Everything You Always Wanted to Know About Cohorts (But Were Afraid to Ask)

Introducing the User Metrics API
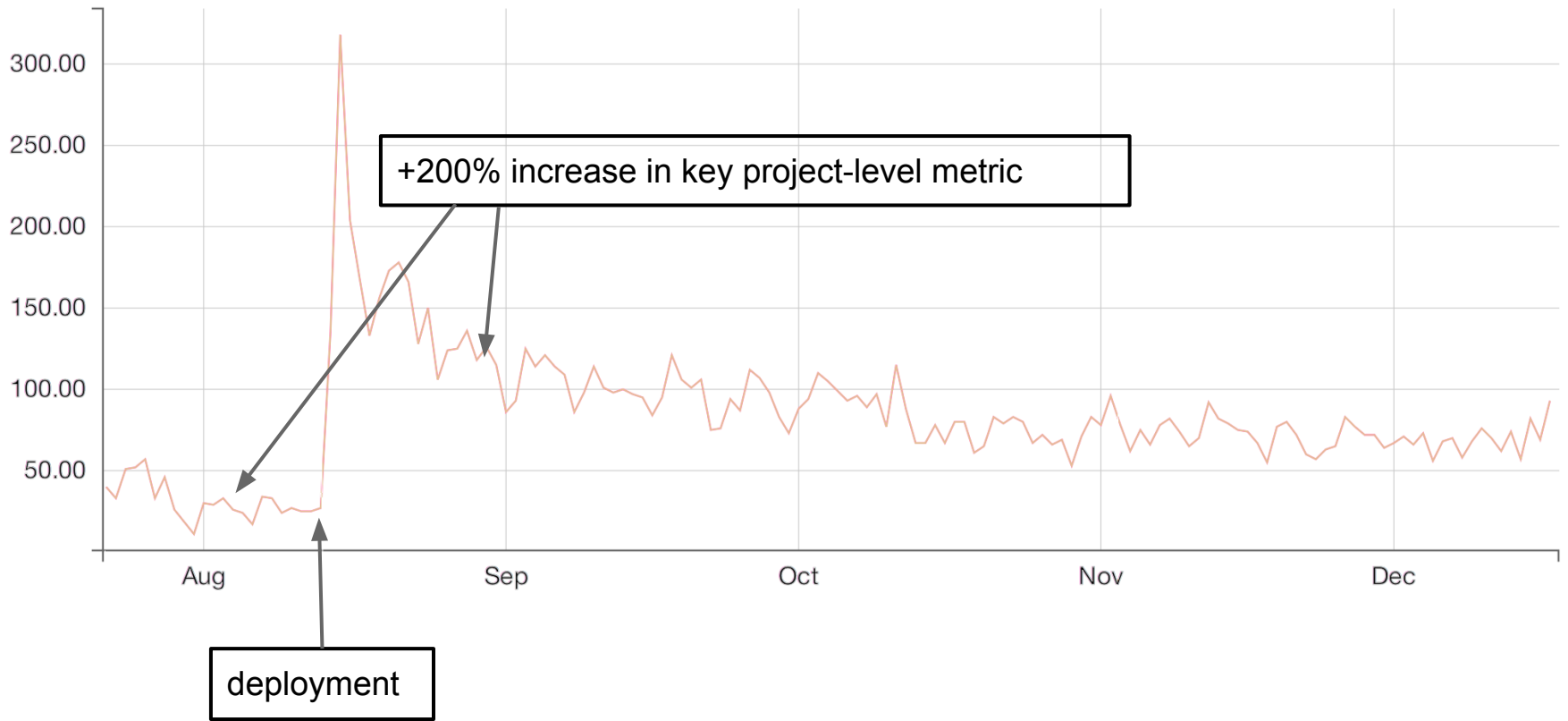
Ryan Faulkner   Dario Taraborelli
rfaulkner@wikimedia.org   dario@wikimedia.org

Monthly Metrics Meeting, March 7 2013

Problem

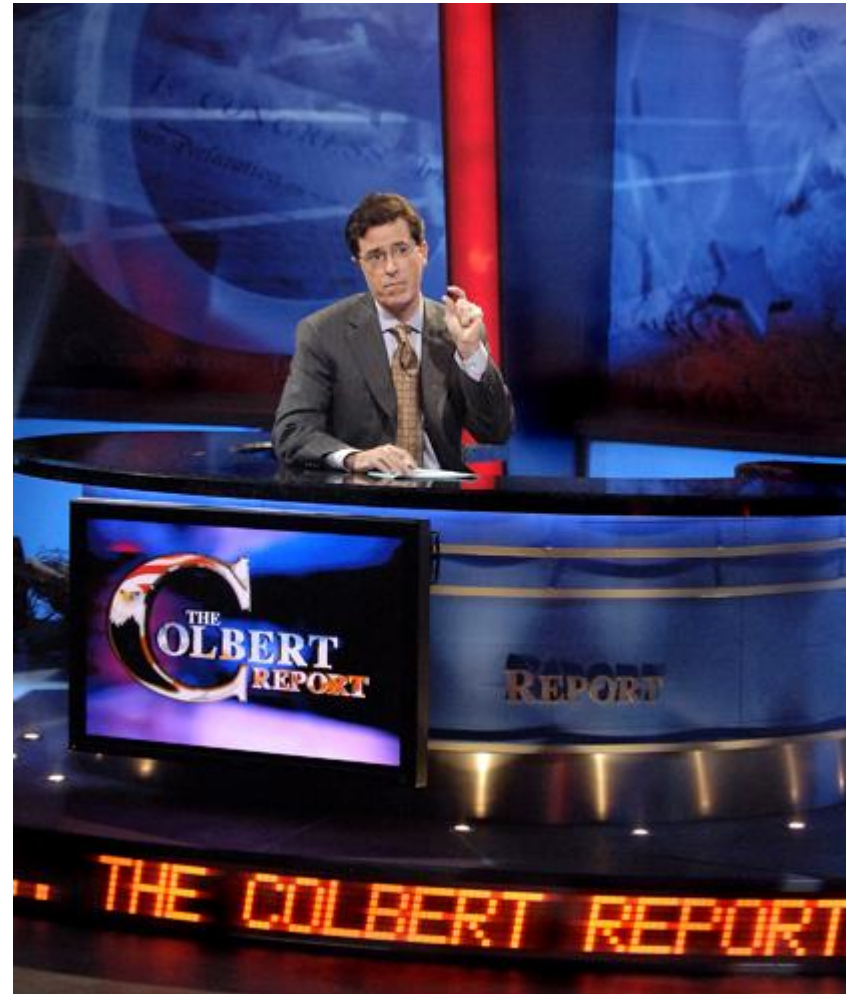How do we know if any {program, experiment, new feature} we run to engage new editors *is actually working*?

did we really cause this?

what if an army of bots did

what if some external event did

Cohort analysis

Measure the productivity of a cohort: a group of users *sharing the same characteristic* or *undergoing the same treatment*

What if we could

1. define a **cohort** from a list of editors

Examples

- users in an E3 experimental group
- students enrolled in a Global Education course
- members of a Wikiproject
- new editors with 10+ edits in their first month
- new users registered on mobile devices
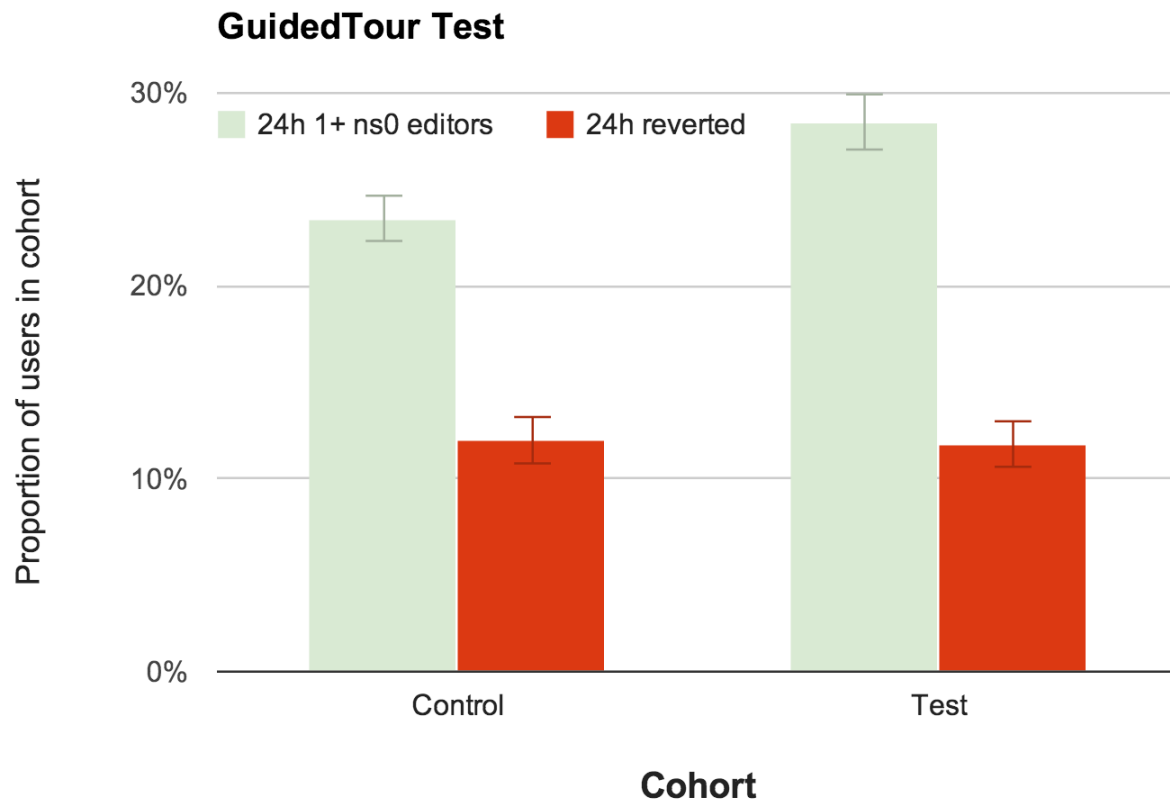- VisualEditor adopters

What if we could

2. measure the {quality, productivity, retention} of these users via a **standard set of metrics**

> **threshold**
>
> measures whether a user reached a given level of activity (a number n of edits) within a specific time t since registration, in a given namespace.

What if we could

3. **compare** this cohort to a (meaningful) baseline.

Introducing the User Metrics API

Introducing the User Metrics API

• **Standardized set of metrics**, backed by extensive research and crash-tested with multiple experiments run by the E3 team.

• Supports **cohorts from any Wikimedia project**, beyond enwiki (including referral campaigns).

• Optimized **performance** for fast extraction of computationally demanding metrics from thousands of users.

• Returns both **individual and aggregate metrics** thanks to multiple aggregator methods. Supports **time series**.

• Easily **extensible** and fully parametrized (metrics can be fine-tuned, default parameters overridden as needed)

Get in touch

Ryan Faulkner

rfaulkner@wikimedia.org

Dario Taraborelli

dario@wikimedia.org

Project home

https://github.com/wikimedia/user_metrics