Longitudinal Analysis of Collaboration in Forked Open Source Software Development Projects

Amirhosein (Emerson) Azarbakht
School of Electrical Engineering and Computer Science
Oregon State University
azarbaka@oregonstate.edu

Great Projects

Assumption: You know what open source is.

- Technical quality
- Operational health
- Survivability
  - Bus #

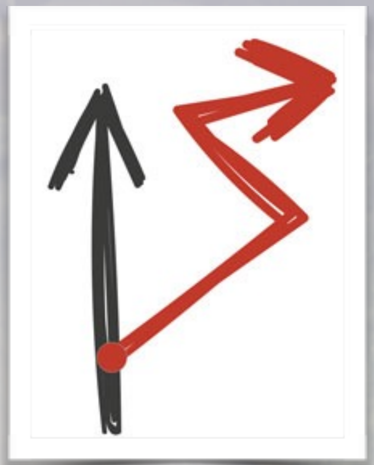
Open source is great
but not all the time

- Not everything works smoothly all the time in open source projects
  - **Problems**:
    - Uncertainty
    - “Industry afraid of "fly-by-night" open source packages & vendors

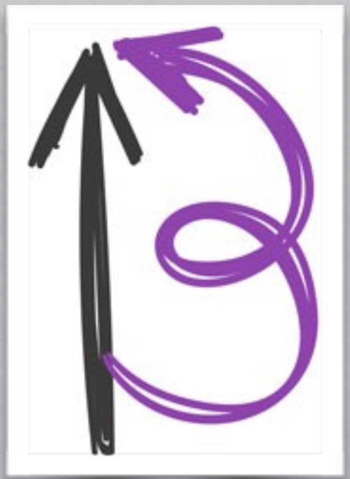
Forking Categories

Communities dissolve or evolve for a variety of different reasons:

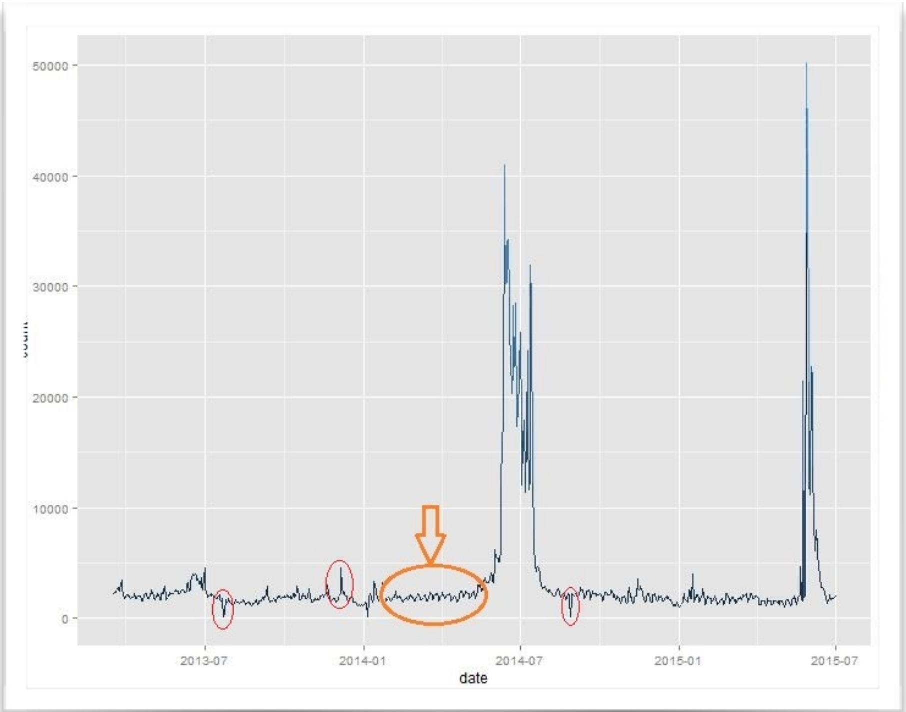
- **Conflict-driven**
  - Destructive
  - Lead to resentment
  - Become competitors; “you’re either with us, or against us”



- **Non-conflict-driven**
  - Developers interested in exploring different functionalities
  - Not necessarily creating an all-or-nothing type of thing;
  - Developers participate in multiple communities, as long as it’s a constructive work

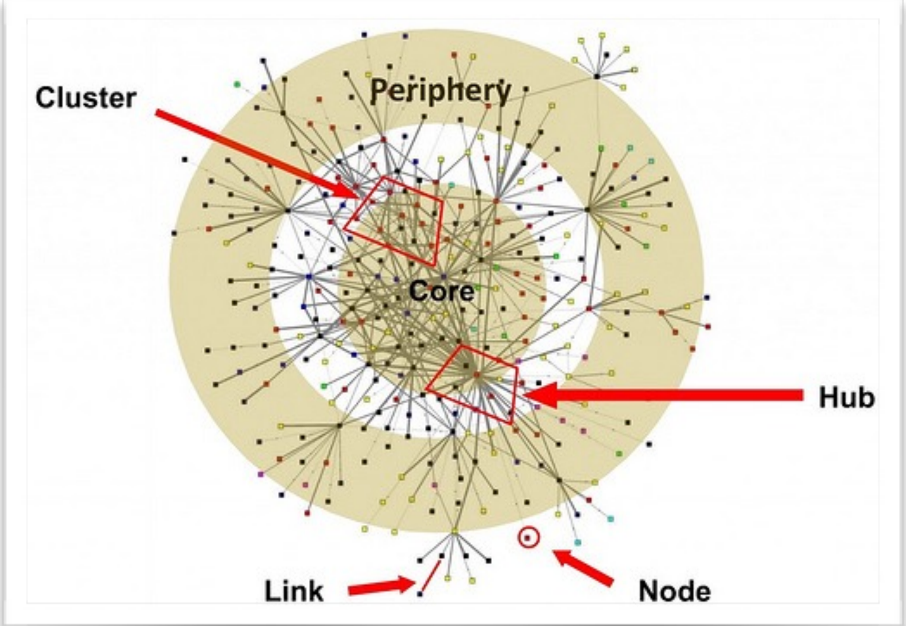


Methodology



Two analyses:

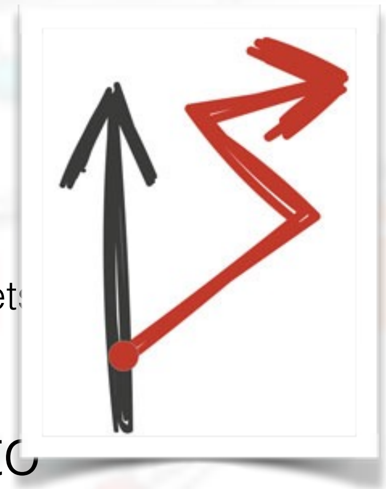
- Time series analysis of developers' communication sentiments
- Social network analysis of developers' interaction graphs

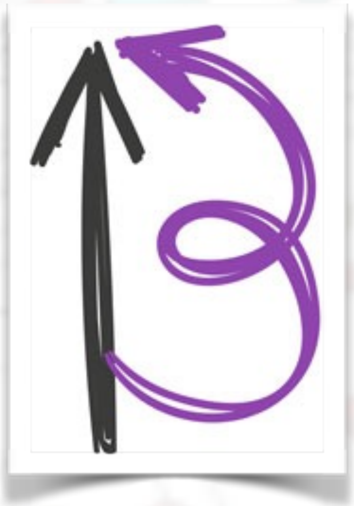


Results

Social network analysis of open source developers' interaction graphs

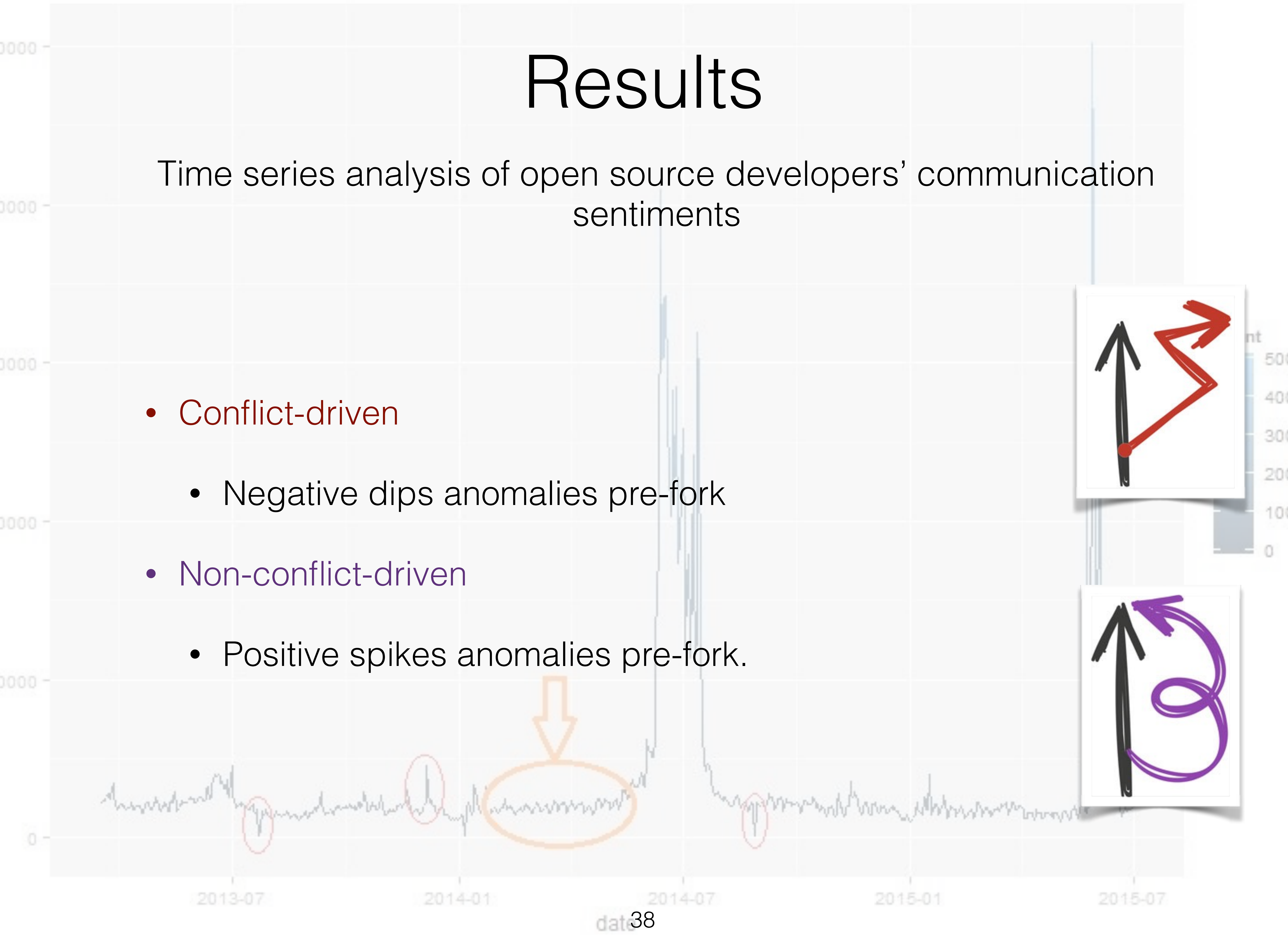
- Conflict-driven
  - Local hierarchy (Negative three-cycles (opposite of hierarchy) and positive transitive triplet
  - Tendency of heavy mailing-list poster developers to be tied to other heavy mailing-list poster developers (Out-out degree assortativity)
- Non-conflict-driven
  - Tendency of heavy code contributors to be tied to other heavy code contributors (Developer's source code activity)



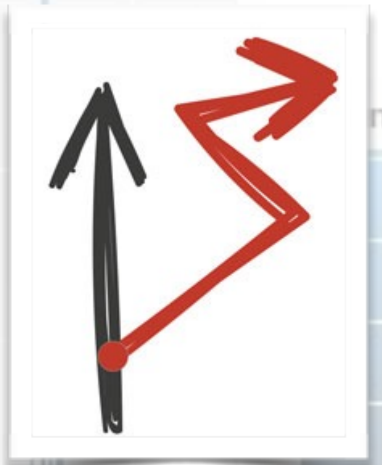


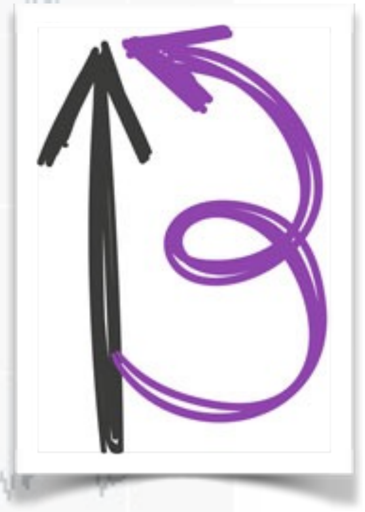
Results

Time series analysis of open source developers' communication sentiments



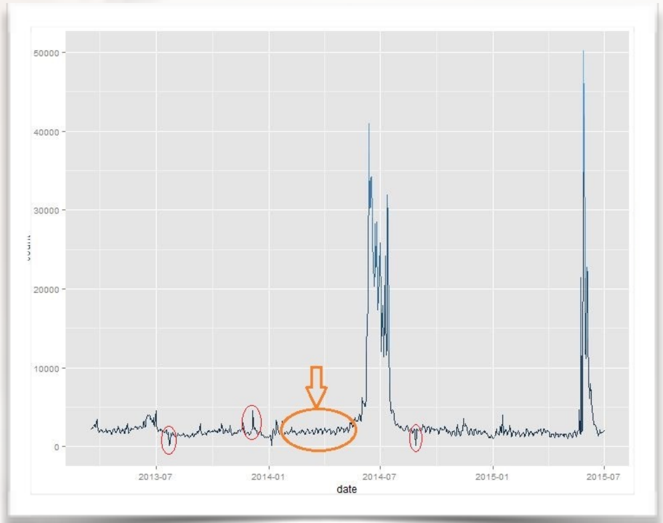
- Conflict-driven
    - Negative dips anomalies pre-fork
- Non-conflict-driven
    - Positive spikes anomalies pre-fork.

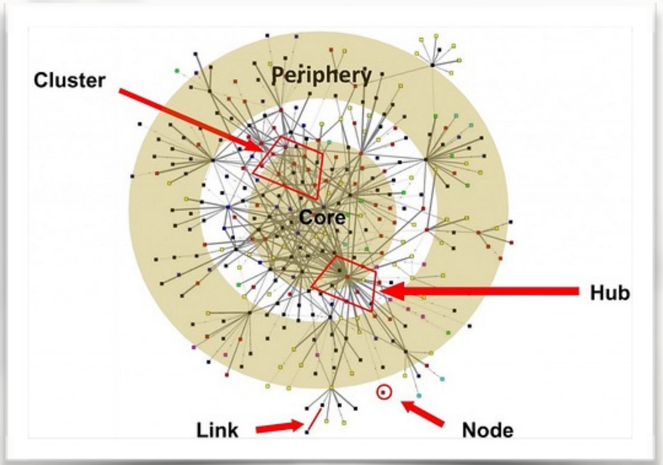




Future Work

- More projects needed
- Does more transparency result in project self-regulation?


date


Cluster
Periphery
Core
Hub
Link
Node