22, 2024. doi:10.1145/3638245.