

Paid vs Volunteer Work in Open Source Projects

**Dirk Riehle, Philipp Riemer,
Carsten Kolassa, Michael Schmidt
Friedrich-Alexander University**

HICSS 47 – 2014-01-08

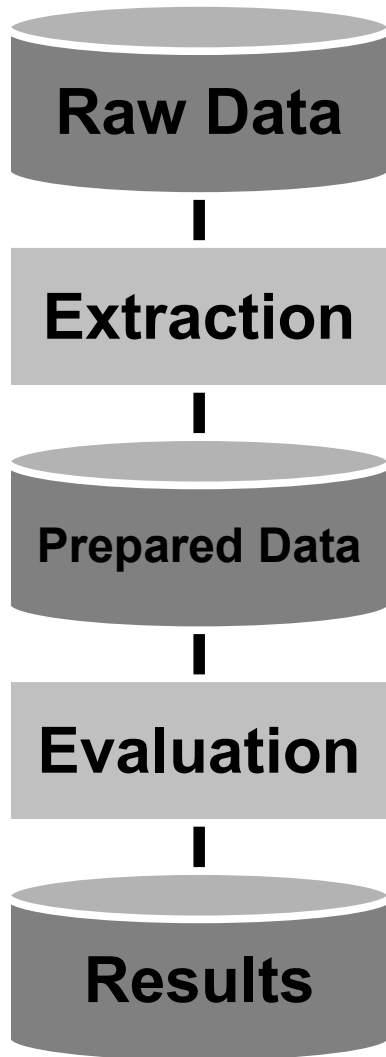
Research Question

- 1. How much commercial contribution (to open source projects) is healthy?**
- 2. How much commercial contribution (to open source projects) is there?**

Research Approach

- 1. Empirical assessment**
- 2. Descriptive analysis**
- 3. Exploratory modeling**
- 4. Dimensional break-out**

Research Process



- Raw data source
 - Local database (ohloh.net, crawled sources)
 - Web services access (ohloh.net, sf.net, others)
- Extraction (pre-processing)
 - Database queries using SQL and SQL scripts
 - Uses Java for computationally heavyweight filters
- Prepared Data (intermediate representation)
 - Output of pre-processing stage for specific tasks
 - Aggregation speeds up analytical processing
- Evaluation (analytical processing)
 - Mines data for insight, hypothesis testing
 - Basic processing, Java programming, R-project
- Analysis output and results
 - Results of processing: averages, distributions, correlations
 - Presented as models, tables, graphs, charts, etc.

Linux Kernel

- In git with clean timezone data
- Author/committer distinction
- From 2005 to 2011
- One project

Ohloh 2008 Snapshot

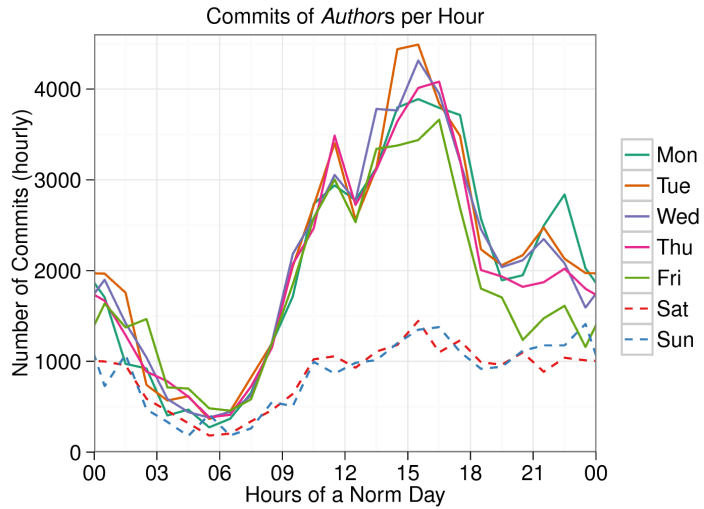
- All kinds of CMS, no TZ data
- Only committer information
- From 2000 to 2008
- 9192 projects in database

Methodological Challenges

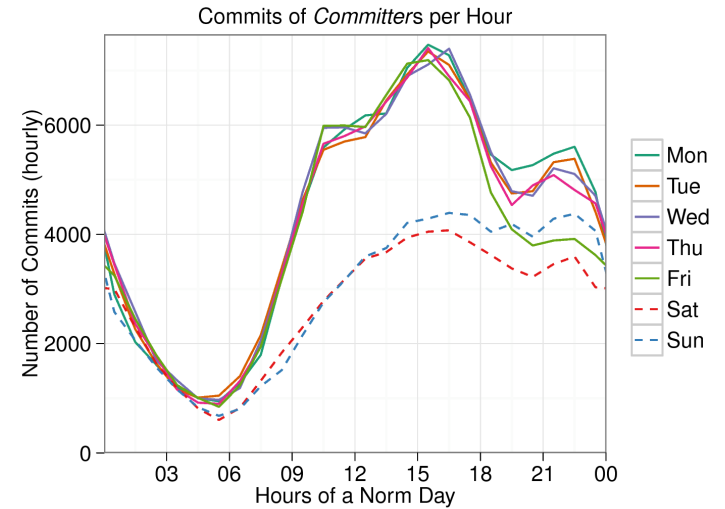
- Time-zone data not available in Ohloh
 - Determined committer timezones for 8% of work using location data
 - Found no reason to question representativeness smaller sample
 - *Worked with two sets, known and extended committers*
- Cultural bias using Western definitions
 - Review of open source activity by time-zone shows Western dominance
 - *Our results are a lower bound estimate*

Overview Result (All Data Sets)

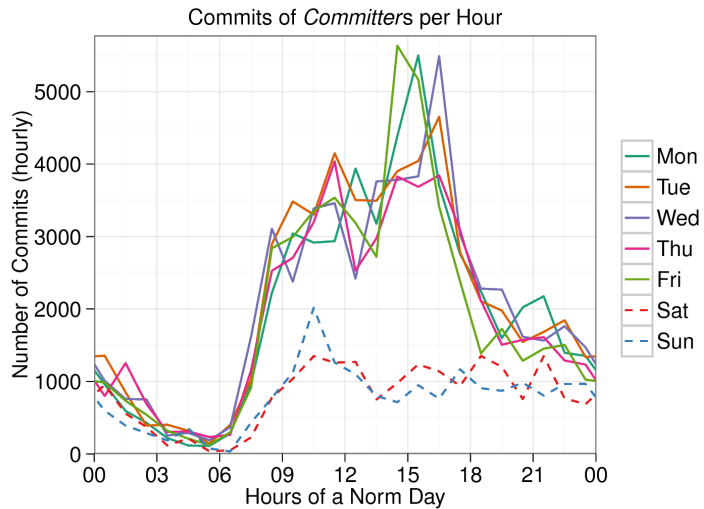
Linux Kernel Authors



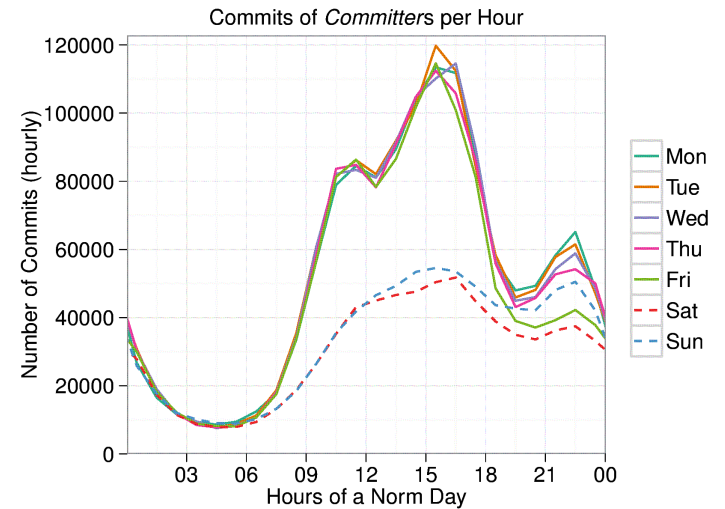
Ohloh Known Set



Linux Kernel Committers

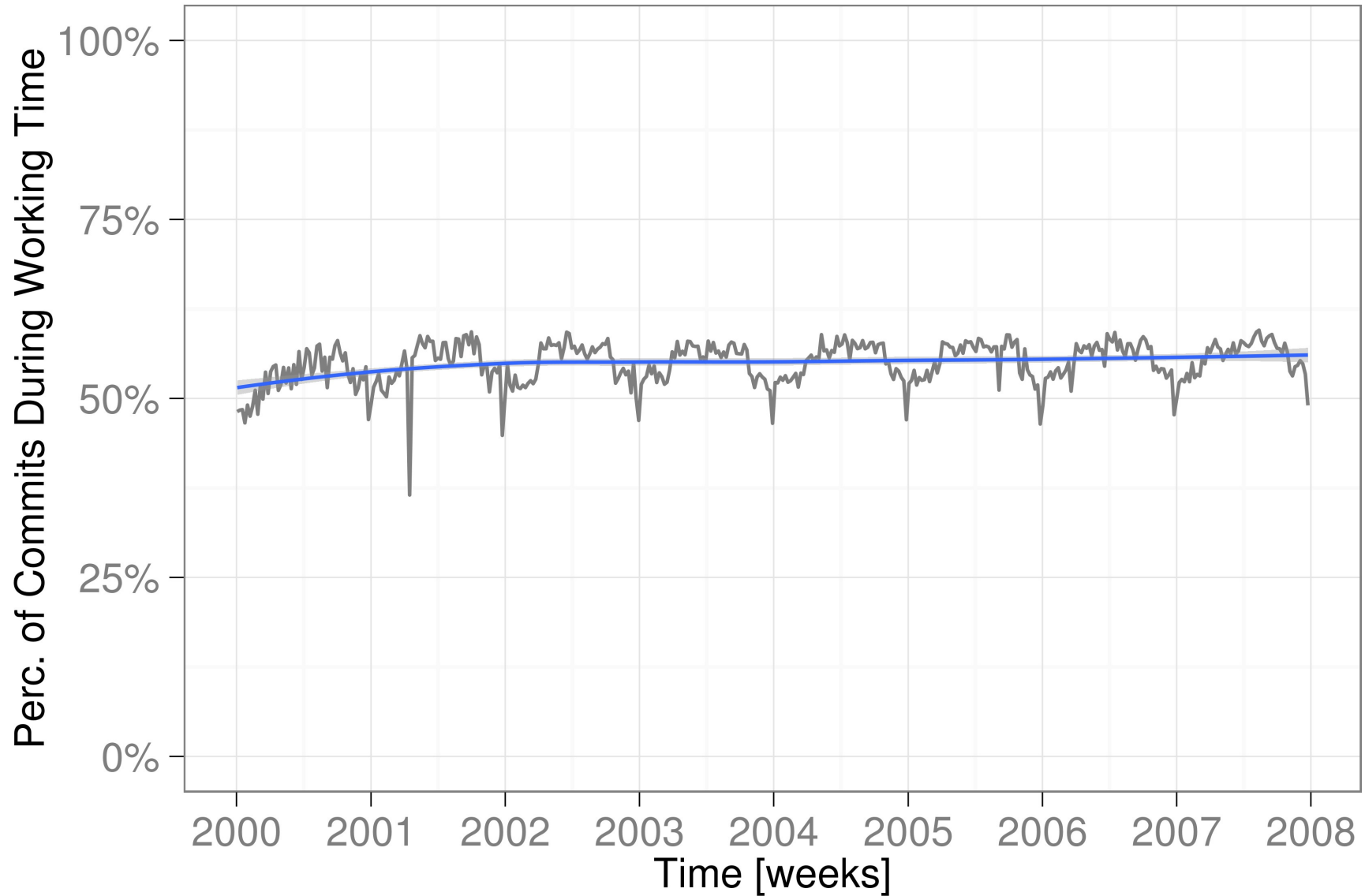


Ohloh Extended Set



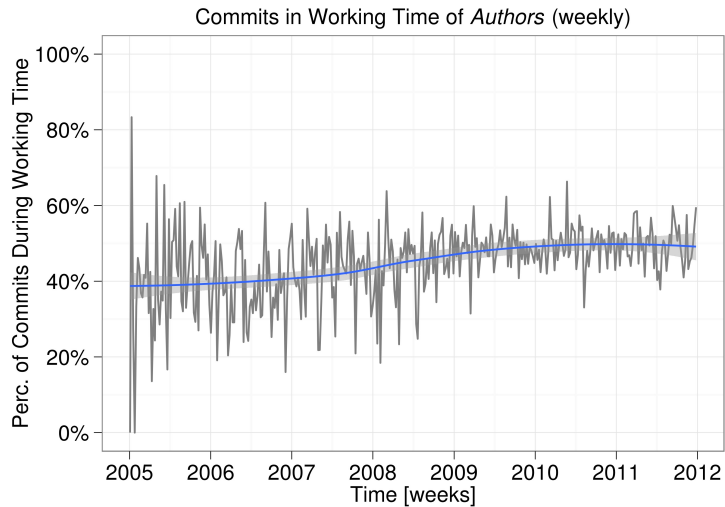
Ohloh Extended Committer Set

Commits in Working Time of *extended Committers* (weekly)

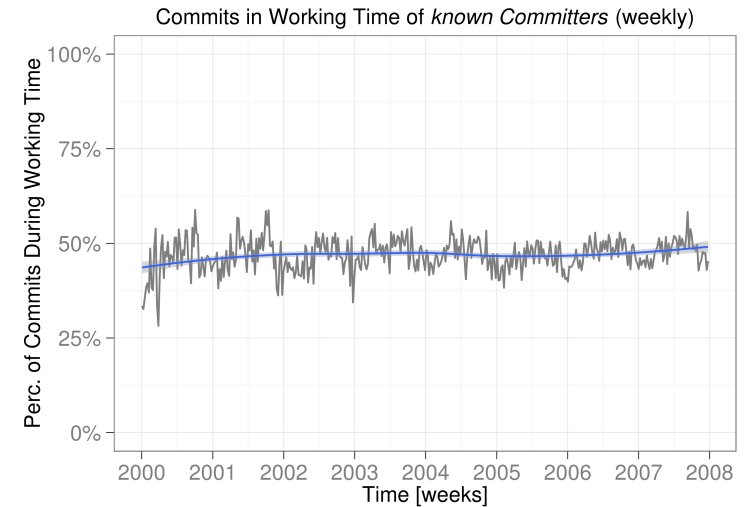


Trend Line (All Data Sets)

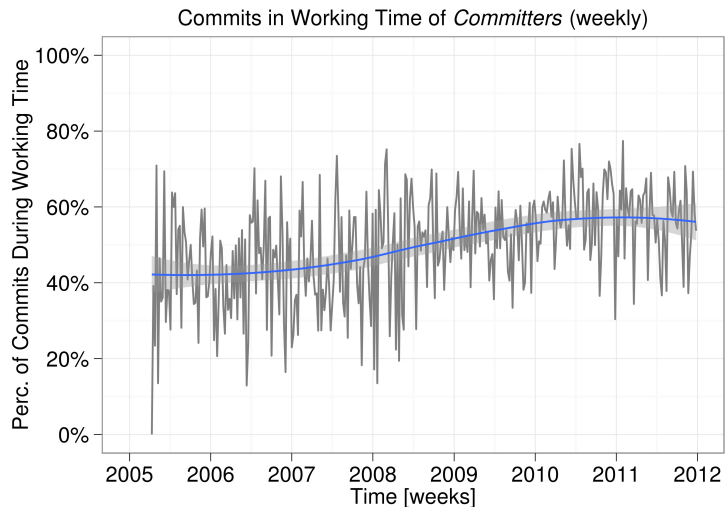
Linux Kernel Authors



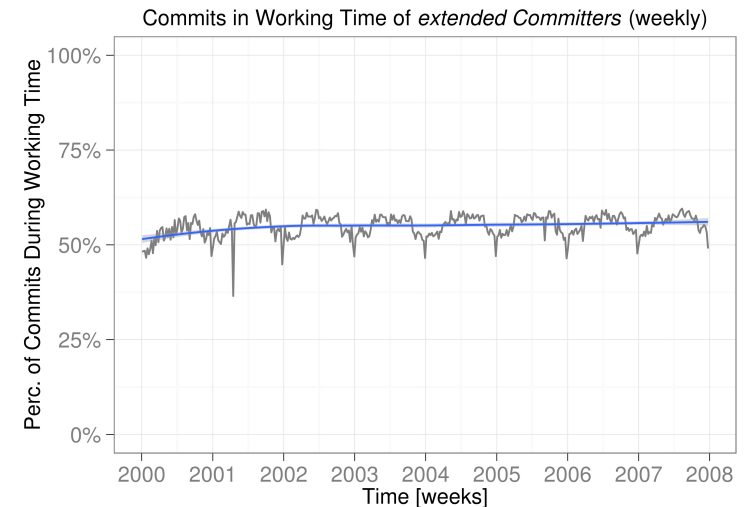
Ohloh Known Set



Linux Kernel Committers



Ohloh Extended Set



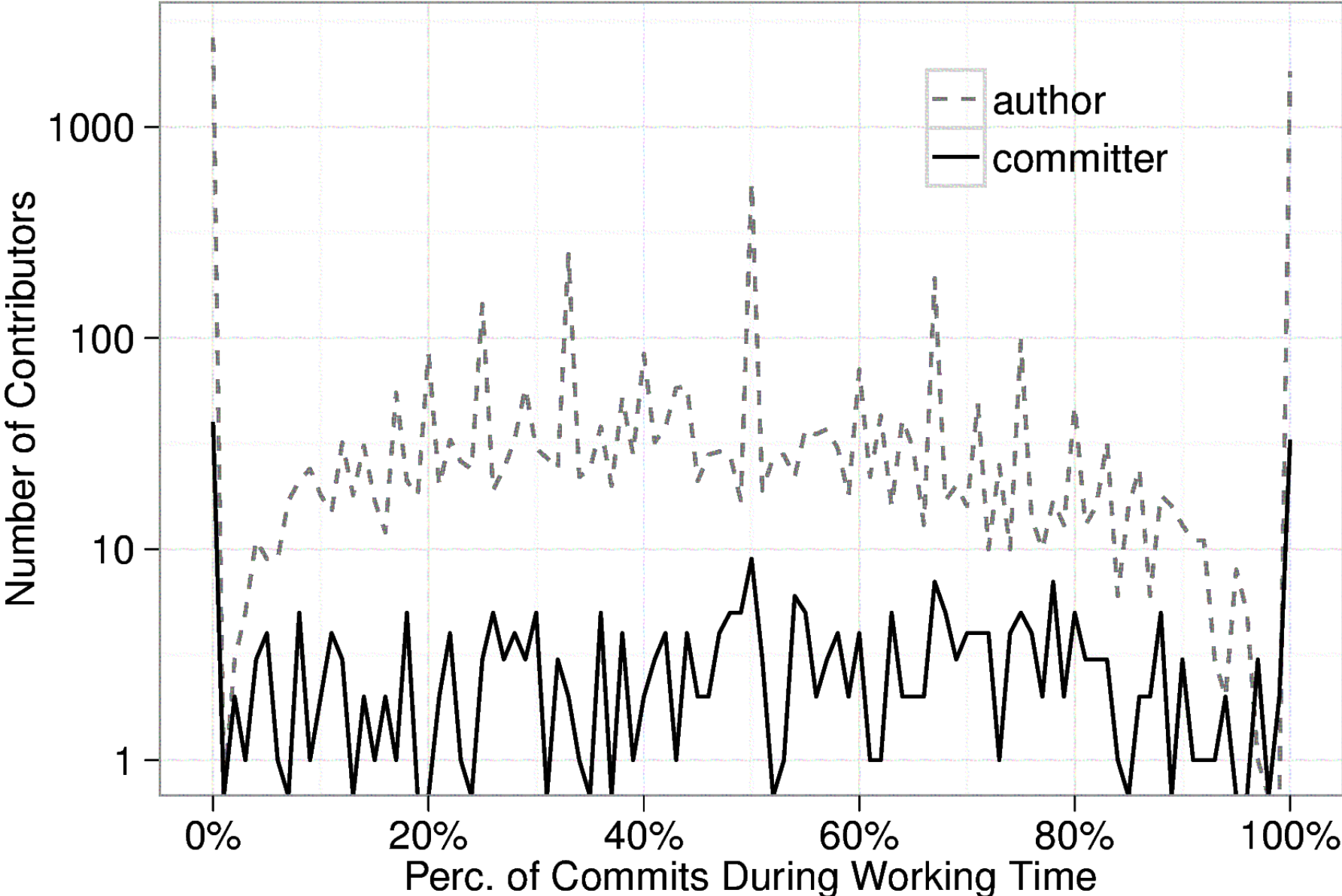
About 50% of all work contributed to open source software projects has been provided Monday to Friday, between 9am and 5pm.

Detail Aggregate Results

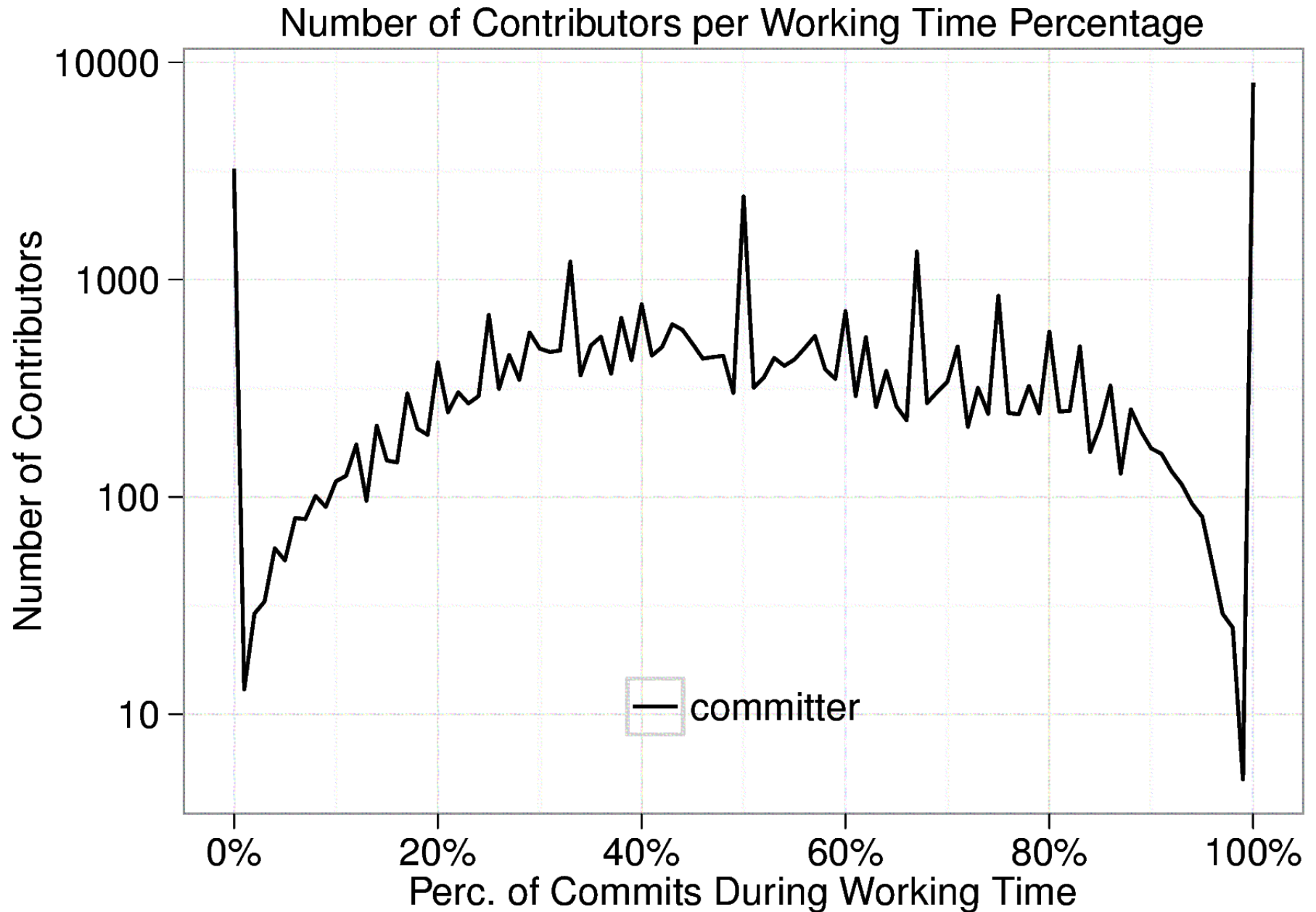
		Percentage of total commits made during working time
Linux Kernel	author	45.00%
	committer	51.36%
Ohloh Projects	known committer	47.3% (min. 28,2%, max. 58,8%)
	extended committers	55.4% (min. 36.5%, max. 59.5%)

Linux Kernel Work Time Percentage

Distribution of Contributors per Perc. During Working Time



Ohloh Work Time Percentage



Work Time Percentage Data

		Volunteer (Spare Time) Work		Mixed	Paid (Working Time) Work	
		0%	0.01%-5%	5.01%-94.99%	95%-99.99%	100%
Linux Kernel	author	33.06%	0.35%	43.45%	0.17%	22.98%
	committer	11.59%	3.05%	74.09%	1.52%	9.76%
Ohloh Projects	known committers	2.41%	1.21%	95.69%	0.00%	0.69%
	extended committers	7.04%	0.4%	74.58%	0.41%	17.56%

- Note the inverse relationship Linux Kernel author / committer

Future Work

- Dimensional break-outs
 - By type of software (infrastructure, productivity, entertainment)
 - By time-zone and (implied) culture (what about China?)
- More current data sets

Thank you! Questions?

dirk.riehle@fau.de – <http://osr.cs.fau.de>

dirk@riehle.org – <http://dirkriehle.com> – [@dirkriehle](#)