Evidence for the Pareto Principle in Open Source Software Activity

Mathieu Goeminne & Tom Mens
Goals

- Present our last publication
- Get your opinion
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)
Wine has a different behavior (typical in case of fork/migration/incomplete data source?)
Conclusion

• For each criteria we studied, the activity effort tends to be more and more inequal.
• Projects are quickly inequal.
• Maybe a stabilization after a time.
The Pareto principle is respected (cumulative ci, ml, br in Evince)
Study of main stakeholders (Nov. 2010)

Brasero
Evince
Wine

committers
mailers

bug report changers

committers
mailers

committers
mailers

bug report changers

Brasero
Evince
Wine

committers
mailers

bug report changers
Conclusion

• We need good algorithms able to detect and merge real persons.

• Most active persons are active in 2 (or more) activities.

• 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 responsible 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