How co-located events enhance online community building & long-term production

Thomas Maillart : maillart@berkeley.edu

Berkeley UNIVERSITY OF CALIFORNIA

Motivation : Understand how the subtle and fast-paced social interactions occurring during co-located events (e.g., hackathon, conferences) influence long-term online production and collaboration. Following Orlikowski and Yates (2002), focus on *special* (*Kairos)* vs. *routine* (*Kronos)* moments, and how the former influence the latter.

Focus : 5 hackathons in the field of astrophysics (organized between 2014 and 2016), for which large parts of resulting knowledge artifacts produced have been posted publicly on GitHub.

Data : Github activity data
- Repository creations (contribution cascade initializations)
- Contributions per repository (contribution cascade sizes)

Method :
- Repositories contributed by a single individual C_i and collectively C_s are considered separately.
- Cascade sizes (number of contributions per repository) are considered similarly.

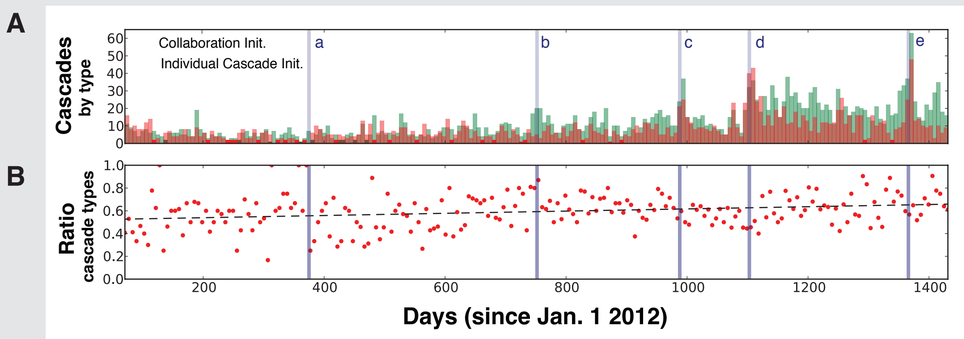
Preliminary Results :

- Hackathons boost repository creations, following $\sim (t - t_h)^{0.7}$ with t_h the date of the hackathon. However, some hackathons boost individual contribution cascades, while others boost collective work immediately **(Figure A)**.
- Overall, hackathons contribute significantly to the long term relative increase of collaboration as defined by $r = C_s / (C_i + C_s)$ **(Figure B).**
- Collective repositories generate cascades of highly skewed distributed sizes, in comparison with individual contribution cascades (log-normal distribution **(Figure C).**

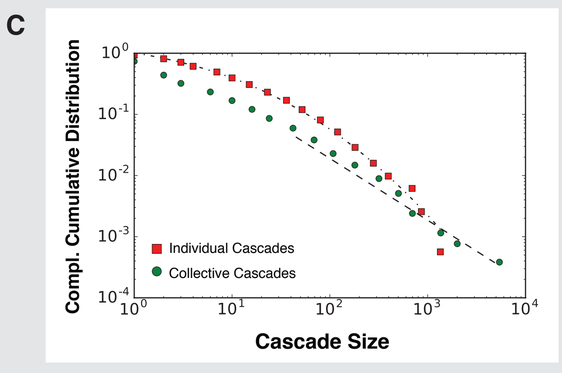
Importance : Even though social interactions during *special* co-located events can hardly be measured and quantified, one can measure their online long-term implications on *routine* production and on the establishment of fruitful long-term collaboration. Collaborative endeavors are likely to generate much larger, potentially orders magnitude more contributions, compared to individual initiatives. Here, 5 hackathons have contributed to the long-term establishment of collaboration ties in the community of astrophysicists, and on average, the relative share of collaboration projects can be precisely measured (3.5% increase annually).

Co-Located Events :
- pair programming, informal discussions, social time.

Follow up Online Production :
- Individual vs. Collaborative Cascade Initializations



- Individual vs. Collaborative Cascade Sizes



The tail distribution of individual cascade sizes is best represented by a log-normal distribution with shape = 1.83 and scale = 5.5 obtained by Maximum Likelihood Estimation (MLE)

The tail distribution of collective cascade sizes is best represented by a power law distribution with exponent = 1 (Zipf's law) for sizes > 50 events.

In the latter case, the distribution seems unbounded, and the two first statistical moments (mean and variance) diverge has the sample grows. Hence, as time passes, it is increasingly likely that much larger cascades will occur.