Visualizing Work Process in Software Engineering with Developer Rivers

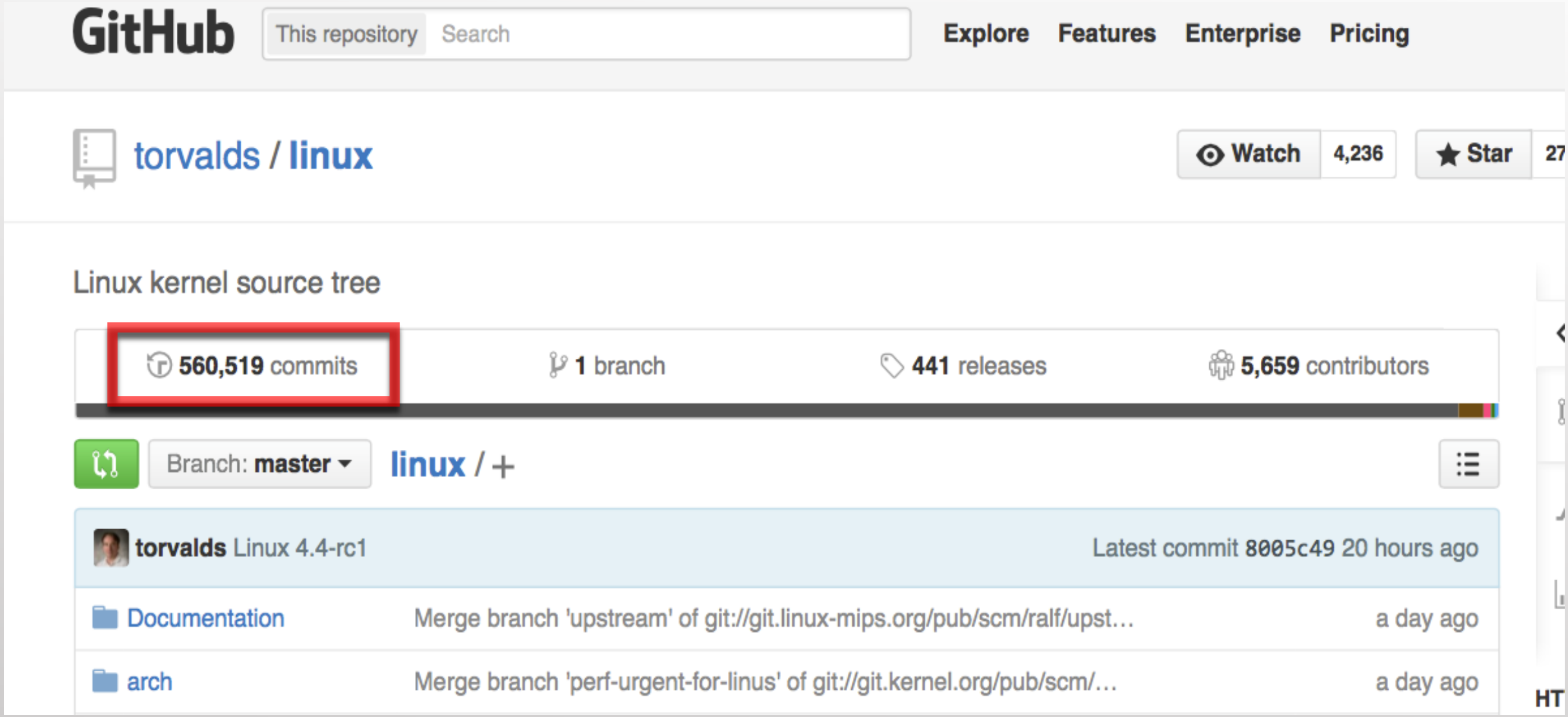
Michael Burch, Tanja Munz, Fabian Beck, and Daniel Weiskopf
VISUS, University of Stuttgart, Germany

Presenter: Arthur Sun

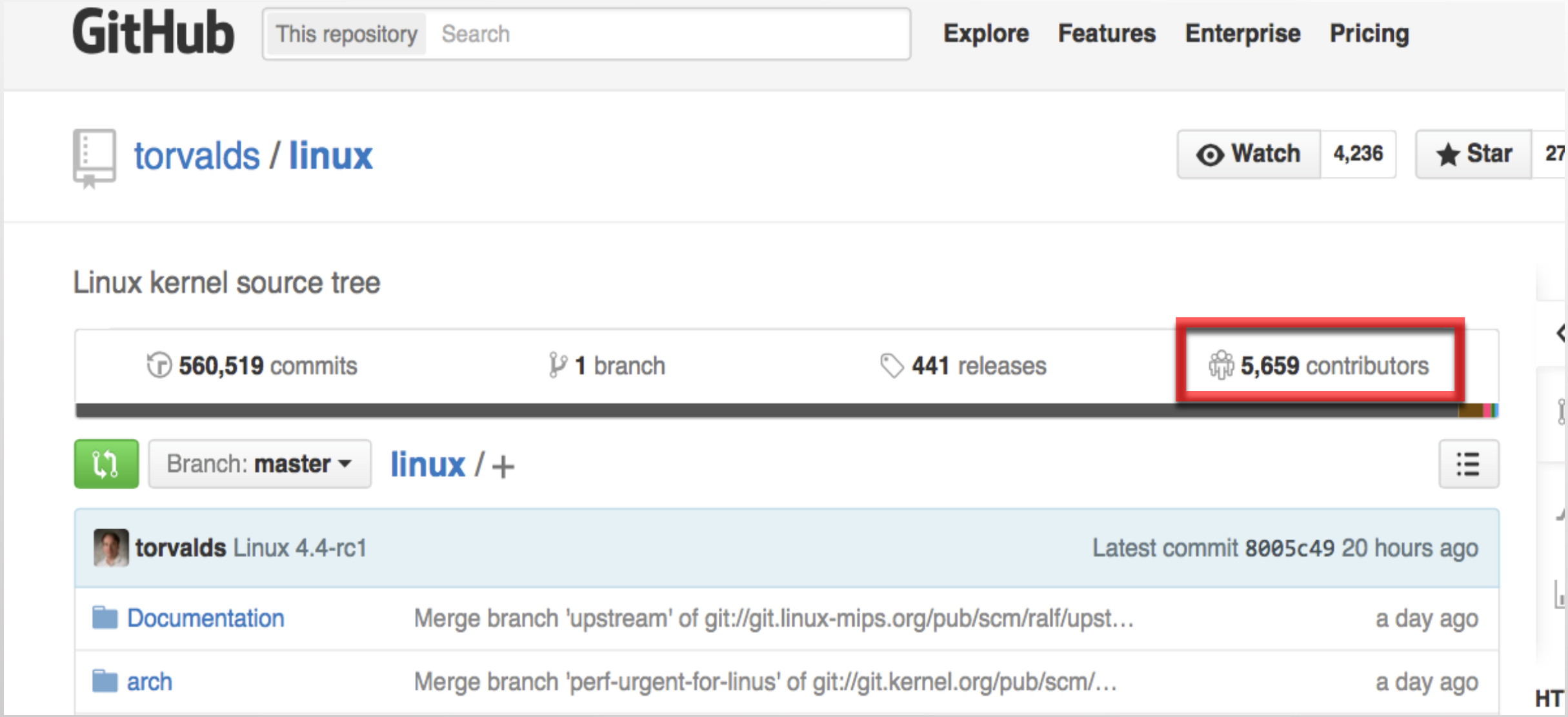
Outline

- What's the current problem for large software projects
- What's the previous solution for large projects InfoVis
- What the paper presents
  - DataSet
  - InfoVis Encoding Technique
  - Visualization Method
  - Sample Usage
- Future Improvement

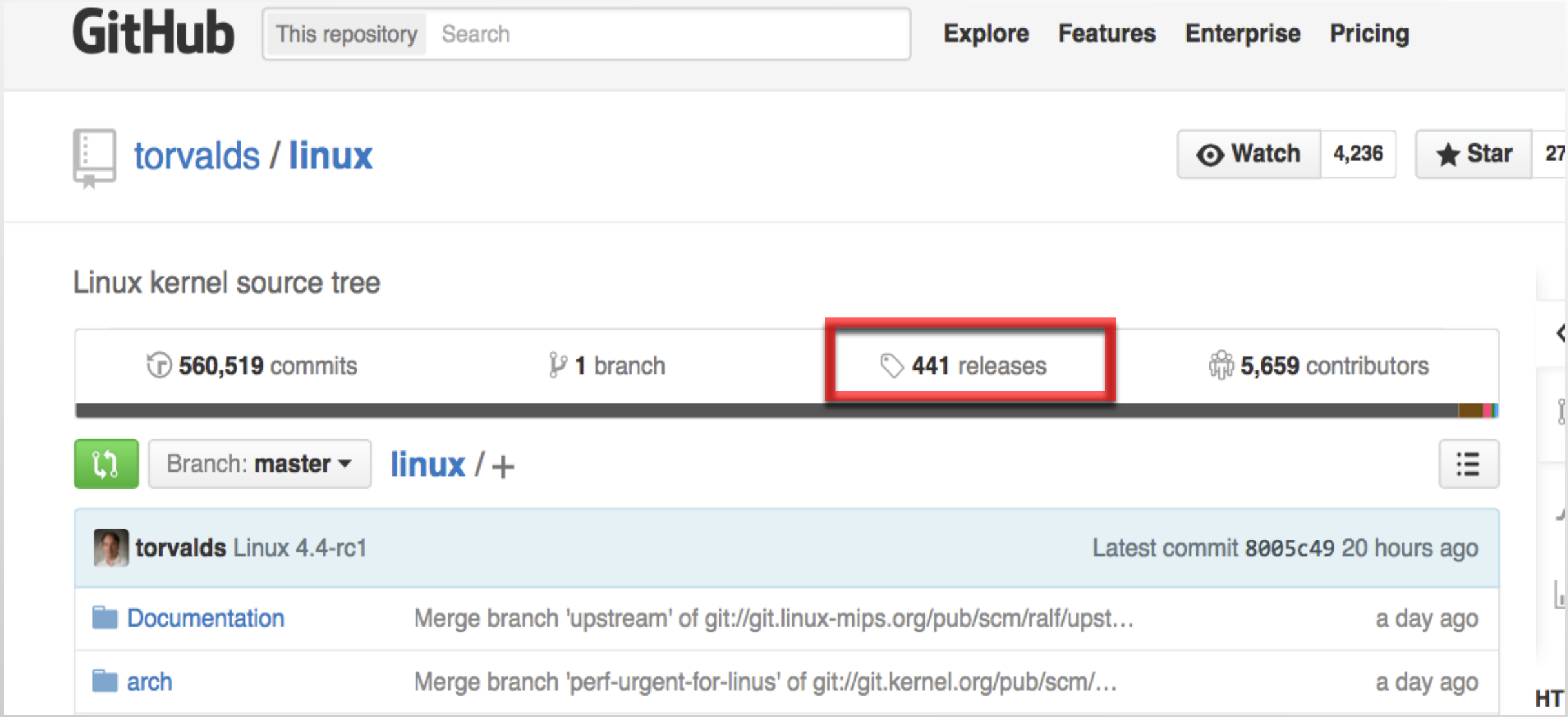
Large open-source projects: 560,519 commits



Large open-source projects: 5659 contributors



Large open-source projects: 441 releases



apache / storm

mirrored from git://git.apache.org/incubator-storm.git

Watch 482 | Star 2,541

Mirror of Apache Storm

5,320 commits | 59 branches | 18 releases | 187 contributors

Branch: master ▾ storm / +

revans2 Added STORM-1190 to Changelog — Latest commit 5a79ba5 3 days ago

bin	Merge branch 'STORM-1155' of https://github.com/tgravescs/storm into …	5 days ago
conf	Merge branch 'STORM-1155' of https://github.com/tgravescs/storm into …	5 days ago
dev-tools	bump timeout by 50% due to intermittent travis build failures	5 days ago
docs	Merge branch 'STORM-1155' of https://github.com/tgravescs/storm into …	5 days ago
examples/storm-starter	STORM-1161: Add License headers and add rat checks to builds	11 days ago
external	Merge branch 'patch-6' of https://github.com/vesense/storm into STORM…	10 days ago

Cod

Pull

Puls

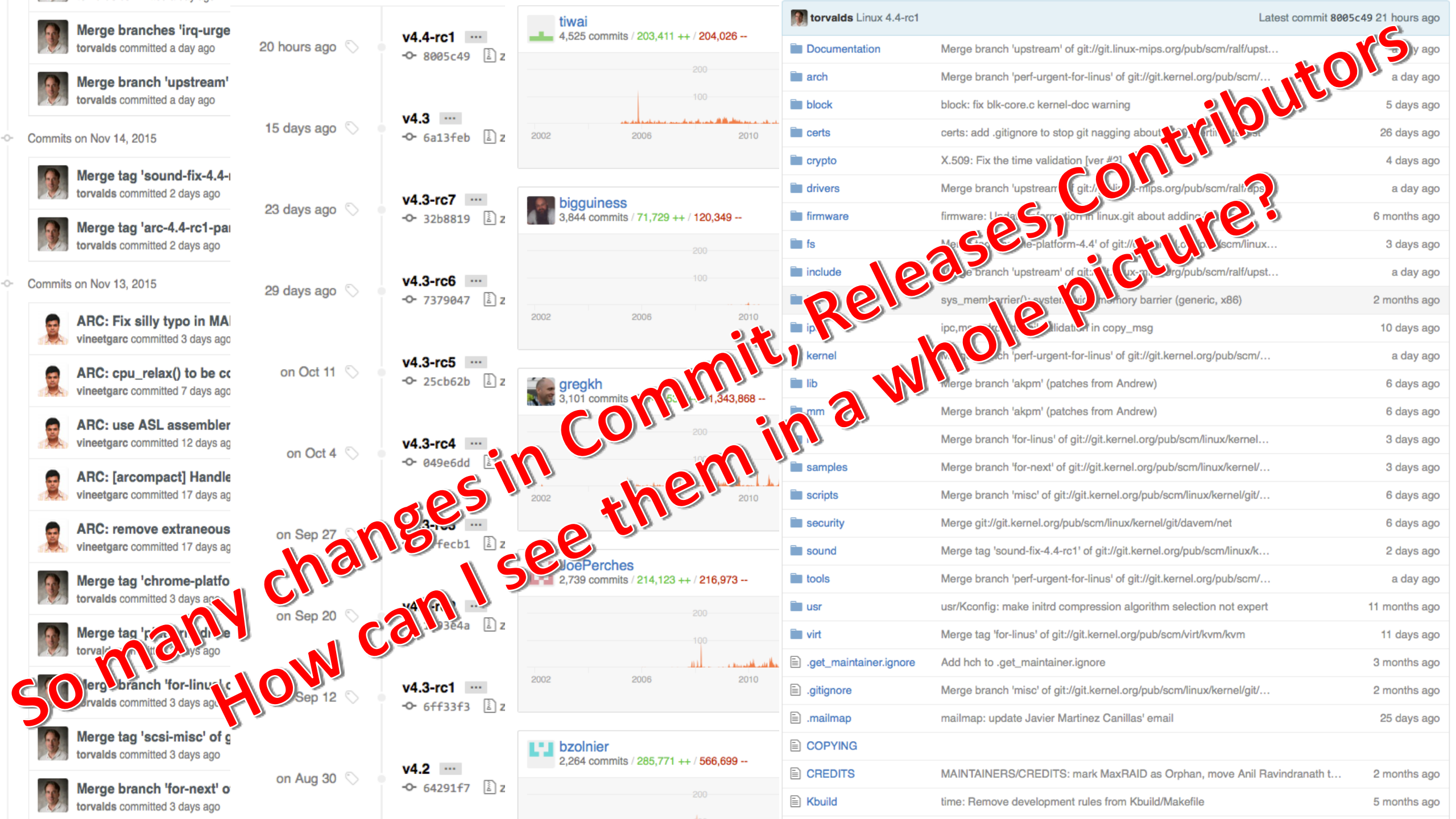
Gra

HTTPS cl

https:/

You can cl
Subversio

Cl


Merge branches 'irq-urge
torvalds committed a day ago
Merge branch 'upstream'
torvalds committed a day ago
Commits on Nov 14, 2015
Merge tag 'sound-fix-4.4-
torvalds committed 2 days ago
Merge tag 'arc-4.4-rc1-pa
torvalds committed 2 days ago
Commits on Nov 13, 2015
ARC: Fix silly typo in MA
vineetgarc committed 3 days ago
ARC: cpu_relax() to be cc
vineetgarc committed 7 days ago
ARC: use ASL assembler
vineetgarc committed 12 days ag
ARC: [arcompact] Handle
vineetgarc committed 17 days ag
ARC: remove extraneous
vineetgarc committed 17 days ag
Merge tag 'chrome-platfo
torvalds committed 3 days ag
Merge tag 'r
torval
Merge branch 'for-linu
orvalds committed 3 days ag
Merge tag 'scsi-misc' of g
torvalds committed 3 days ago
Merge branch 'for-next' o
torvalds committed 3 days ago
20 hours ago
15 days ago
23 days ago
29 days ago
on Oct 11
on Oct 4
on Sep 27
on Sep 20
Sep 12
on Aug 30
v4.4-rc1
8005c49
v4.3
6a13feb
v4.3-rc7
32b8819
v4.3-rc6
7379047
v4.3-rc5
25cb62b
v4.3-rc4
049e6dd
fecb1
9364a
v4.3-rc1
6ff33f3
v4.2
64291f7
tiwai
4,525 commits / 203,411 ++ / 204,026 --
bigguiness
3,844 commits / 71,729 ++ / 120,349 --
gregkh
3,101 commits
1,343,868 --
JoePerches
2,739 commits / 214,123 ++ / 216,973 --
bzolnier
2,264 commits / 285,771 ++ / 566,699 --
200
100
2002
2006
2010
torvalds Linux 4.4-rc1
Latest commit 8005c49 21 hours ago
Documentation
Merge branch 'upstream' of git://git.linux-mips.org/pub/scm/ralf/upst…
arch
Merge branch 'perf-urgent-for-linus' of git://git.kernel.org/pub/scm/…
a day ago
block
block: fix blk-core.c kernel-doc warning
5 days ago
certs
26 days ago
crypto
X.509: Fix the time validation
4 days ago
drivers
a day ago
firmware
6 months ago
fs
3 days ago
include
a day ago
2 months ago
10 days ago
kernel
a day ago
lib
Merge branch 'akpm' (patches from Andrew)
6 days ago
mm
Merge branch 'akpm' (patches from Andrew)
6 days ago
3 days ago
samples
Merge branch 'for-next' of git://git.kernel.org/pub/scm/linux/kernel/…
3 days ago
scripts
Merge branch 'misc' of git://git.kernel.org/pub/scm/linux/kernel/git/…
6 days ago
security
Merge git://git.kernel.org/pub/scm/linux/kernel/git/davem/net
6 days ago
sound
Merge tag 'sound-fix-4.4-rc1' of git://git.kernel.org/pub/scm/linux/k…
2 days ago
tools
a day ago
usr
usr/Kconfig: make initrd compression algorithm selection not expert
11 months ago
virt
Merge tag 'for-linus' of git://git.kernel.org/pub/scm/virt/kvm/kvm
11 days ago
.get_maintainer.ignore
Add hch to .get_maintainer.ignore
3 months ago
.gitignore
Merge branch 'misc' of git://git.kernel.org/pub/scm/linux/kernel/git/…
2 months ago
.mailmap
mailmap: update Javier Martinez Canillas' email
25 days ago
COPYING
CREDITS
MAINTAINERS/CREDITS: mark MaxRAID as Orphan, move Anil Ravindranath t…
2 months ago
Kbuild
time: Remove development rules from Kbuild/Makefile
5 months ago
So many changes in Commit, Releases, Contributors
How can I see them in a whole picture?

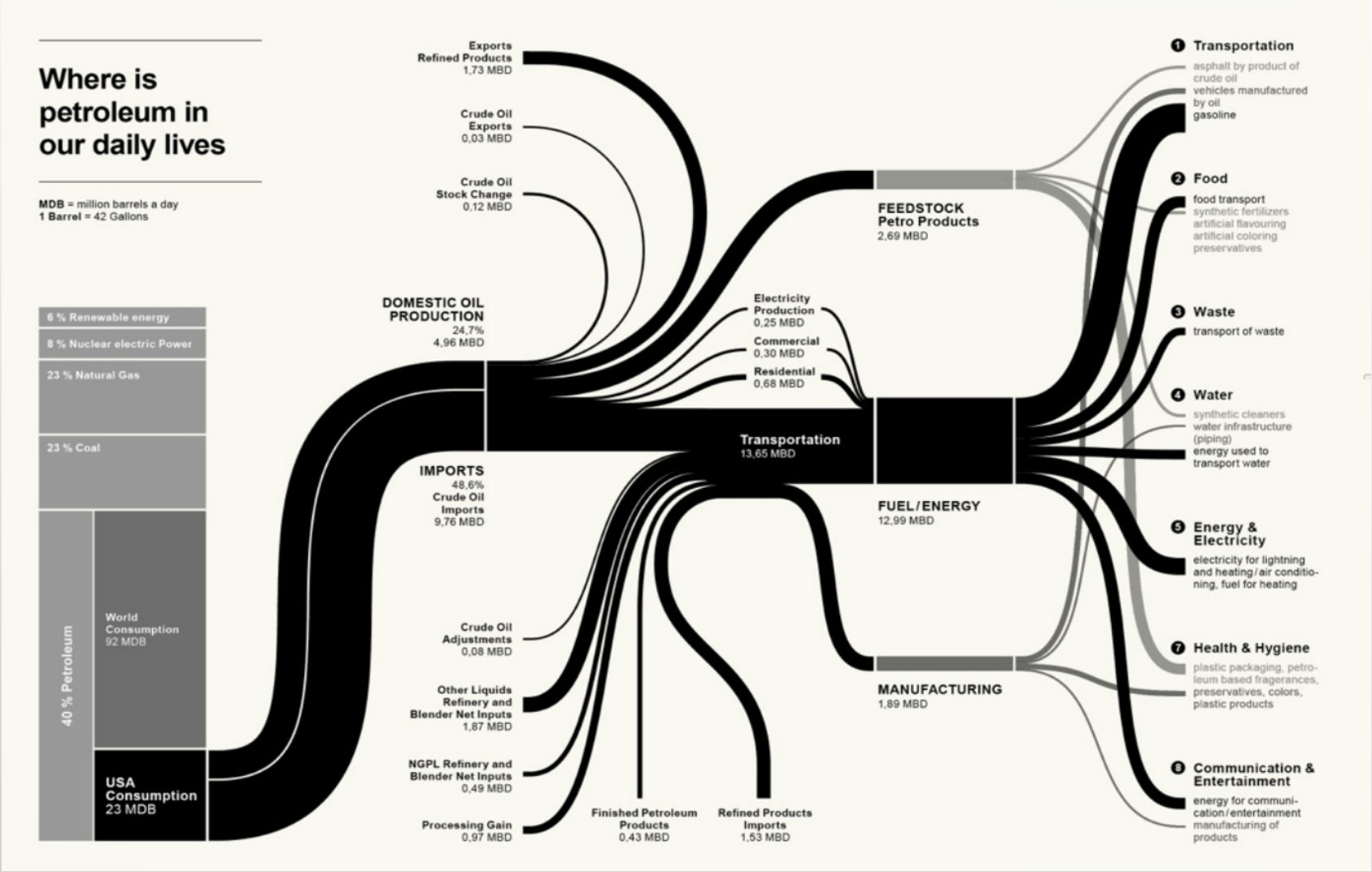
What’s do we want

A whole picture of the overall progress of extreme large software engineering project proceeding with time frame in detailed visualization for major participants, their contribution to respective work, how much amount of work they did and their work change

Previous work – Sankey Diagram

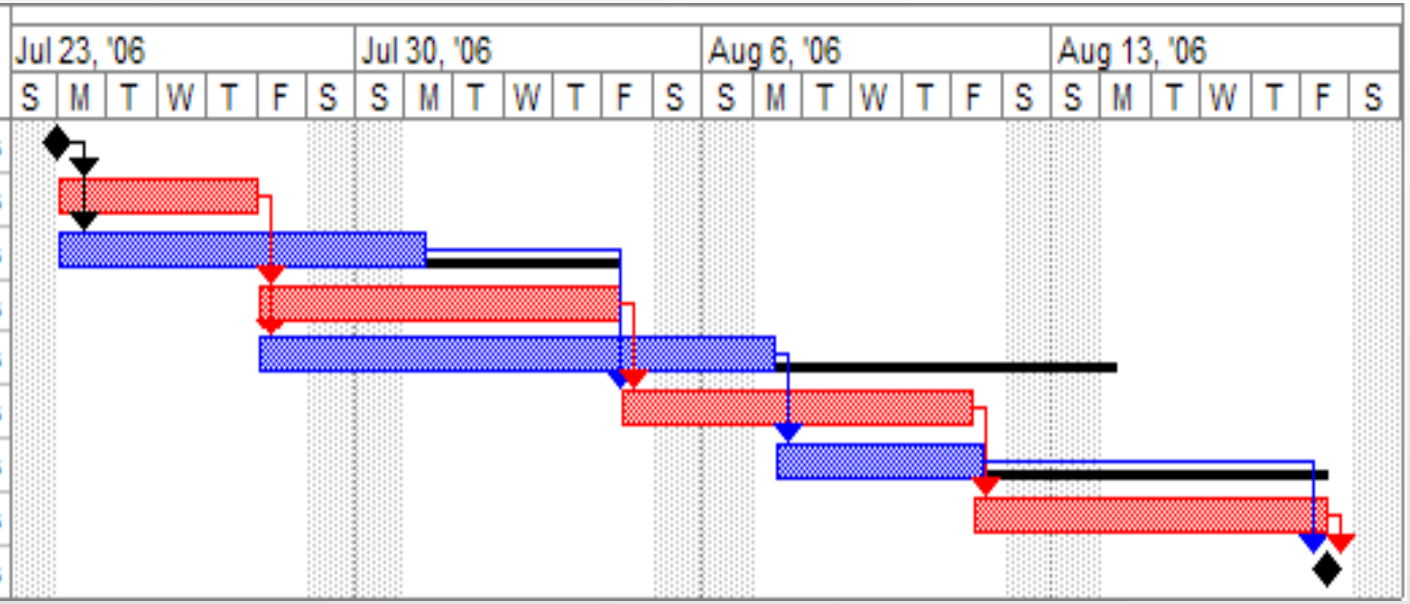
Sankey diagrams are a specific type of flow diagram, in which the width of the arrows is shown proportionally to the flow quantity.

Problem:
No Time Frame



Previous work – Gantt Chart

ID	Task Name	Predecessors	Duration
1	Start		0 days
2	a	1	4 days
3	b	1	5.33 days
4	c	2	5.17 days
5	d	2	6.33 days
6	e	3,4	5.17 days
7	f	5	4.5 days
8	g	6	5.17 days
9	Finish	7,8	0 days



A Gantt chart is a type of bar chart that illustrates a project schedule. Gantt charts illustrate the start and finish dates of the involved tasks. Modern Gantt charts also show the dependency

Problem: 1. Doesn't show how many people/resources involved in project

2. Don't have a whole picture about the project

What’s the author propose

- A graph flow which can not only show the interconnection of different modules of development along with the timeframe but also the programmer who took part in the whole project with vivid color to show difference

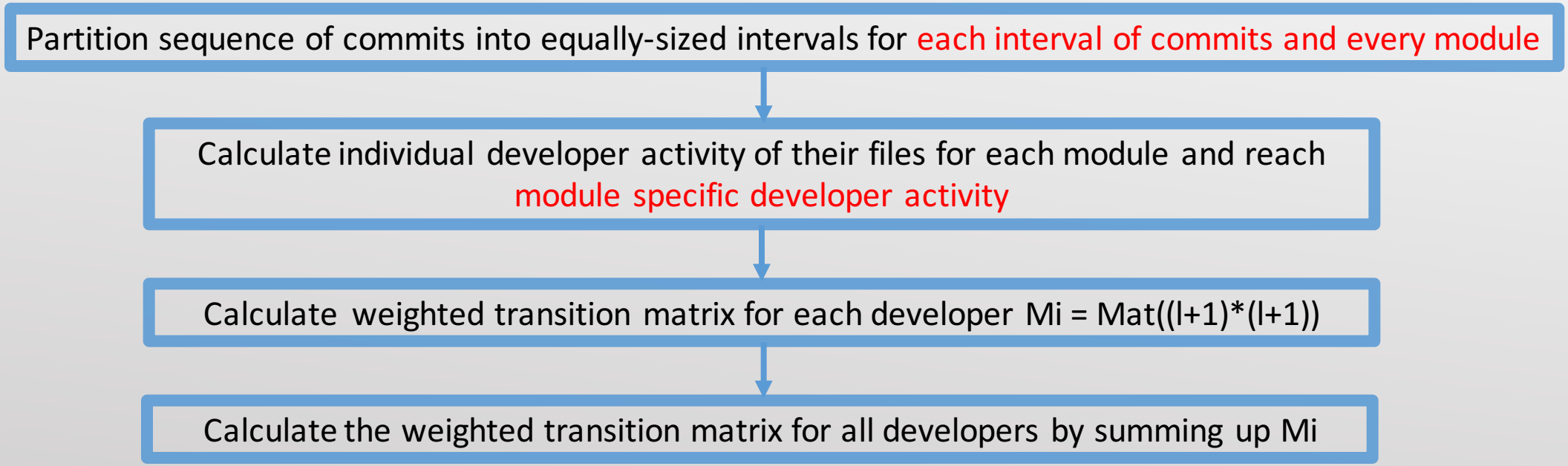
Visualization Technique

DataSet: Developer Activity Model

Encoding: Develop River for Time-Varying Developer Activities

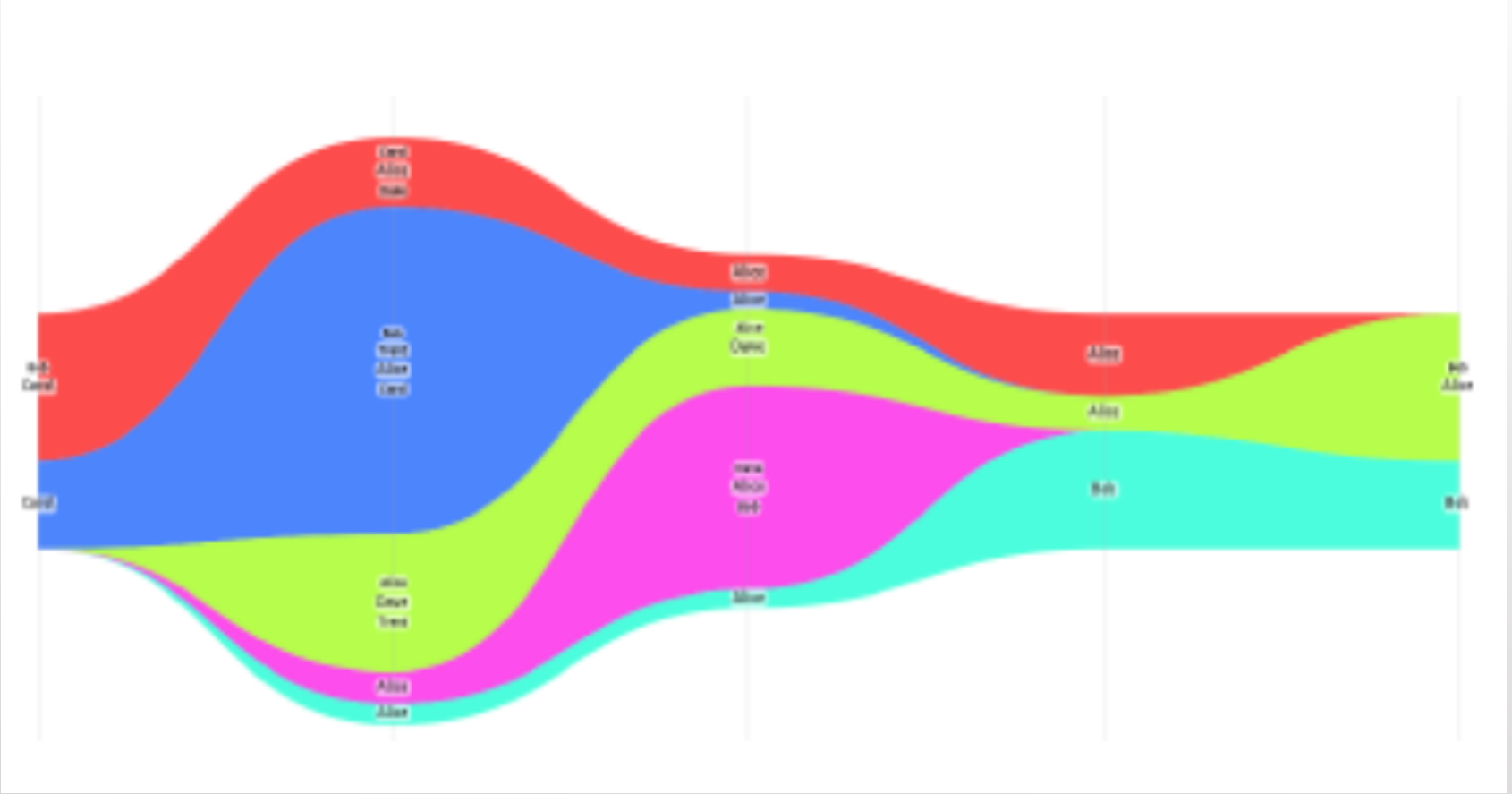
Dataset: Developer Activity Model

1. Abstract commit as c, time as t, developer name as d, files as f, all files as F, file modules hierarchy as H



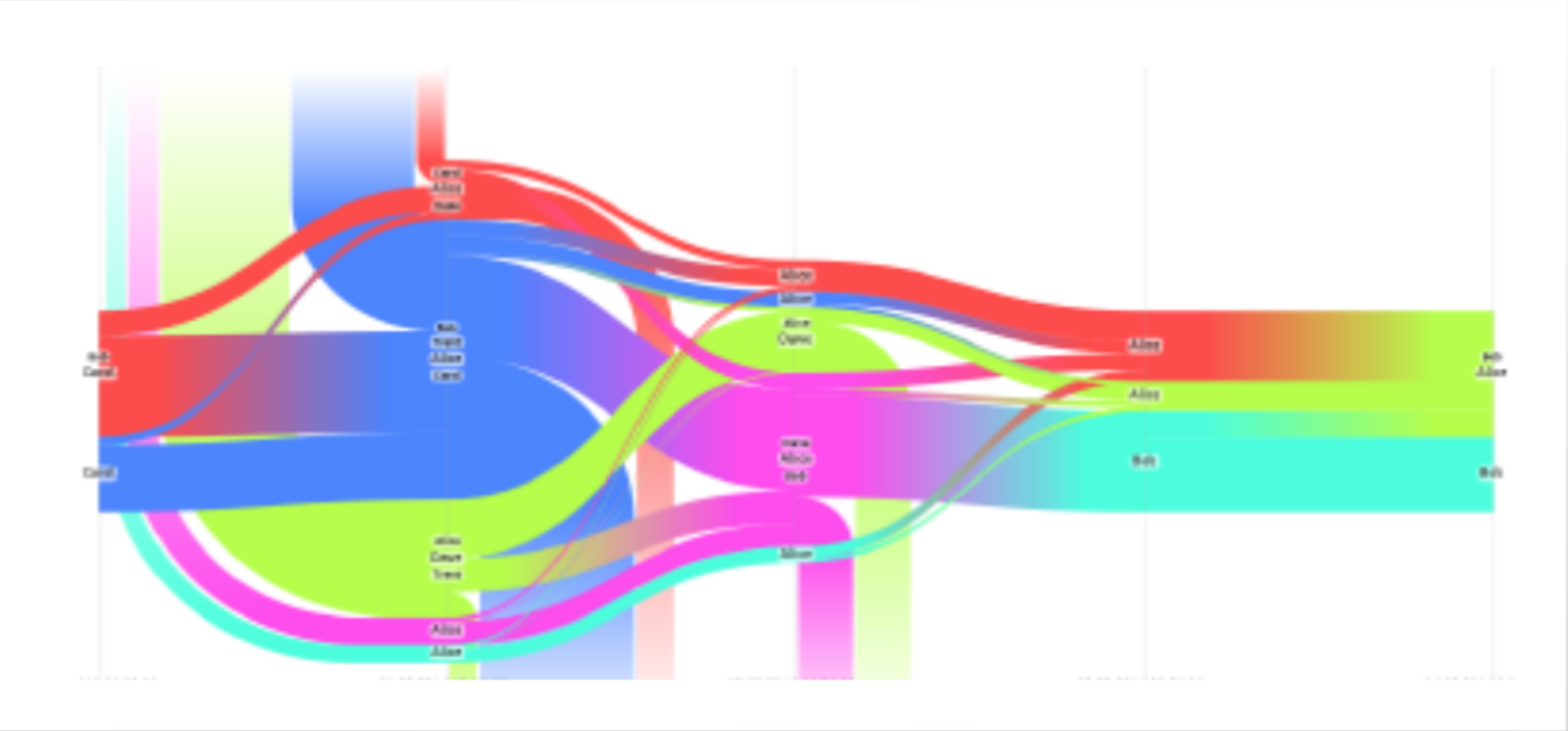
Paper didn’t show how to map real data into Activity Model Matrix

Developer Rivers



Mapping Activity Model Matrix into Develop Rivers without intersection

Developer Rivers

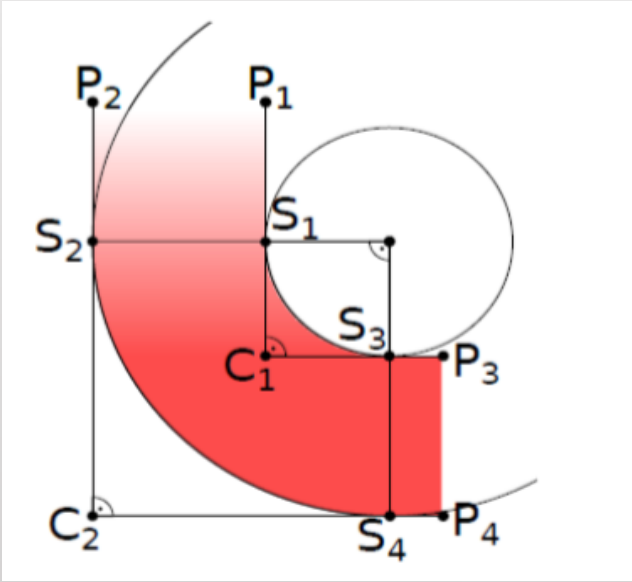


Mapping Activity Model Matrix into Develop Rivers with intersection

Developer Rivers Curves

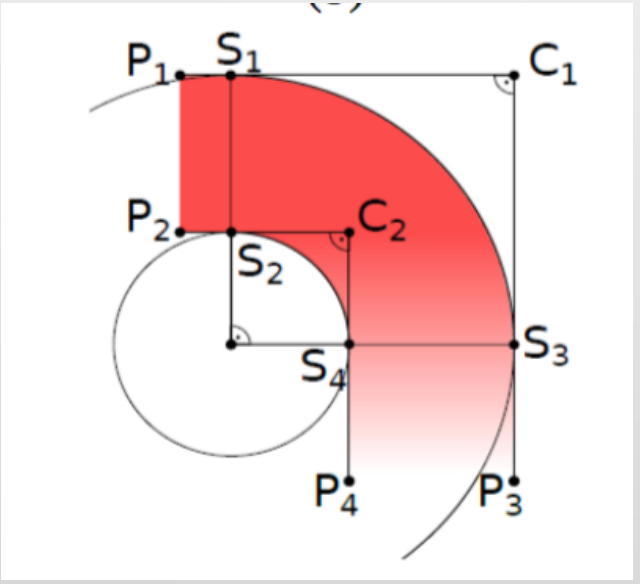
1. Transition: how developers change their behavior between different module groups using cubic Bezier Curves
2. Transition color is a linear gradient from color of start module to target module

Paper didn't show how to link Matrix Data with Bezier curve creation



Influents

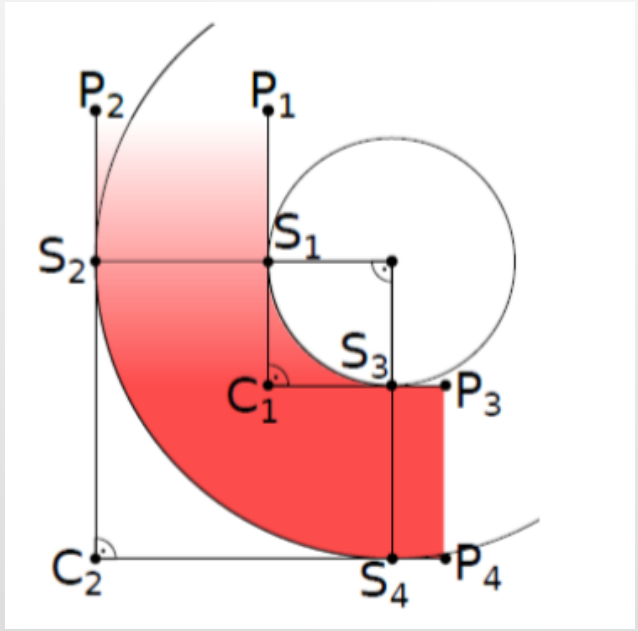
Developer join current step



Effluents

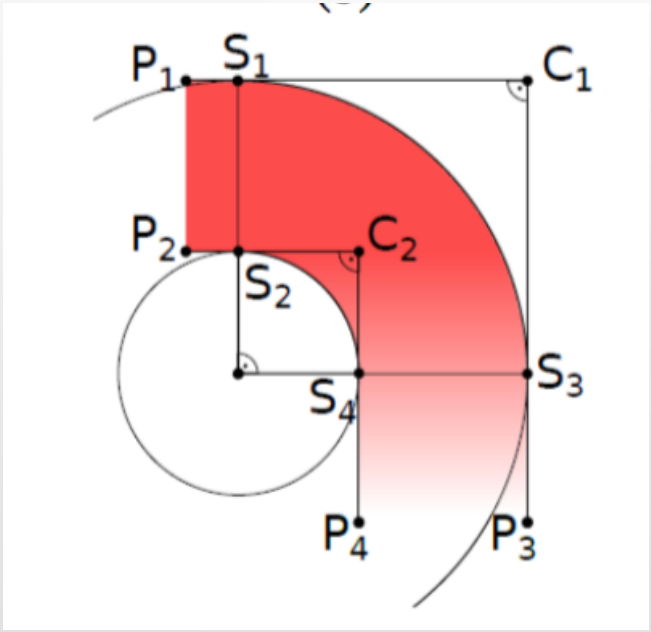
Level the main river

Developer Rivers Curves



Effluents

Leave the main river



Effluents

Leave the main river

Diagram

- Inflow/Outflow: A transition from or to the outside of the diagram identifies a developer enter or leaving the project
- Constant Flow: An intra-transition with a constant width indicating a group of developers constantly working on the same module
- Growth/Decline: An intra-transition with an increasing or decreasing strength hints at a group of developers that keep working on a module but with changing total effort
- Split/Merge: A module that is split into or merged from multiple flows shows a qualitative change of developer activity (i.e., developers' relative focus switches between modules). While at least one inter-transition is required for this pattern, one of the flows can be an intra-transition.
- Exchange: A pair of intra-transitions connecting two modules in opposite directions at the same time is a specific qualitative change of activity: some developers move between the two modules in both directions.

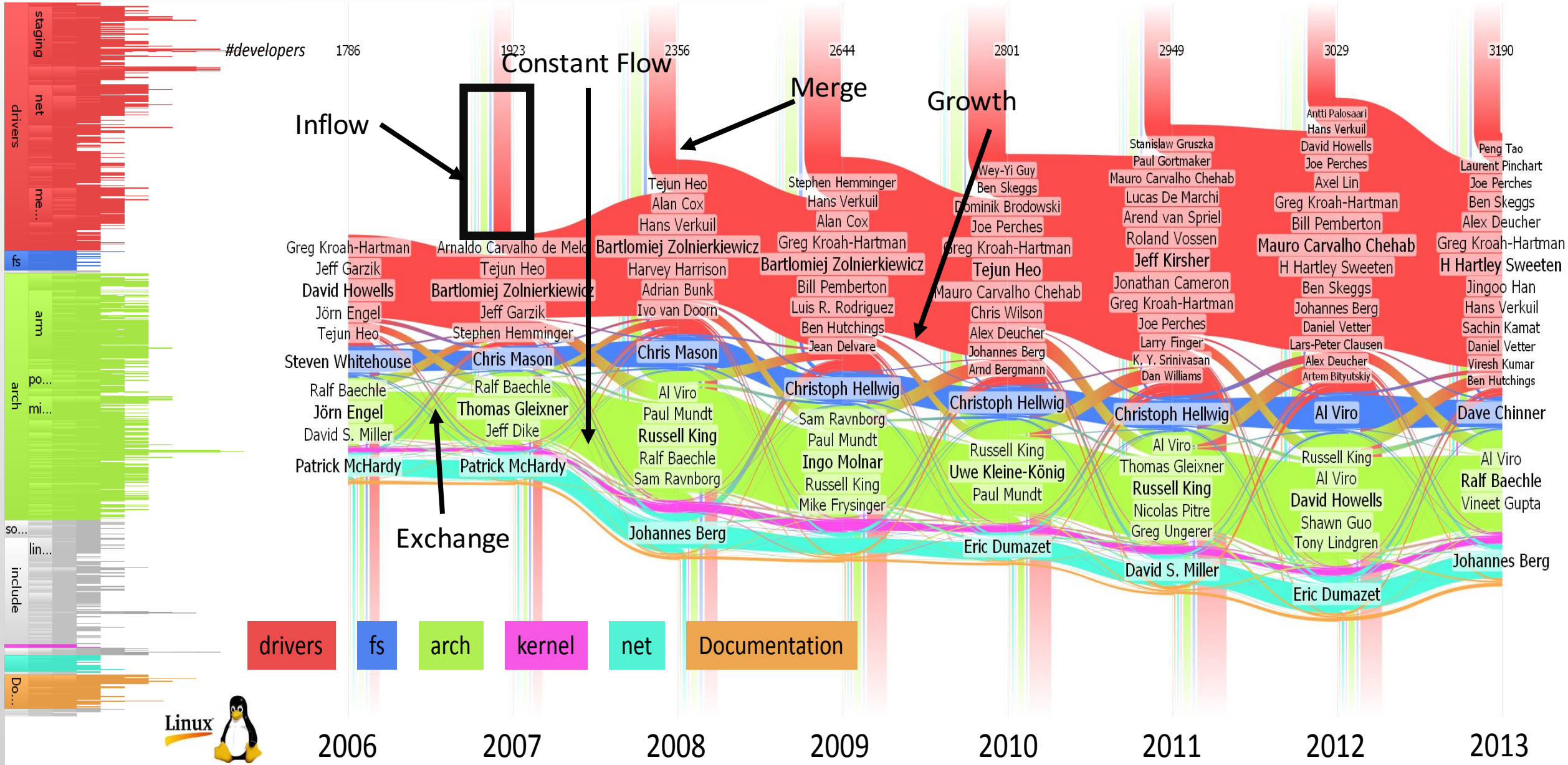
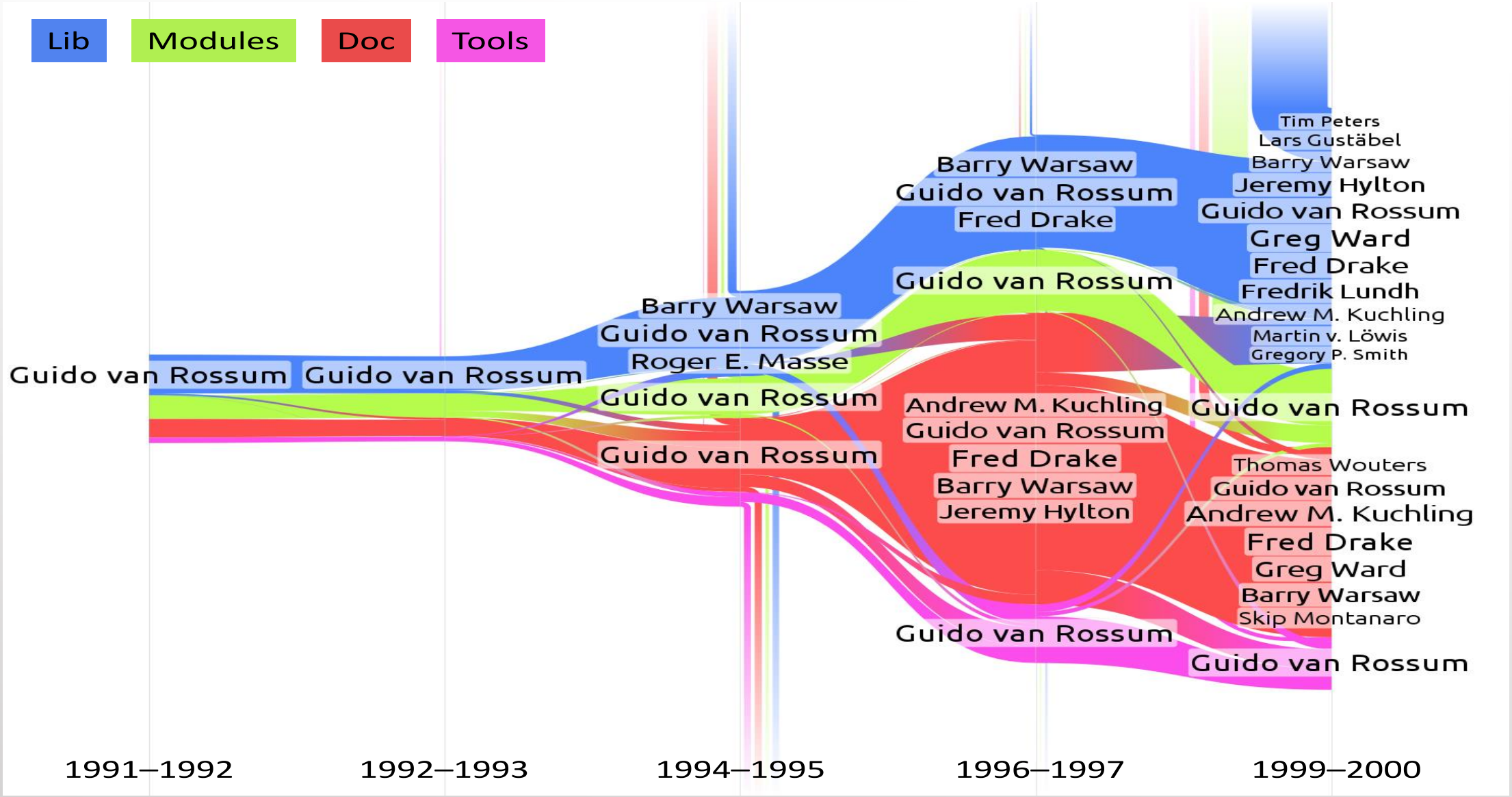


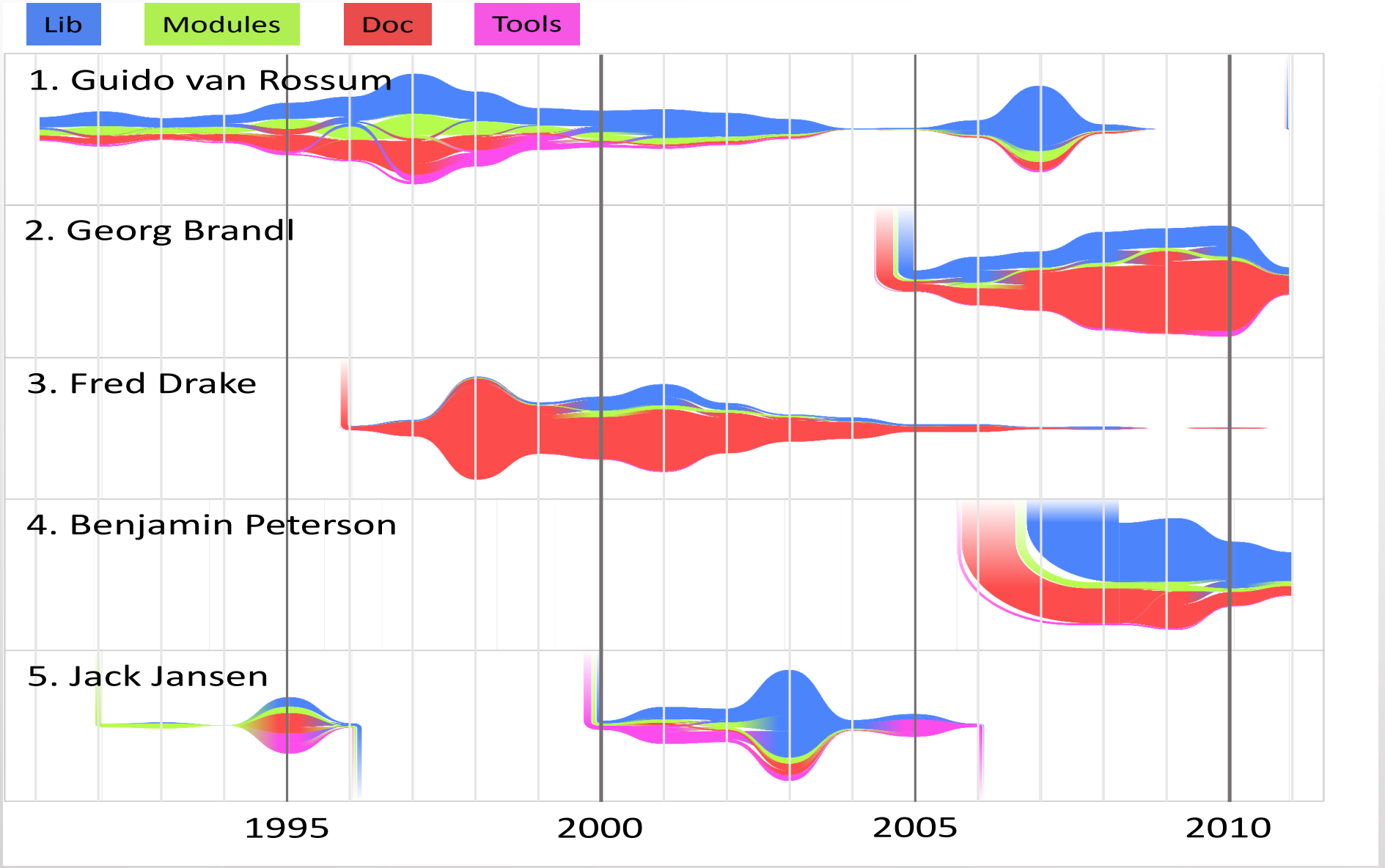
Fig. 8. Linux *main module overview*, 2006–2013, 1-year interval.

Visual Patterns

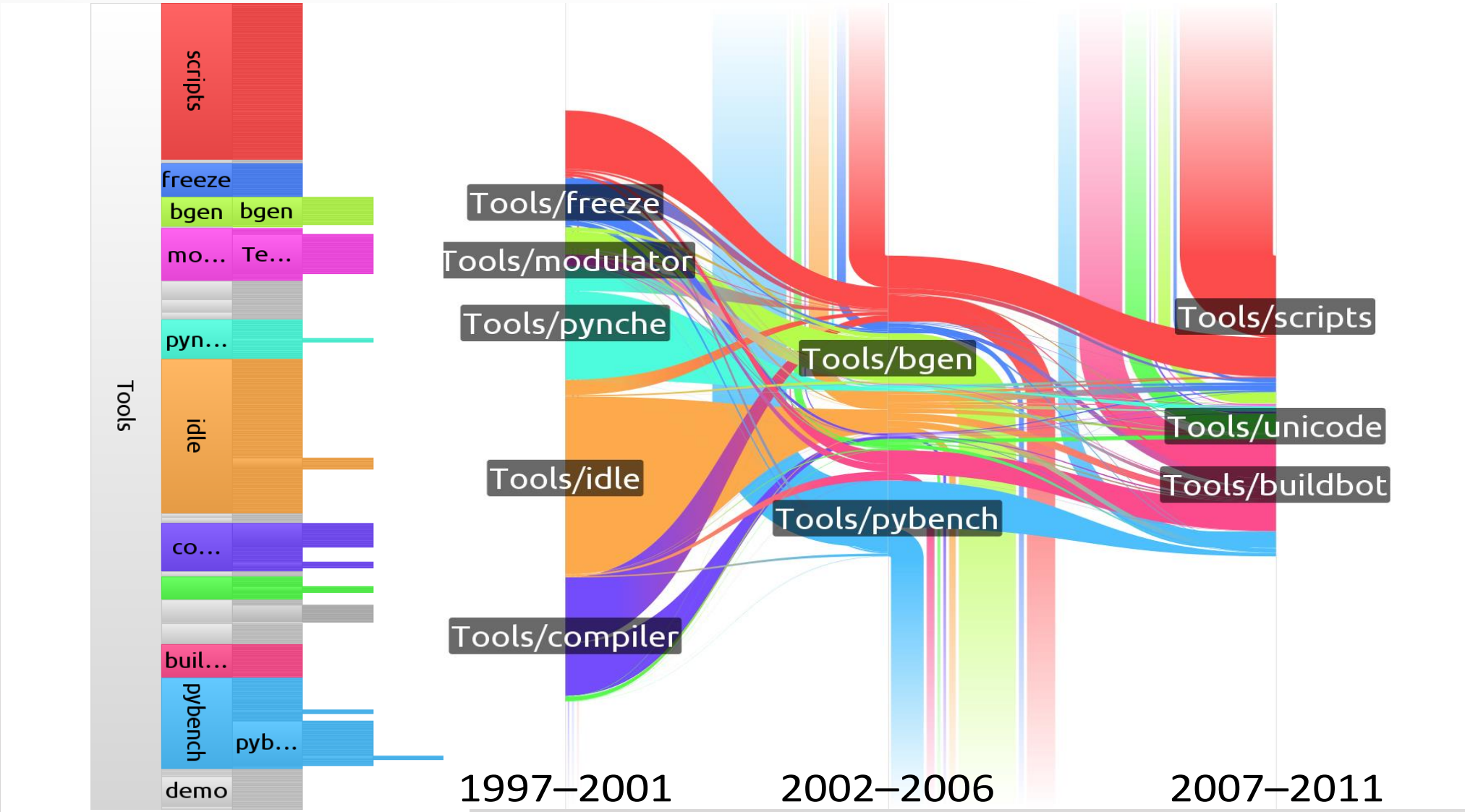
- Main Module Overview
  - Consists main directories, developers and their contributors
- File Type Overview
  - Automatic definition of modules by file types
- Developer Sparklines
  - Highlight top 5 star developer contributing most to the whole project
- Subsystem Details
  - Modules in a subdirectory of the system shows details of a specific system



Python main module overview



Python Developer Sparkline of top 5 developers



Python Subsystem details of Tools Directory

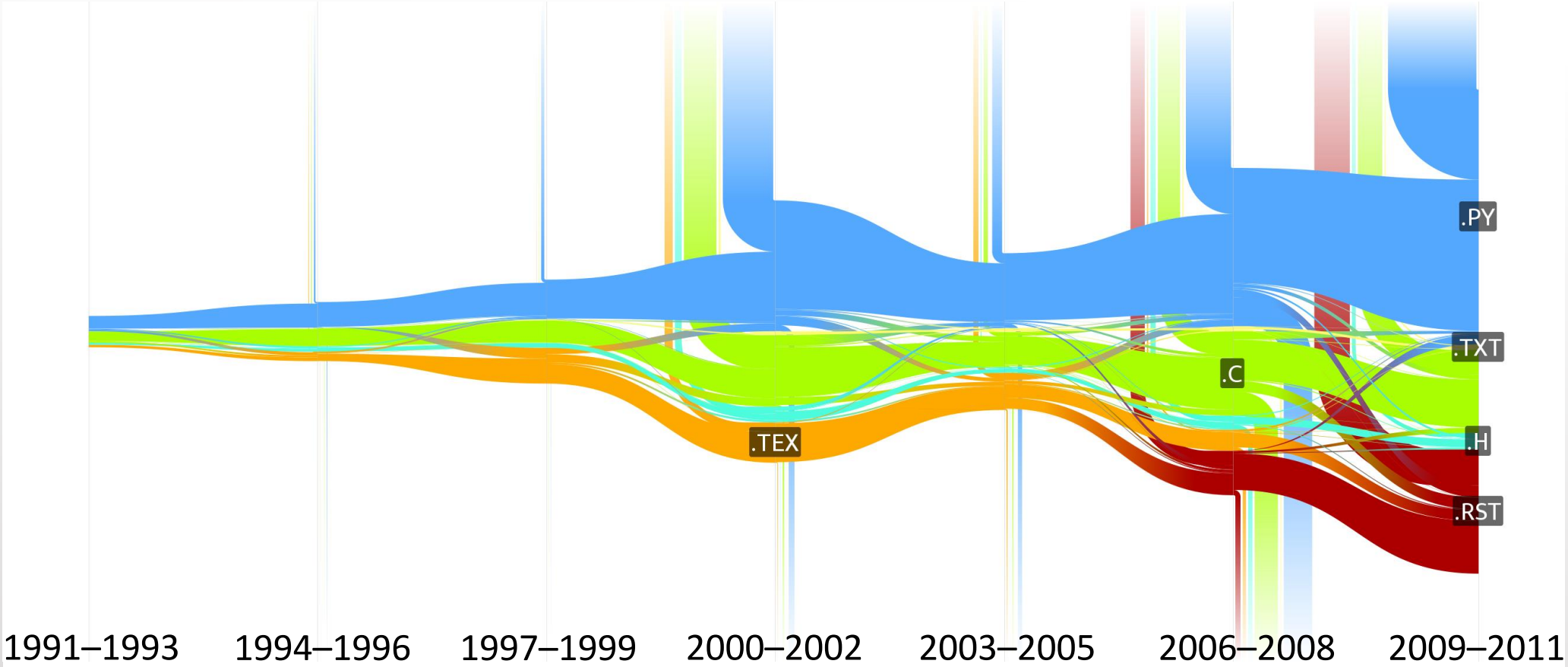
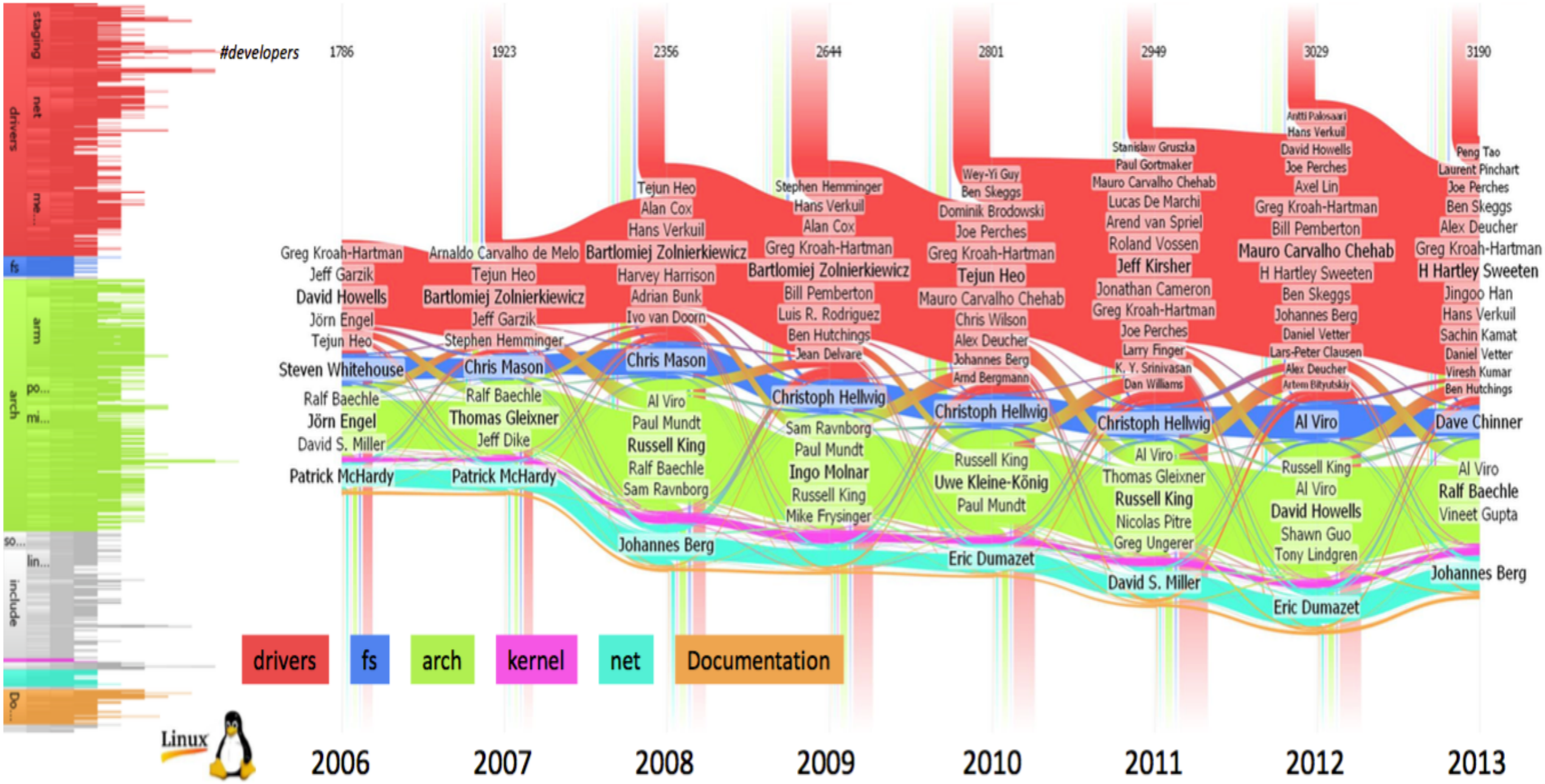
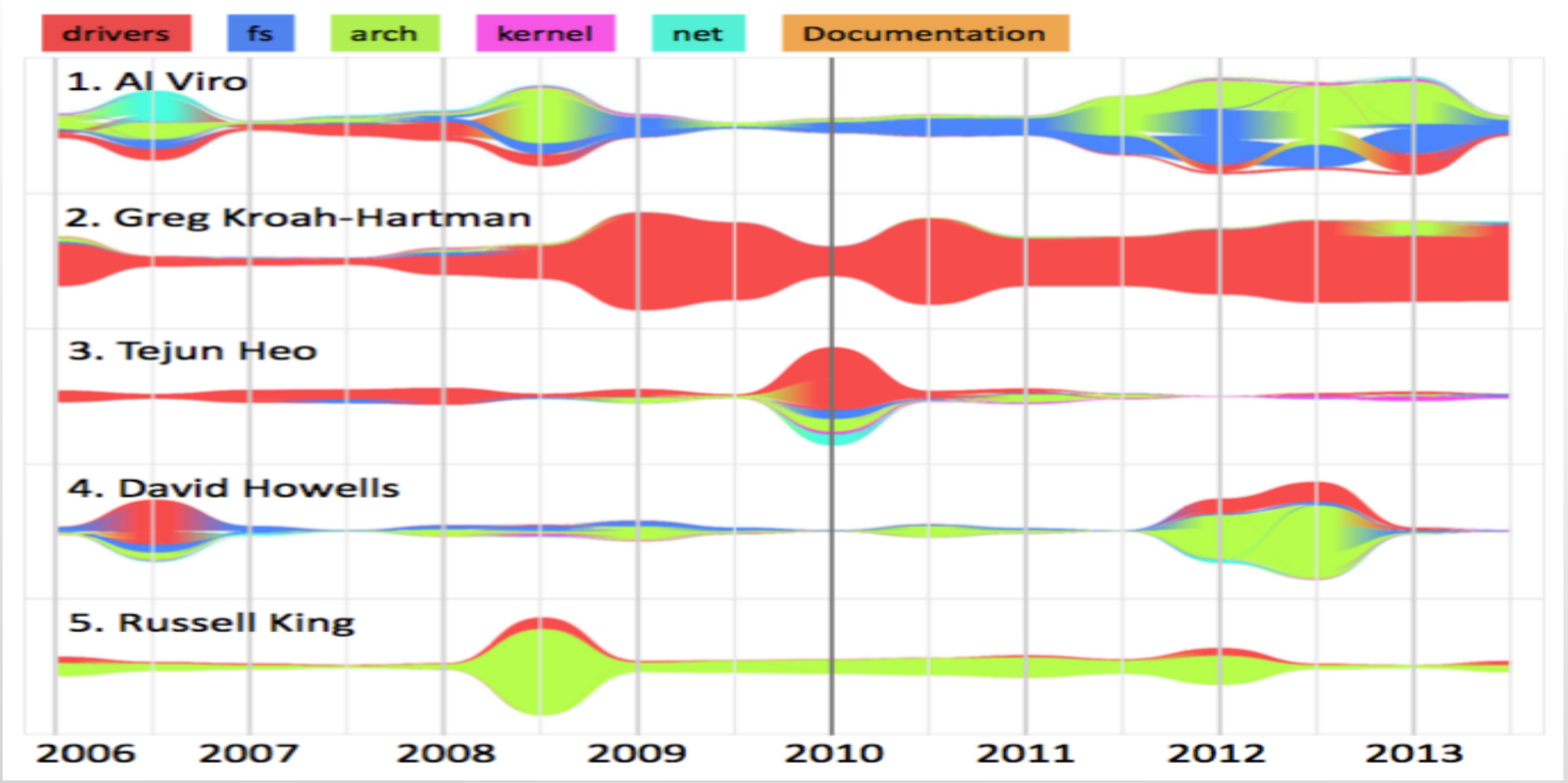


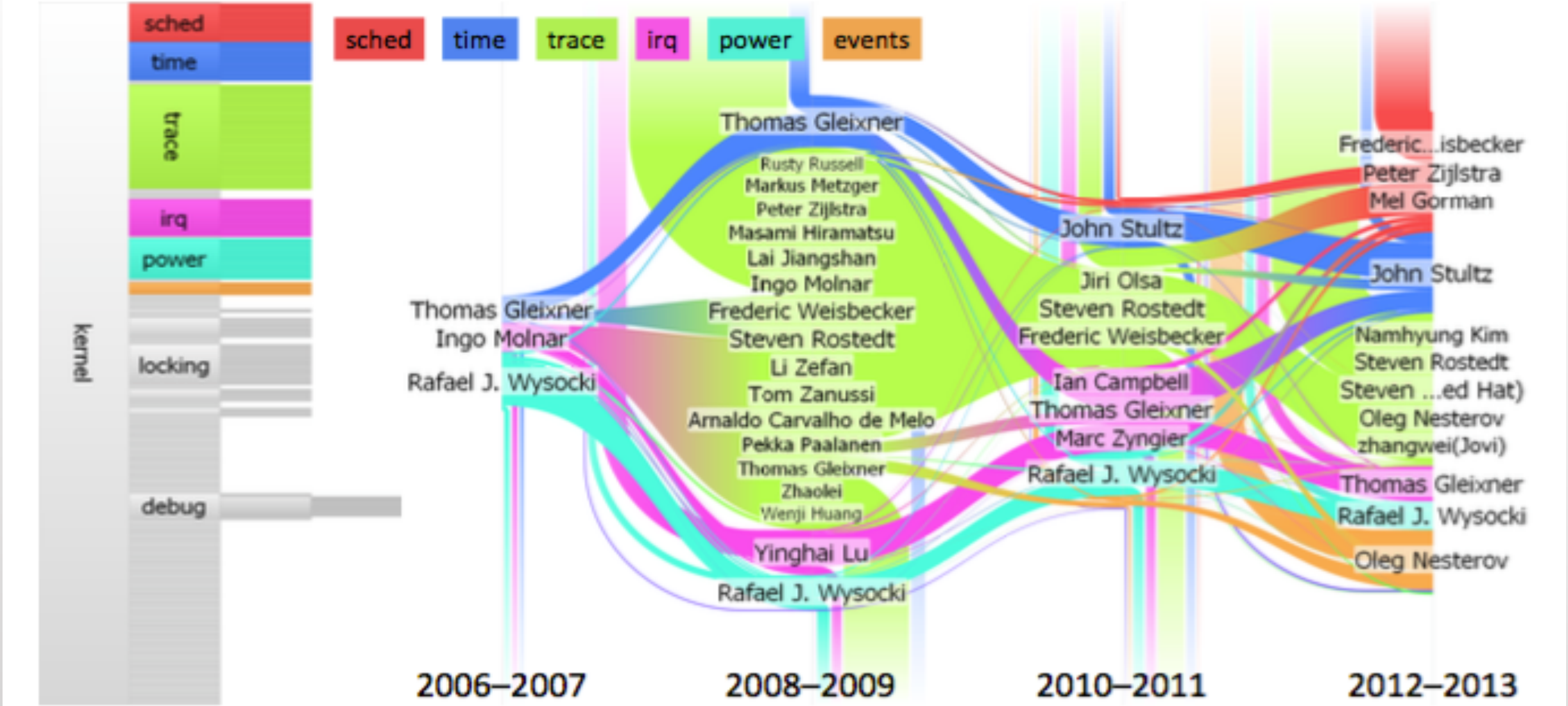
Fig. 4. Python *file type overview*, 1991–2011, 3-years interval.

Python file type overview


#developers
1786
1923
2356
2644
2801
2949
3029
3190
drivers
fs
arch
kernel
net
Documentation
2006
2007
2008
2009
2010
2011
2012
2013
Linux
Greg Kroah-Hartman
Jeff Garzik
David Howells
Jörn Engel
Tejun Heo
Steven Whitehouse
Ralf Baechle
Jörn Engel
David S. Miller
Patrick McHardy
Arnaldo Carvalho de Melo
Tejun Heo
Bartlomiej Zolnierkiewicz
Jeff Garzik
Stephen Hemminger
Chris Mason
Ralf Baechle
Thomas Gleixner
Jeff Dike
Patrick McHardy
Tejun Heo
Alan Cox
Hans Verkuil
Bartlomiej Zolnierkiewicz
Harvey Harrison
Adrian Bunk
Ivo van Doorn
Chris Mason
Al Viro
Paul Mundt
Russell King
Ralf Baechle
Sam Ravnborg
Johannes Berg
Stephen Hemminger
Hans Verkuil
Alan Cox
Greg Kroah-Hartman
Bartlomiej Zolnierkiewicz
Bill Pemberton
Luis R. Rodriguez
Ben Hutchings
Jean Delvare
Christoph Hellwig
Sam Ravnborg
Paul Mundt
Ingo Molnar
Russell King
Mike Frysinger
Johannes Berg
Wey-Yi Guy
Ben Skeggs
Dominik Brodowski
Joe Perches
Greg Kroah-Hartman
Tejun Heo
Mauro Carvalho Chehab
Chris Wilson
Alex Deucher
Johannes Berg
Arnd Bergmann
Christoph Hellwig
Russell King
Uwe Kleine-König
Paul Mundt
Eric Dumazet
Stanislaw Gruszka
Paul Gortmaker
Mauro Carvalho Chehab
Lucas De Marchi
Arend van Spriel
Roland Vossen
Jeff Kirsher
Jonathan Cameron
Greg Kroah-Hartman
Joe Perches
Larry Finger
K. Y. Srinivasan
Dan Williams
Christoph Hellwig
Al Viro
Thomas Gleixner
Russell King
Nicolas Pitre
Greg Ungerer
David S. Miller
Eric Dumazet
Antti Palosaari
Hans Verkuil
David Howells
Joe Perches
Axel Lin
Greg Kroah-Hartman
Bill Pemberton
Mauro Carvalho Chehab
H Hartley Sweeten
Ben Skeggs
Johannes Berg
Daniel Vetter
Lars-Peter Clausen
Alex Deucher
Artem Bityutskiy
Al Viro
Russell King
Al Viro
David Howells
Shawn Guo
Tony Lindgren
Eric Dumazet
Peng Tao
Laurent Pinchart
Joe Perches
Ben Skeggs
Alex Deucher
Greg Kroah-Hartman
H Hartley Sweeten
Jingoo Han
Hans Verkuil
Sachin Kamat
Daniel Vetter
Viresh Kumar
Ben Hutchings
Dave Chinner
Al Viro
Ralf Baechle
Vineet Gupta
Johannes Berg
staging
drivers
net
me...
fs
arch
arm
po...
mi...
so...
include
lin...
Do...



Linux Developer Sparkline of top 5 developers



Linux Subsystem details of Tools Directory

Future Improvements

- Show us how do the author organize the data(Data->Matrix)
- Show how to transfer the data into influents and effluents (Matrix->Influents)
- Provide tool ready for practitioners who can use developer river directly(No description about how to tackle the dataset)
- Distinguished colors may be up to 10 colors, otherwise graph may be hard to see
- Transfer the way to study software engineering research into social-technical aspects of engineering research