

Evidence for the Pareto Principle in Open Source Software Activity

Mathieu Goeminne & Tom Mens





- Present our last publication
- Get your opinion

Université de Mons Mathieu Goeminne & Tom Mens

We study the evolution of persons involved in OSS. Questions:

- Is the development activity balanced?
- Is the balance/imbalance evolving over time?
- Is there a bus factor?

Empirical study

- 3 OS projects : Brasero, Evince, Wine
- 3 notions of activity : #commits, #mails, #BR changes
- 3 aggregate metrics : Theil, Gini, Hoover

The aggregate metrics are similar (Evince Commits)



Université de Mons Mathieu Goeminne & Tom Mens

Wine has a different behavior (typical in case of fork/migration/ incomplete data source?)



Université de Mons Mathieu Goeminne & Tom Mens

Conclusion

- For each criteria we studied, the activity effort tends to be more and more inequal.
- Projects are quickly inequals.
- Maybe a stabilization after a time.

The Pareto principle is respected (cumulative ci, ml, br in Evince)



Université de Mons Mathieu Goeminne & Tom Mens



Université de Mons Mathieu Goeminne & Tom Mens

Conclusion

- We need good algorithms able to detect and merge *real* persons.
- Most active persons are active in 2 (or more) activies.
- Less structured projects (Wine) make most active persons less visible.

General conclusion

- The Pareto principle is respected
- After a short time, project activities are strongly imbalanced
- Locally most active persons generally form a bus factor, because they are responsable of the most part of the global activity.

Open questions

- Are the most active persons the same over time? Is the core team the same over time?
- Are distributions strictly respect a Pareto/ power law?
- Can we generalize our observations for all FLOSS? And for closed-source software?

Thank you

Université de Mons Mathieu Goeminne & Tom Mens