

Hirsch index for OSS

Pavel Aitkulov

ajtkulov@gmail.com

May 13, 2025

$$6 > 4$$

Better?

6 ? 4

Better?

6 ? 4
It depends

banana ? apple

banana ? apple
'banana' > 'apple'

banana ? apple
'banana' > 'apple'
banana ? Sunday

banana ? apple
'banana' > 'apple'
banana ? Sunday
multidimensional, embeddings

We want to hire best engineers!

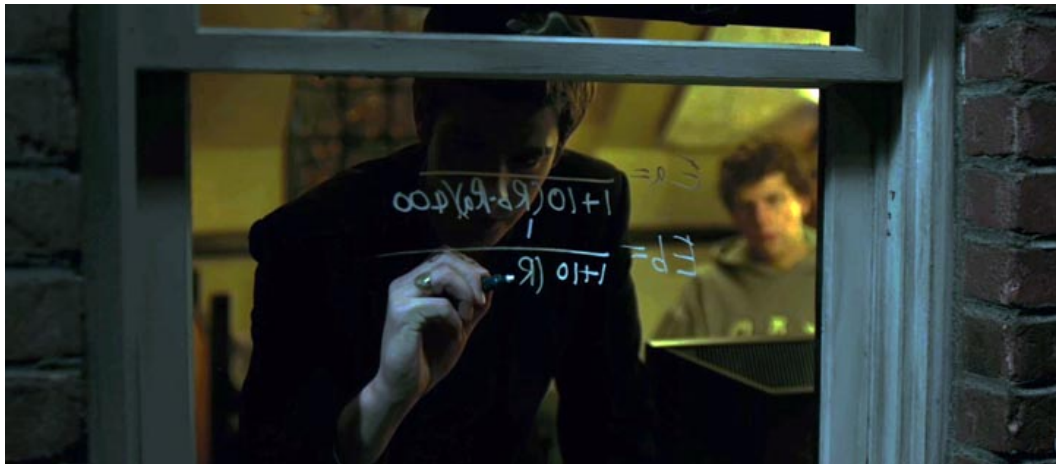
- ① Lines of code
- ② Commits
- ③ PR
- ④ Resolved issues
- ⑤ Different projects

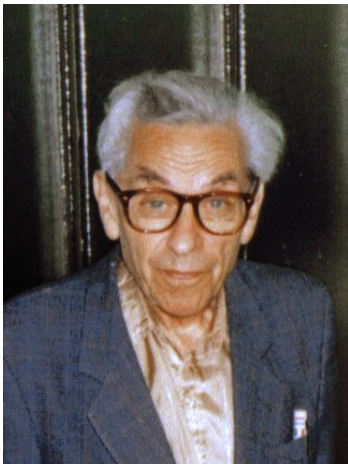
We want to hire best engineers!

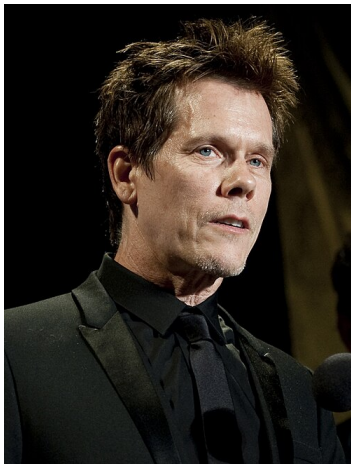
- ① Lines of code
- ② Commits
- ③ PR
- ④ Resolved issues
- ⑤ Different projects
- ⑥ Not an asshole

We want to hire best engineers!

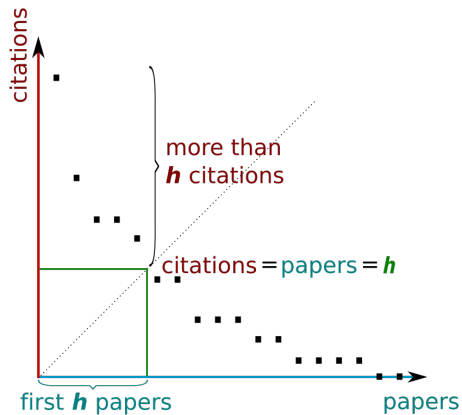
- ① Lines of code
- ② Commits
- ③ PR
- ④ Resolved issues
- ⑤ Different projects
- ⑥ Not an asshole
- ⑦ For a little money, and fire them in one day







Combination of Erdős + Bacon + ...



Strong correlation between H-index and Nobel Prize winners.

<https://stats.getprog.ai/git?profile=jdegoes>

<http://localhost:3000/jdegoes>

<https://hirsch-ui.vercel.app/github/torvalds>

<https://stats.getprog.ai/repo?repo=apache/spark>

<https://stats.getprog.ai/git?profile=jdegoes>

<http://localhost:3000/jdegoes>

<https://hirsch-ui.vercel.app/github/torvalds>

<https://stats.getprog.ai/repo?repo=apache/spark>

Github stats: 7M non trivial H-index, 400M repos, 120M profiles

OSS index doesn't mean ...



<https://www.npmjs.com/package/is-number> 72M downloads weekly

<https://arxiv.org/pdf/2412.13459> 4.5 Million (Suspected) Fake Stars in GitHub: A Growing Spiral of Popularity Contests, Scams, and Malware

Fake it till ...

- 1 Skripal poisoning case, March 2018
- 2 Downloads, views, likes, stars (top instagram post)
- 3 H-index
- 4 Citations

- ❶ Fake libs in Maven Central
https://mvnrepository.com/artifact/com.pvful/cats-effect_sjs1_3/5.1 and
https://mvnrepository.com/artifact/com.typesafe.akka/shapeless_2.9.2/1.9 [404 both]
- ❷ Fake accounts on Github, Linkedin, botnets
- ❸ What if I get access to <https://www.npmjs.com/package/is-number> [70M] and add malicious code?
- ❹ What if I bought a fake Github stars for competitors

- 1 Github
- 2 Stackoverflow
- 3 Linkedin
- 4 Kaggle
- 5 Maven Central, Pypi, NPM
- 6 DockerHub (I know the link)

About myself

- ① H-index = 1
- ② OSS H-index = 0
- ③ Marathon PB = 3h22m
- ④ Chess Elo (lichess.org, blitz, highest) = 2118

About myself

- ① H-index = 1
- ② OSS H-index = 0
- ③ Marathon PB = 3h22m
- ④ Chess Elo (lichess.org, blitz, highest) = 2118
- ⑤ Bacon + Adult film = 0, in my hope

About others

- 1 Demis Hassabis
- 2 Martin Odersky
- 3 John M. Jumper
- 4 Matei Zaharia

- 1 Comparison of high-dimensional things is hard and wrong

Conclusion

- 1 Comparison of high-dimensional things is hard and wrong
- 2 Classes, be alike - better

Conclusion

- ① Comparison of high-dimensional things is hard and wrong
- ② Classes, be alike - better
- ③ you can fake a lot
- ④ faking everything is expensive

Conclusion

- 1 Comparison of high-dimensional things is hard and wrong
- 2 Classes, be alike - better
- 3 you can fake a lot
- 4 faking everyting is expensive
- 5 some animals are more equal than others

Conclusion

- 1 Comparison of high-dimensional things is hard and wrong
- 2 Classes, be alike - better
- 3 you can fake a lot
- 4 faking everyting is expensive
- 5 some animals are more equal than others
- 6 some repos are more equal than others

Conclusion

- 1 Comparison of high-dimensional things is hard and wrong
- 2 Classes, be alike - better
- 3 you can fake a lot
- 4 faking everyting is expensive
- 5 some animals are more equal than others
- 6 some repos are more equal than others
- 7 punish bad guys, appreciate good one

Conclusion

- 1 Comparison of high-dimensional things is hard and wrong
- 2 Classes, be alike - better
- 3 you can fake a lot
- 4 faking everything is expensive
- 5 some animals are more equal than others
- 6 some repos are more equal than others
- 7 punish bad guys, appreciate good one
- 8 to filter out an asshole is hard problem

That's all



Actually not, Quiz

