On the Importance of Co-Located Events for Community Building and Long-Term Online Production

THOMAS MAILLART, UC Berkeley

Nearly all online communities organize co-located meetings, such as conferences, un-conferences, and hackathons. These events are short, fast-paced, yet they are intended to enable social interactions and fast-circulation of informal knowledge between attendants. There is however a dearth of knowledge on the contribution of co-located events to community enhancement and long-term online production. Here, I study a community of astrophysicists involved in open and reproducible data science. Over the span of data collected (4 years), five co-located meetings were organized. Each meeting triggered contrasted immediate effects regarding collaboration, but all of them had significant long-term enhancing effects on community building and online knowledge production. These results illustrate how punctual co-located meetings change the way contributors engage with their community once they have resumed their routine work online.

1. INTRODUCTION

Early on, organization science researchers have pointed out the importance of special *"kairos"* moments, with their own purpose and genuine timing, by opposition to *"kronos"* routine moments [Orlikowski and Yates 2002]. The development of online collaboration tools has created a striking contrast between these moments: routine work is mostly performed online, whereas co-located meetings address completely different problems, ranging from prototyping and showcasing [Trainer et al. 2014] to strengthening social ties [Ducheneaut 2005], teaching tools and on-boarding newcomers [Von Krogh et al. 2003], deliberating on an agenda for future work, as well as for solving community governance issues [O'Mahony and Ferraro 2007]. On the contrary, routine work consists mostly in a stream of prioritized issues with some degree of required collective coordination to solve them [Dabbish et al. 2012]. Thus, co-located events are less about immediate knowledge production than about organizing online collective production on the long-term, once community members have gone back home, sometimes all over the country or the world. Co-located events are characterized by numerous and fast-paced social interactions or pair-programming, involving small-talk as well as rapid exchange of knowledge between tens if not hundreds of people over short periods of time, usually ranging from one day to a week.

Although direct observation, quantification and modeling of subtle inter-personal interactions during physical is usually complicated [Choudhury and Pentland 2002; Pentland et al. 2005; Onnela et al. 2014], co-located events trigger cascades of contributions with long-term memory, which allow measuring their after-effects on community and building and follow-up online production. Furthermore, co-located meetings, and hackathons in particular, may be associated with super-linear productive bursts, which stem from *critical, self-sustained, cascades of contributions* among developers [Sornette et al. 2014], with similitudes with super-linear laws of productivity (e.g., number of patents) and interactions (e.g., phone calls) in cities [Bettencourt et al. 2007; Schläpfer et al. 2014].

Here, I study how co-located meetings generate cascades of contributions. These cascades may be of two kinds: Either they occur on a personal project (i.e., code repository), or they are contributed on a shared project. While the former contribution cascades reflect single-minded concentration on a problem, the latter cascades display the intention and capacity to interact with others. Both cascade types co-exist, but the amount of contribution cascades initialized on shared code repositories is a signature of a community having built social ties, and vividly sharing problems and solutions, during or following co-located events.

2. DATA AND METHOD

The present study is carried on for the American Astronomical Society (AAS) HackDays (2013, 2014 and 2015) and Astro Hack Weeks (2014 and 2015). The former events are intended to *"write code or work on some other project, fast, [and] fully execute it in one day"* and the latter events correspond to a *"week-long summer school, hack week, un-conference focused on astrostatistics and data-intensive astronomy"*.[1] The activity (i.e., both number of contributors and projects contributed to) and production (i.e., code and issue submissions) per time unit have been queried from the GitHub Archive.[2] And community members considered here are selected for their participations to at least one Astro Hack Week.

Considering the aforementioned co-located hackathons, I first compare the number of contribution cascades initiated on *personal* and *shared* code repositories, and how they have evolved over the period of scrutiny (from January 1st, 2012 until December 31st, 2015). I then show that contribution cascades are heavy-tailed (with a significant difference between personal and shared contributions). I finally provide insights on the overall flow of contributions by the community, again with a distinction between personal and shared contributions.

3. RESULTS

The community of astrophysicists involved in open and reproducible data science [Perez 2013] has been active for four years at least. Since January 1st, 2012, five co-located meetings were organized to shape a community (64 members in this study), and to build collectively programming and data handling tools to modernize their practice of astrophysics.

Each hackathon had its own impact on the community as measured by the number of contribution cascades initiated on personal and shared code repositories (resp. c_i and c_s; see Figure 1A). The first AAS HackDay **(a)** had limited immediate measurable impact but contributed to increase the baseline of c_i and c_s over the next year, while the following hackathons had a significant immediate impact on contribution cascade initializations, followed by a slow decay until the next co-located meeting. The decays are best fitted by power laws $\sim t^{-\alpha}$ with $\alpha \approx 0.7(1)$ for **(c)**, **(d)** and **(e)** (both for personal and shared contribution cascades), which is reminiscent of critical cascades in social media [Crane and Sornette 2008].

The ratio $r = c_s/(c_i + c_s)$ of contribution cascade initializations in shared repositories c_s over all cascade initializations increased linearly (3.5% per year; $p < 0.001$) and from $r_0 \approx 0.5$ in 2012 (see Figure 1B). The contribution of each hackathon to r is however contrasted: HackDays 2014 and 2015 (resp. **b** and **d**) had a long-standing impact, while Astro Hack Week 2014 **(c)** had rather a negative immediate

[1] Astro Hack Day, http://www.astrobetter.com/wiki/AAS_HackDay_2014; and Astro Hack Week, http://astrohackweek.github.io
[2] GitHub Archive, https://www.githubarchive.org/.

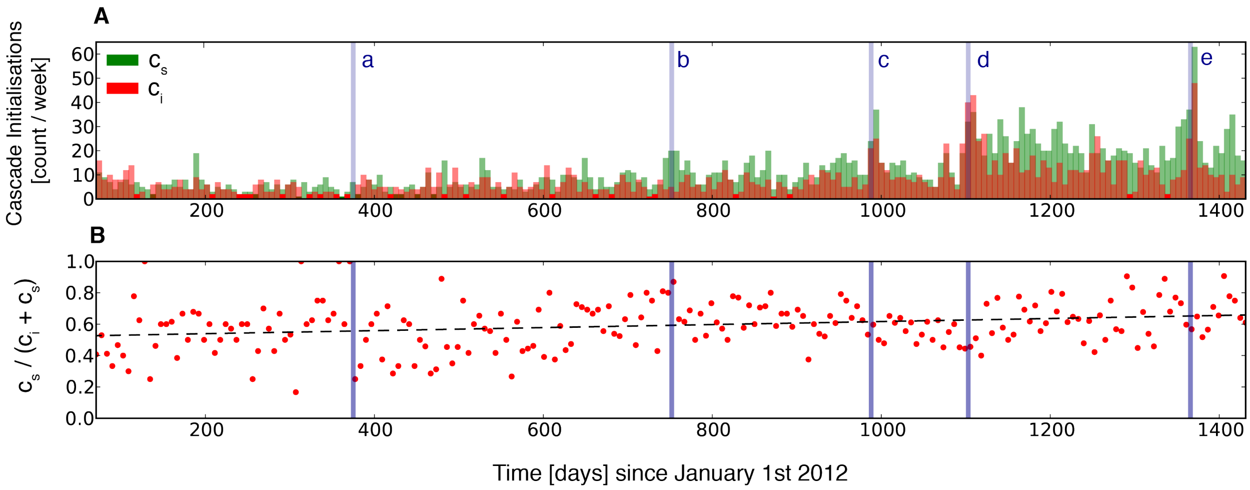

Fig. 1. **A.** Progression of individual vs. collective contribution cascade initializations (resp. c_i and c_s), punctuated by 5 hackathons: **(a)** HackDay 2013, **(b)** HackDay 2014, **(c)** Astro Hack Week 2014, **(d)** HackDay 2015, and **(e)** Astro Hack Week 2015. **B.** The share of collective contribution cascades initiated c_s over all cascade initializations ($c_i + c_s$) increases linearly (3.5%/year; $p < 0.001$; dashed line). It is important to note that $r = c_s/(c_i + c_s) > 0.5$ on average. As time passes, each hackathon triggers a larger impulse on cascade initializations.

impact on r and the situation for Astro Hack Week 2015 **(e)** remains unclear.

While contribution cascade initializations illustrate the intention to work individually or to share improvements, the number of contributions (i.e., the actual cascade size) after initialization helps quantify the long-term commitment to respectively individual or collective projects. The distributions of contribution cascades to *individual* and *shared* projects are respectively log-normal and power law (not represented). On average contribution cascades to individual projects are longer ($\mu_i = 29.60$) compared to contributions to shared projects ($\mu_s = 16.33$). However, the former are less skewed ($\gamma_i = 8.02$) compared to the latter ($\gamma_s = 31.54$) with maximum number of contributions being respectively $1,348$ and $5,415$. In other words, contribution cascades to shared projects in the community of astrophysicists are much more extreme. This reflects the difficulty to organize collective work (compared to own work), but when it is established, collective work is likely to be considerably more productive.

4. CONCLUSION

Co-located meetings are special moments of importance for community enhancement and long-term online production. Here, I have investigated how five hackathons have fostered online collaboration in a community of astrophysicists. I find that the four most recent hackathons have triggered productive bursts of contribution both to shared and individual projects. Their relative contributions to collective work cascades initializations are however contrasted. Finally, long-term online production is embodied by the size of contribution cascades: contribution cascades to shared projects in the community of astrophysicists are much more extreme, reflecting the difficulty to organize collective work, and at the same time, its amazing productivity when collective work cascades are unlocked.

REFERENCES

Luís MA Bettencourt, José Lobo, Dirk Helbing, Christian Kühnert, and Geoffrey B West. 2007. Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the national academy of sciences* 104, 17 (2007), 7301–7306.

Tanzeem Choudhury and Alex Pentland. 2002. The sociometer: A wearable device for understanding human networks. In *CSCW'02 Workshop: Ad hoc Communications and Collaboration in Ubiquitous Computing Environments*.

Riley Crane and Didier Sornette. 2008. Robust dynamic classes revealed by measuring the response function of a social system. *Proceedings of the National Academy of Sciences* 105, 41 (14 Oct. 2008), 15649–15653. DOI:http://dx.doi.org/10.1073/pnas.0803685105

Laura Dabbish, Colleen Stuart, Jason Tsay, and Jim Herbsleb. 2012. Social coding in GitHub: transparency and collaboration in an open software repository. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*. ACM, 1277–1286.

Nicolas Ducheneaut. 2005. Socialization in an open source software community: A socio-technical analysis. *Computer Supported Cooperative Work (CSCW)* 14, 4 (2005), 323–368.

Siobhán O'Mahony and Fabrizio Ferraro. 2007. The emergence of governance in an open source community. *Academy of Management Journal* 50, 5 (2007), 1079–1106.

Jukka-Pekka Onnela, Benjamin N Waber, Alex Pentland, Sebastian Schnorf, and David Lazer. 2014. Using sociometers to quantify social interaction patterns. *Scientific Reports* 4 (2014).

Wanda J Orlikowski and JoAnne Yates. 2002. It's about time: Temporal structuring in organizations. *Organization science* 13, 6 (2002), 684–700.

Alex Pentland, Tanzeem Choudhury, Nathan Eagle, and Push Singh. 2005. Human dynamics: computation for organizations. *Pattern Recognition Letters* 26, 4 (2005), 503–511.

Fernando Perez. 2013. An ambitious experiment in Data Science takes off: a biased, Open Source view from Berkeley. http://blog.fperez.org/2013/11/an-ambitious-experiment-in-data-science.html. (2013). Accessed: May 21, 2015.

Markus Schläpfer, Luís MA Bettencourt, Sébastian Grauwin, Mathias Raschke, Rob Claxton, Zbigniew Smoreda, Geoffrey B West, and Carlo Ratti. 2014. The scaling of human interactions with city size. *Journal of the Royal Society Interface* 11, 98 (2014), 20130789.

Didier Sornette, Thomas Maillart, and Giacomo Ghezzi. 2014. How Much Is the Whole Really More than the Sum of Its Parts? 1Ł 1= 2.5: Superlinear Productivity in Collective Group Actions. *PloS one* 9, 8 (2014), e103023.

Erik H Trainer, Chalalai Chaihirunkarn, Arun Kalyanasundaram, and James D Herbsleb. 2014. Community Code Engagements: Summer of Code & Hackathons for Community Building in Scientific Software. In *Proceedings of the 18th International Conference on Supporting Group Work*. ACM, 111–121.

Georg Von Krogh, Sebastian Spaeth, and Karim R Lakhani. 2003. Community, joining, and specialization in open source software innovation: a case study. *Research Policy* 32, 7 (2003), 1217–1241.