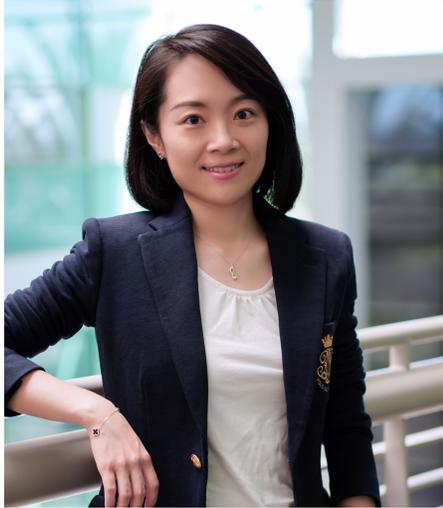


# How Has Forking Changed in the Last 20 Years?

## A Study of Hard Forks on GitHub

Shurui Zhou, Bogdan Vasilescu, Christian Kästner

**Carnegie Mellon University**



**Shurui Zhou**

University of Toronto

Assistant Prof. (Fall 2020)



The Edward S. Rogers Sr. Department  
of Electrical & Computer Engineering  
**UNIVERSITY OF TORONTO**



**Bogdan Vasilescu**

Software Engineering Ph.D. Program

**Carnegie  
Mellon  
University**



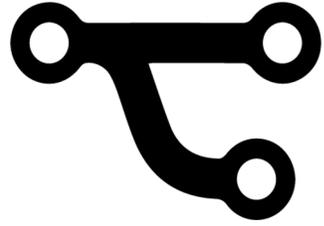
institute for  
**SOFTWARE  
RESEARCH**



**Christian Kästner**

# Forking

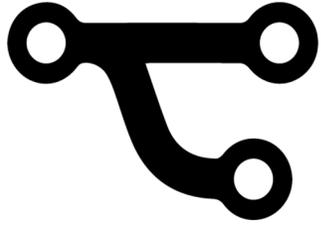
**Upstream**



**Fork/Branch**

# Traditional Notion of Forking

Upstream



Fork/Branch



→ Splitting off a community

A need of a community that was not fulfilled by the original project.

# Motivations for Forking

- Technical reason



# Motivations for Forking

- Technical reason
- Governance disputes



## A Docker Fork: Talk of a Split Is Now on the Table

29 Aug 2016 12:30pm, by [Alex Williams](#) and [Joab Jackson](#)



Hudson

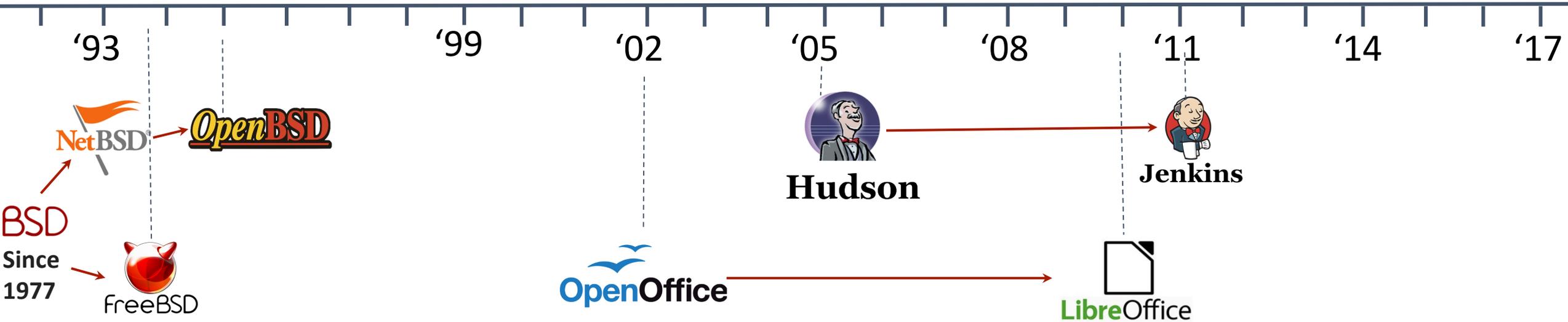


Jenkins

# Motivations for Forking

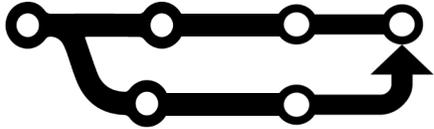
- Technical reason
- Governance disputes
- Discontinuation of the original project
- Commercial forks
- Legal reasons
- Personal reasons

# Timeline of Some Open-Source Forking Events



Fork-Based Development  
Changed Everything

# Fork-Based Development



→ Fork a repository to start **CONTRIBUTE** to a project [1].



[1] Fork a repo. <https://help.github.com/en/github/getting-started-with-github/fork-a-repo>

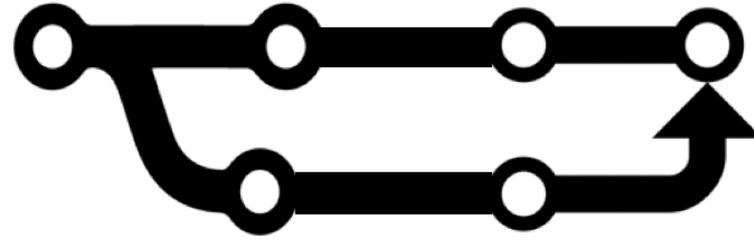
# Fork-based Dev. Becomes Popular

#Forks	#GitHub Projects
>50	114,120
>500	9164
>1,000	2236
>5,000	198
>10,000	72
>100,000	2



[GHTorrent 2019-06]

# Different kinds of Forks



**Hard Fork**



**(Social) Fork**

# Controversial Discussion of Hard forks



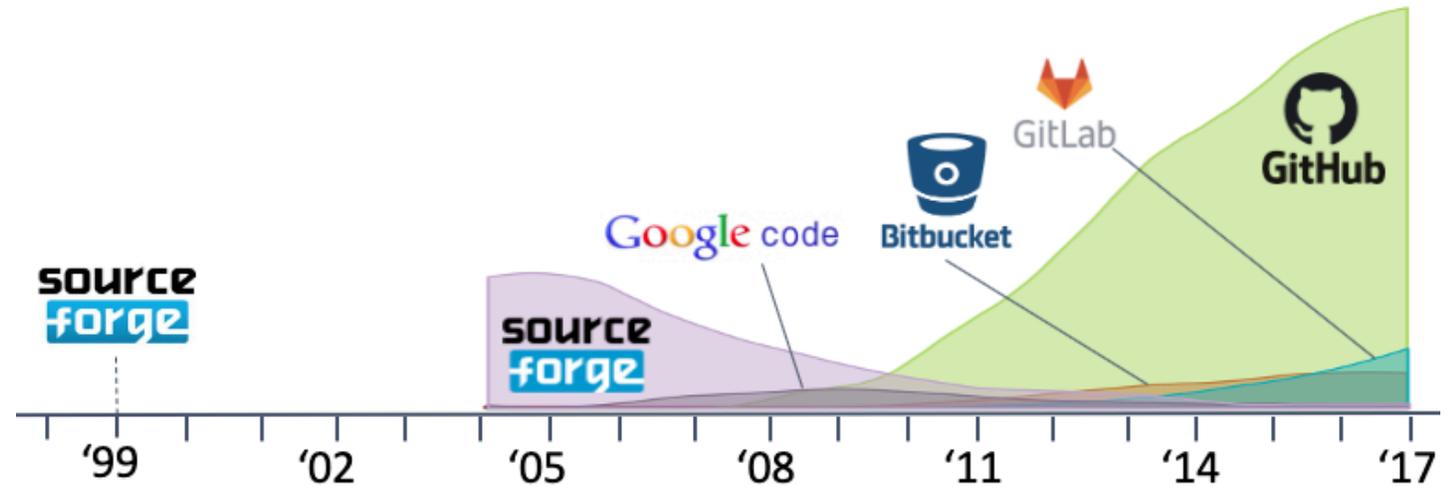
Free and open-source licenses  
Guaranteeing flexibility  
Fostering disruptive innovations



Fragment a community  
Lead to confusion for both maintainer and contributors



# Fork-Based Dev. Changed Everything

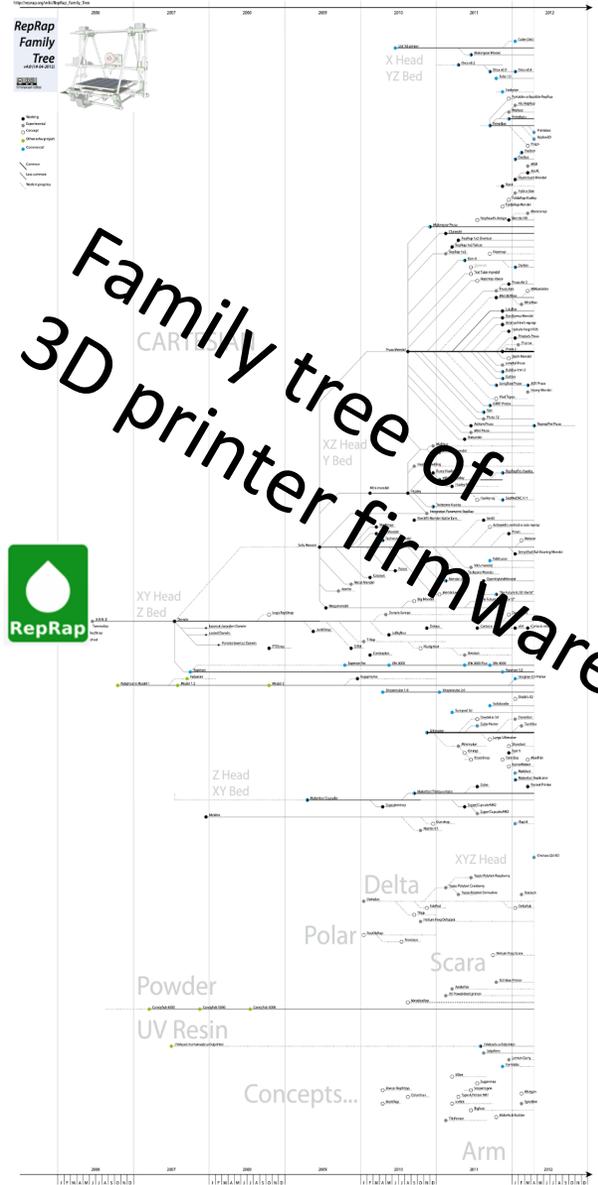


Hard Fork

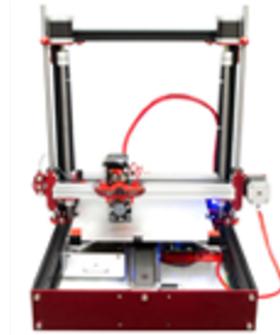
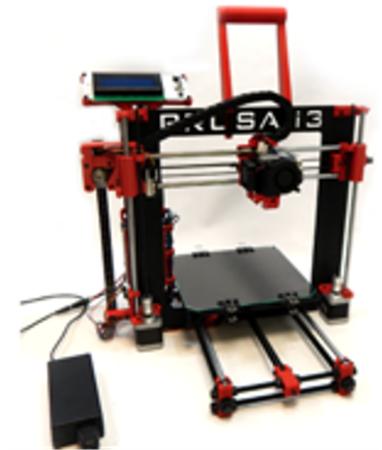
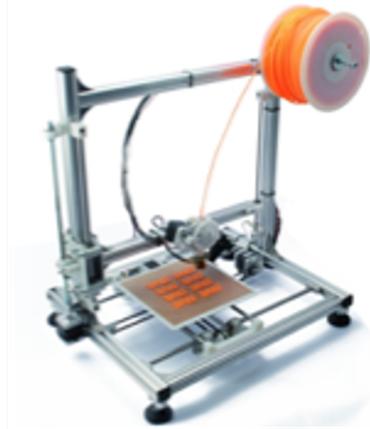
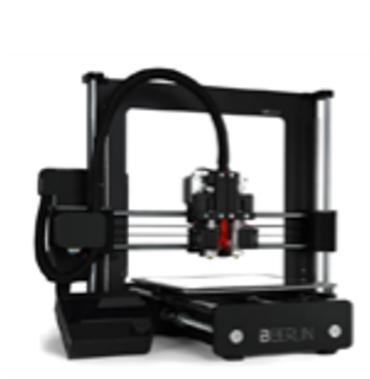


(Social) Fork

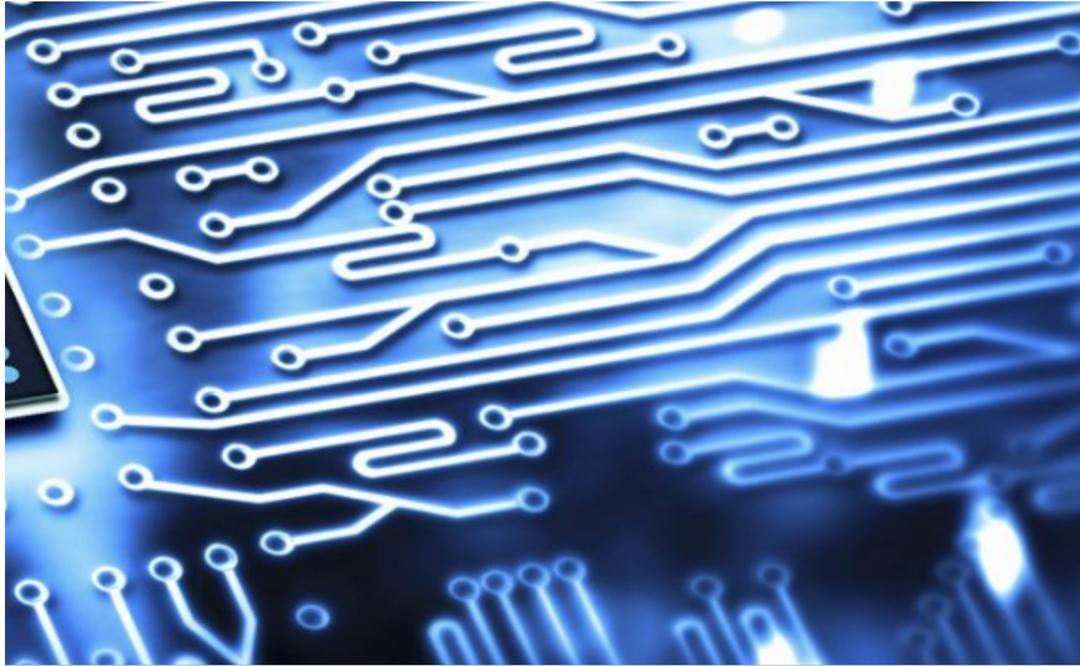
# Hard Forks in Social Coding Era



Family tree of 3D printer firmware



# Hard Forks in Social Coding Era



Behind the Scenes Bytes

## 3D Printer Firmware – Which to Choose and How to Change It?



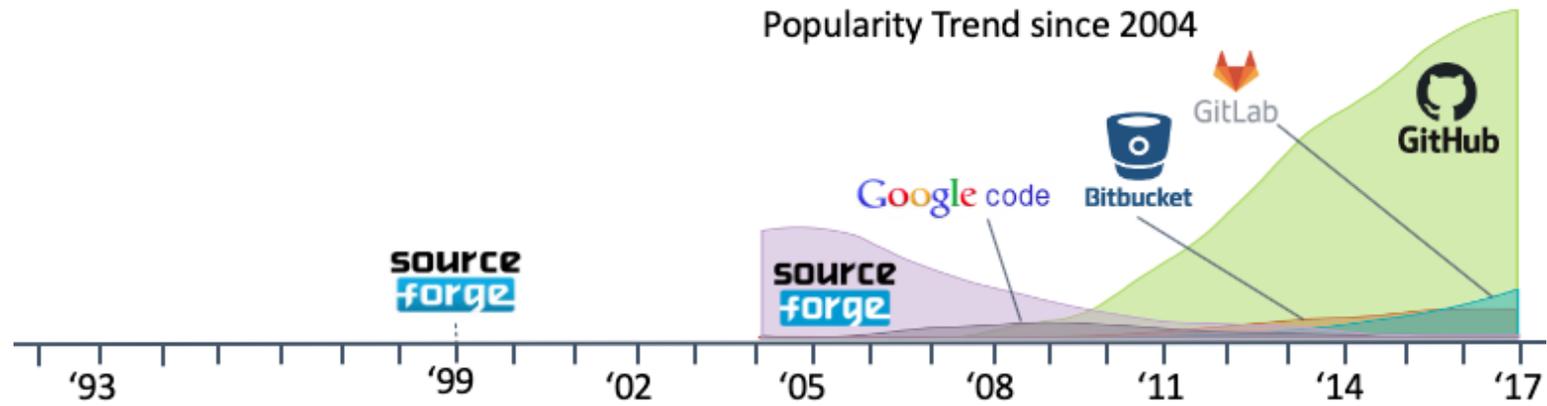
by Michael Jones  
Apr 4, 2018

# Research Question

How have perceptions and practices around hard forks changed?

# Research Question

How have perceptions and practices around hard forks changed?



## Motivations for Forking

- Technical
- Governance disputes
- Discontinuation of the original project
- Commercial forks
- Legal reasons
- Personal reasons



# Mixed Methods



Repository Mining



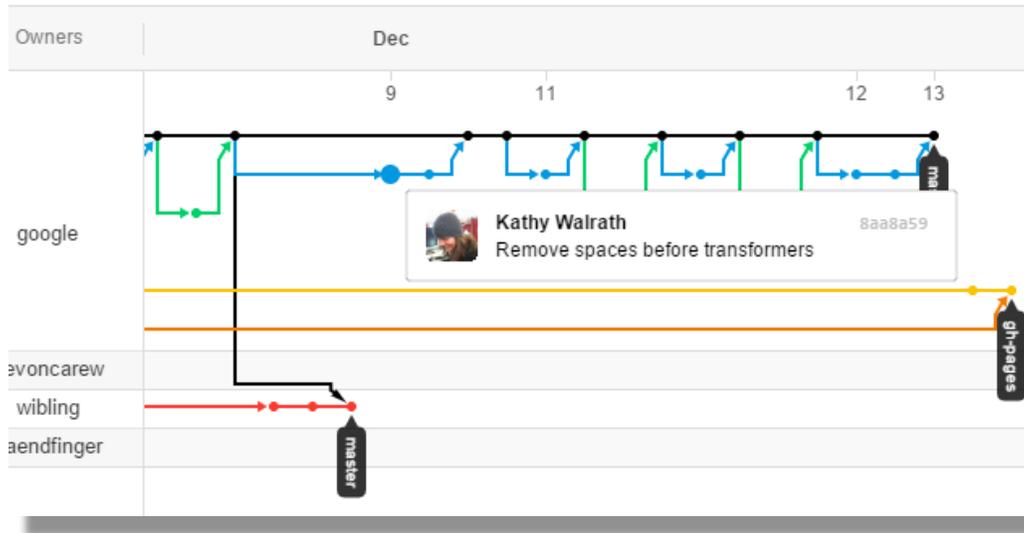
Interview



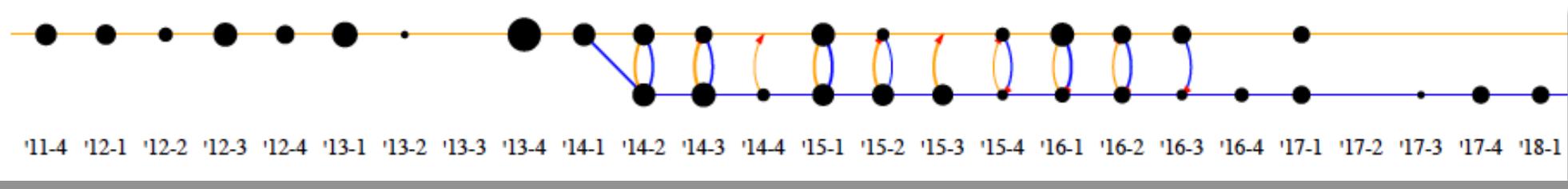
Repository Mining

- Heuristics to identify candidate hard forks
- Filtering false positives
- Card sorting

# Visualizing Fork Activities

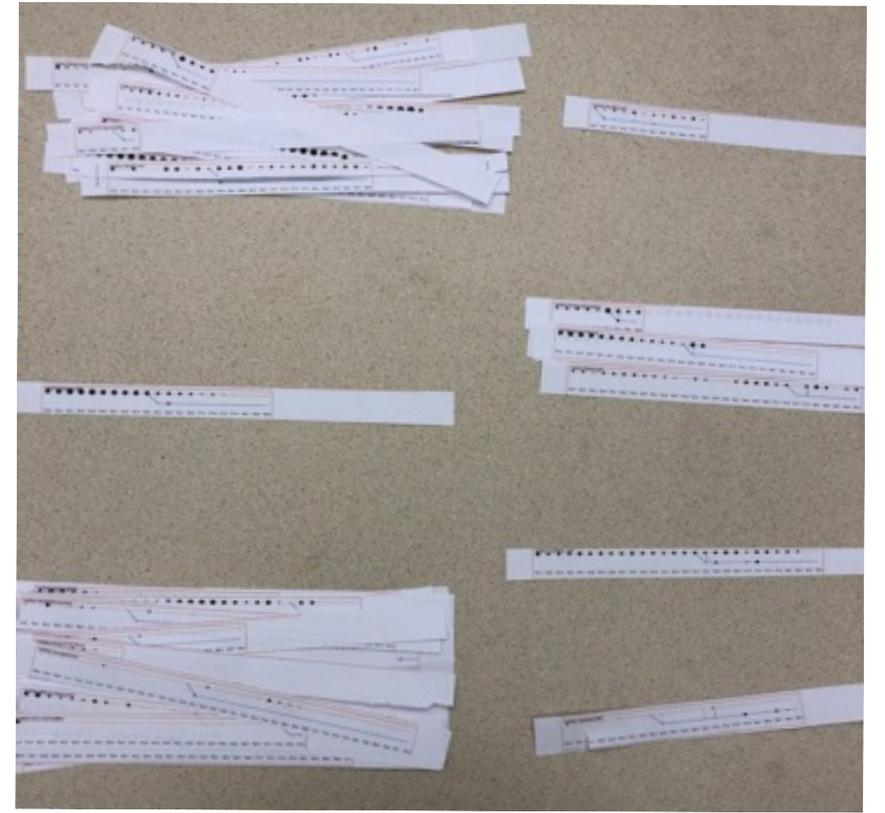


Commit history of both fork and upstream



Commit graph of fork: *tmyroadctfig/jnode*

# Identifying Evolution Patterns (Card Sorting)



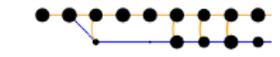
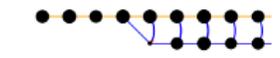
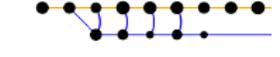
# Identifying Evolution Patterns of Hard Forks

Id	Category	Total	Sub-category	Example	Count
1	Revive Dead Project	623	Success (F. active > 2 Qt)	Upstream remains inactive	576
2			Upstream active again	56	
3			No Success	420	420
4	Both Alive	723	only merge	26	
5			only sync	107	
6			merge & sync	28	
7			no interaction	562	
8	Forking Active Project	7280	only merge	174	
9			only sync	686	
10			merge & sync	107	
11			no interaction	6313	
12	Fork does not out live upstream	6251	only merge	388	
13			only sync	762	
14			merge & sync	199	
15	no interaction	4902			

- 15 evolution patterns
- 15,306 hard forks

Covering 97.7 % of all hard forks

# Result: Frequency of Hard Forks

Id	Category	Total	Sub-category	Example	Count
1	Success	623	Upstream remains inactive		
2	(F. active > 2 Qt)		Upstream active again		
3	No Success	420			
4	Both Alive	723	only merge		26
5			only sync		107
6			merge & sync		28
7			no interaction		562
8	Fork Lived Longer	7280	only merge		174
9			only sync		686
10			merge & sync		107
11			no interaction		6313
12	Fork does not out live upstream	6251	only merge		388
13			only sync		762
14			merge & sync		199
15			no interaction		4902

Most hard forks are created as forks of active projects (14,254 hard forks, 93 %)

# Result: Frequency of Hard Forks

Id	Category	Total	Sub-category	Example	Count
1	Revive Dead Project (F. active > 2 Qt)	623	Success Upstream remains inactive		576
2			Upstream active again		56
3			No Success	420	
4	Both Alive	723	only merge		
5			only sync		
6			merge & sync		
7			no interaction		
8	Forking Active Project Fork Lived Longer	7280	only merge		
9			only sync		
10			merge & sync		
11			no interaction		
12	Fork does not out live upstream	6251	only merge		
13			only sync		
14			merge & sync		
15			no interaction		

A substantial number of cases where hard fork are created to revive a dead project (1,052 hard forks, 6.8 %)

# Result: Frequency of Hard Forks

Id	Category	Total	Sub-category	Example	Count	
1	Success	623	Upstream remains inactive		576	
2	(F. active > 2 Qt)		Upstream active again		56	
3	No Success	420				
4	Both Alive	723	only merge		26	
5			only sync		107	
6			merge & sync		28	
7			no interaction		562	
8				only merge		
9	Forking Active Project	7280	Fork Lived Longer	only sync		
10			merge & sync			
11			no interaction			
12				only merge		
13	Fork does not out live upstream	6251				only sync
14			merge & sync			
15			no interaction			

Both upstream and hard fork remain active for extended periods of time are not common (779 hard forks, 5%)

# Result

Id	Category	Total	Sub-category	Example	Count	Interviewees
1	Revive Dead Project	623	Success (F. active > 2 Qt)	Upstream remains inactive	576	P12
2			Upstream active again	56		
3			No Success	420		
4	Both Alive	723	only merge		26	
5			only sync			
6						
7	Forking Active Project	6251			6313	P6, P8, P11
8			only merge		388	
9			only sync		762	
10			merge & sync		199	
11			no interaction		4902	

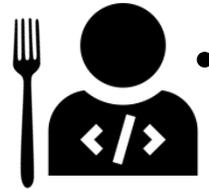
**A rare phenomenon**  
**Only 15,306 hard forks, 0.2 % of GitHub's**  
**47 million forks have 3 or more stars.**

hard forks  
 forks  
 of  
 forks

# Interview 18 Upstream & Hard Fork owners



7% response rate



Fork owner

- decision process that lead to hard fork
- relationship to the upstream project
- future plans



Owners of upstream: “To what extent,...

- aware of/interact with/monitor hard forks
- concern/take steps to avoid hard forks

# Result: Why Hard Forks Are Created

Align well with prior findings.

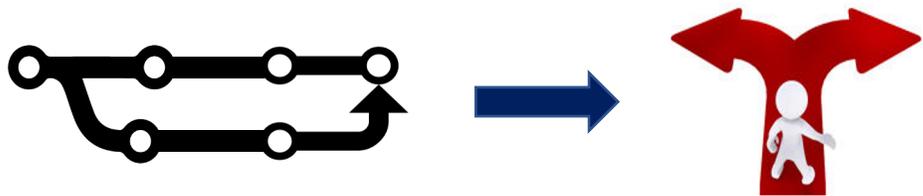


## Motivations for Forking

- Technical
- Governance disputes
- Discontinuation of the original project
- Commercial forks
- Legal reasons
- Personal reasons



# Result: Why Hard Forks Are Created



Common obstacles :

- Unresponsive maintainers (P1, P2, P8)
- Rejected pull requests (P11, P13, P14)

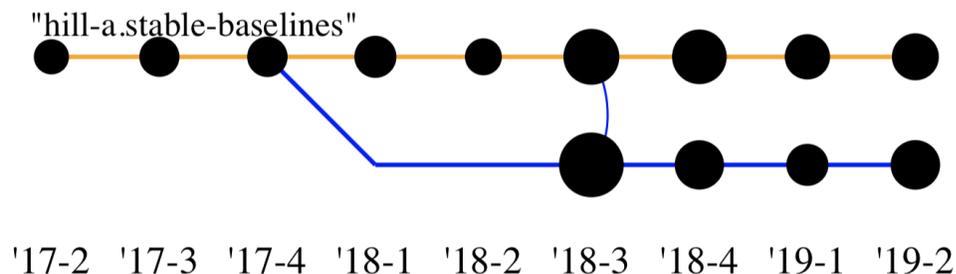


*P2: "before forking, we started by opening issues and pull requests, but there was a lack of response from their part. [We] got some news only 2 months after, when our fork was getting some interest from others."*

**upstream:** *openai/baselines*

**P2:** *hill-a/stable-baselines*

*(has 463 second-level forks)*



# Hard forks are not likely to be avoidable



The stigma  
around hard  
forking is gone!



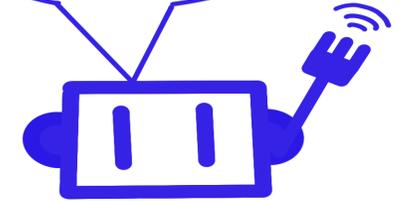
with concern about  
community fragmentation

# Tooling Opportunities

- Considering multiple forked projects as part of a larger community
  - A bot to monitor emerging hard forks
  - Identify the intention behind a fork

The hard fork fixed bug #123 (high priority)!

Found a hard fork!  
*shuiblue/fragment*

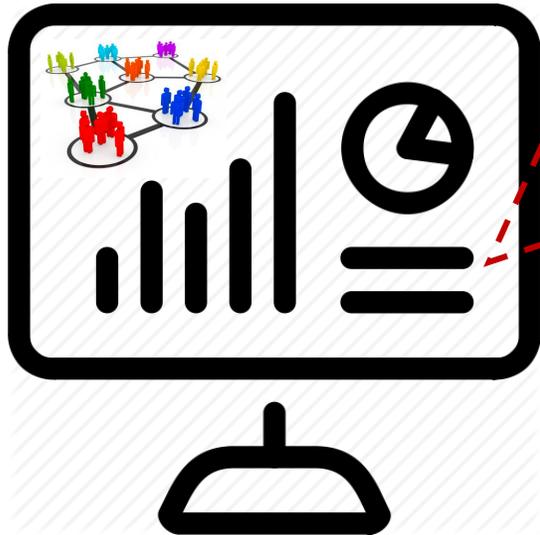
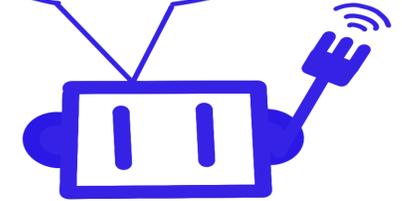


# Tooling Opportunities

- Considering multiple forked projects as part of a larger community.
  - A bot to monitor emerging hard forks
  - Identify the intention behind a fork
  - Dashboard to show how multiple projects and important hard forks interrelate

The hard fork fixed bug #123 (high priority)!

Found a hard fork!  
*shuiblue/fragment*



Date	Activity	Participants
2021-06-11	repo1 cross-referenced 2 PRs to repo2	usr1, usr13
2021-06-13	repo3 has 105 more stars	usr100... usr205
2021-07-01	repo4 submitted PR#234 to repo2 (35 commits), got rejected	usr50, usr89
2021-07-05	12 contributors from repo2 migrate to repo 4	usr20, ...
...	...	...

## Identifying Evolution Patterns of Hard Forks

Id	Category	Total	Sub-category	Example	Count
1	Revive Dead Project	623	Success (F. active > 2 Qt)		576
2			Upstream remains inactive		56
3			No Success		420
4	Both Alive	723	only merge		26
5			only sync		107
6			merge & sync		28
7			no interaction		562
8	Forking Active Project	7280	only merge		174
9			only sync		686
10			merge & sync		107
11			no interaction		6313
12			only merge		388
13	Fork does not out live upstream	6251	only sync		762
14			merge & sync		199
15			no interaction		4902

- 15 evolution patterns
- 15,306 hard forks



Par.	Domain	#Stars(U)	#Stars(F)	LOC	Role	Exp.(yr)
P1	Blockchain	<20	<10	10K	F	19
P2	Reinforcement learning	10K	1K	30K	F	3
P3	Mobile processing	-	70	20K	F	6
P4	Video recording	-	160	300K	F	18
P5	Helpdesk system	2K	<10>	800K	F	5
P6	CRM system	30	260	800K	F	10
P7	Physics engine	-	360	100K	F	15
P8	Social platform	500	230	500K	F	20
P9	Reinforcement learning	<20	<20	30K	2nd-F	3
P10	Game Engine	500	<10	200K	2nd-F	21
P11	Networking	300	160	500K	F	10
P12	Email library	-	10K	20K	F/U	32
P13	Game engine	3K	70	20K	F	11
P14	Machine learning	30K	50	60K	F	8
P15	Image editing	70	<10	20K	F	20
P16	Image editing	70	<10	20K	U	10
P17	Microcontrollers	9K	1K	300K	U	6
P18	Maps	400	<10	100K	U	9

F: Hard Fork Owner; U: Upstream Maintainer; 2nd-F: Fork of the Hard Fork  
 \*Some of the upstream projects are not in GitHub, so the number of stars is unknown.  
 Numbers rounded to one significant digit.



# The stigma around hard forking is gone!

## Tooling Opportunities

- Considering multiple forked projects as part of a larger community.
  - A bot to monitor emerging hard forks
  - Identify the intention behind a fork
  - Dashboard to show how multiple projects and important hard forks interrelate

The hard fork fixed bug #123 (high priority)!

Found a hard fork! shuiblue/fragment



Date	Activity	Participants
2021-06-11	repo1 cross-referenced 2 PRs to repo2	usr1, usr13
2021-06-13	repo3 has 105 more stars	usr100... usr205
2021-07-01	repo4 submitted PR#234 to repo2 (35 commits), got rejected	usr50, usr89
2021-07-05	12 contributors from repo2 migrate to repo 4	usr20, ...
...	...	...

