

Estimating the Causal Effects of Twitter's Interventions on Engagement with Trump's Tweets

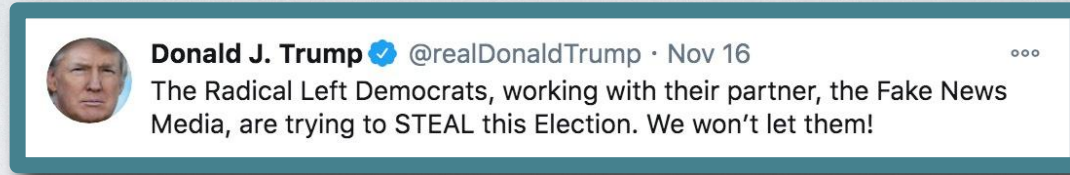
Swapneel Mehta, Jim Bisbee

Richard Bonneau, Joshua Tucker, Jonathan Nagler



Center for
Data Science

How to Limit Misleading Info.?



Viral Tweet with Misinformation

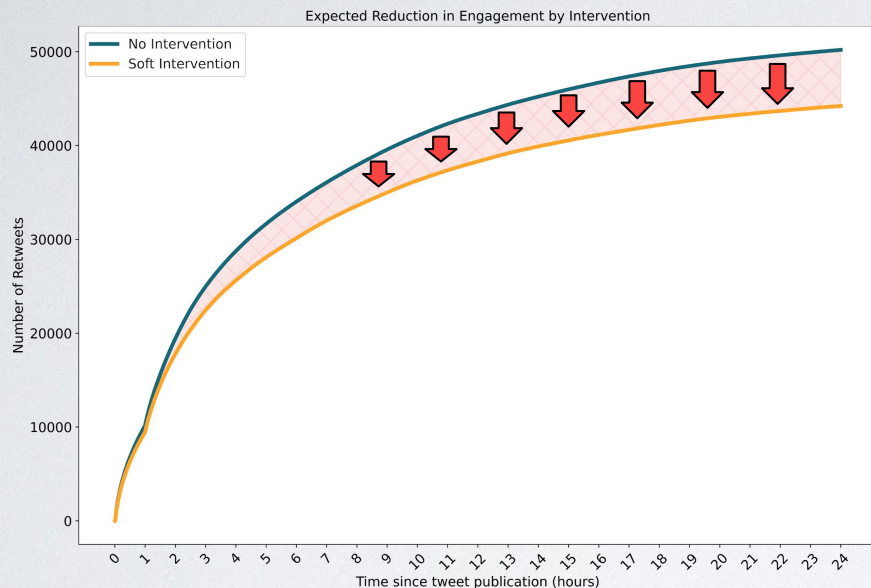


Hard Intervention

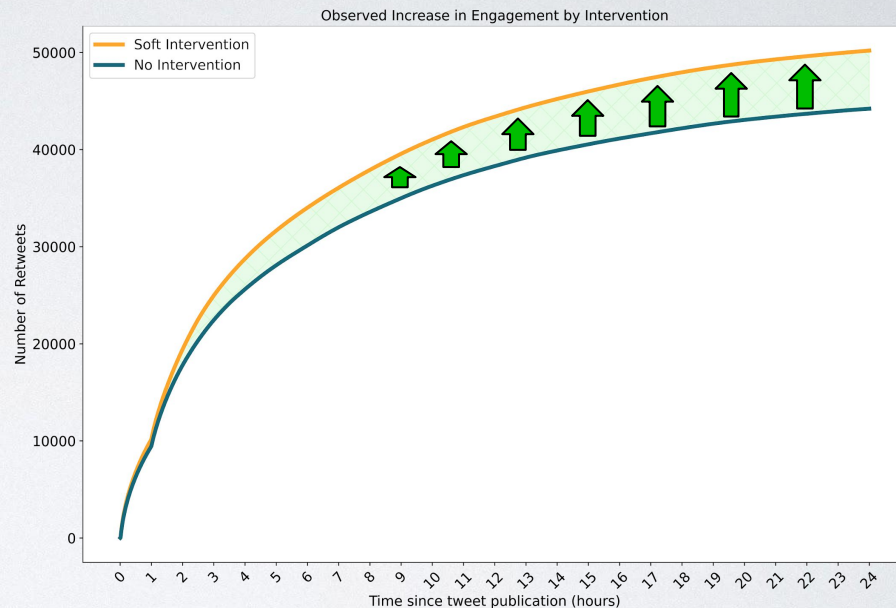


Soft Intervention

What is the Impact of Interventions?

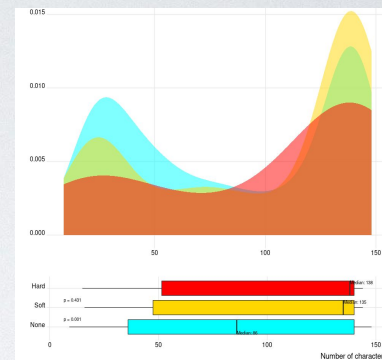
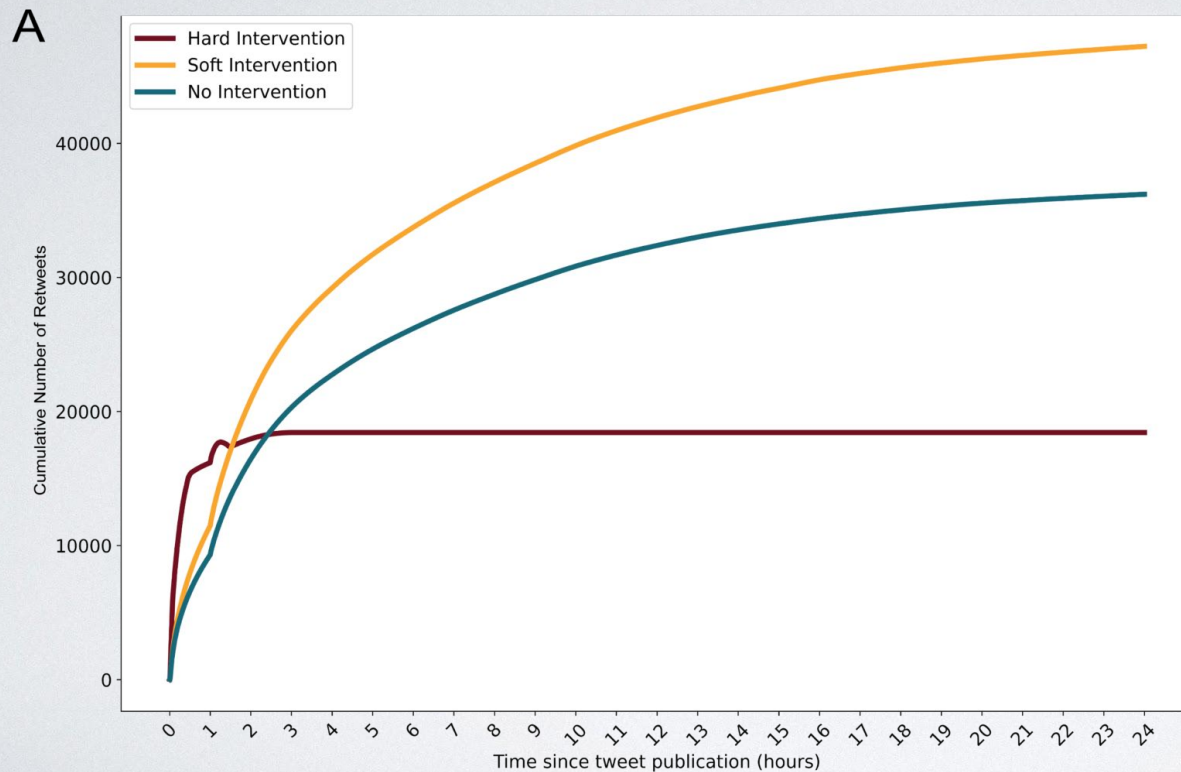


Expectation: **Reduced Engagement**

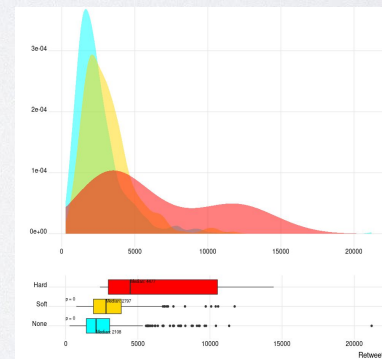


Reality: **Increased Engagement**

Comparing Group-wise Engagement

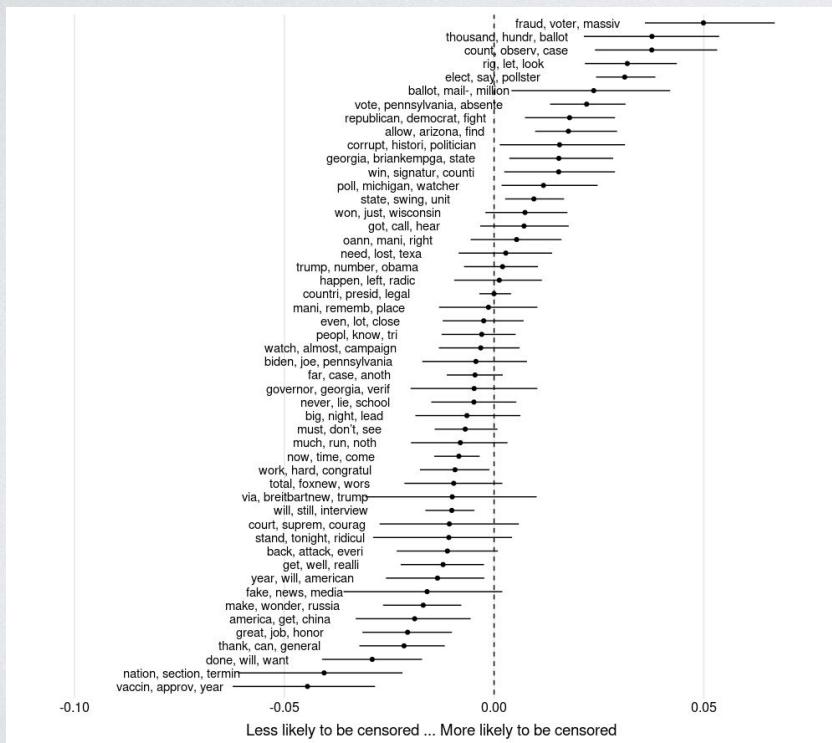


Retweet Distribution <10min



Favorite Distribution <10min

Comparing Group-wise Covariates



- **Intervened-upon tweets are different!**

- More favorited in <10 minutes
- More retweeted in ≤ 10 minutes
- More verbose
- Retweeted by less popular accounts
- More likely to mention “controversial” topics
- Toxicity scores* likely higher

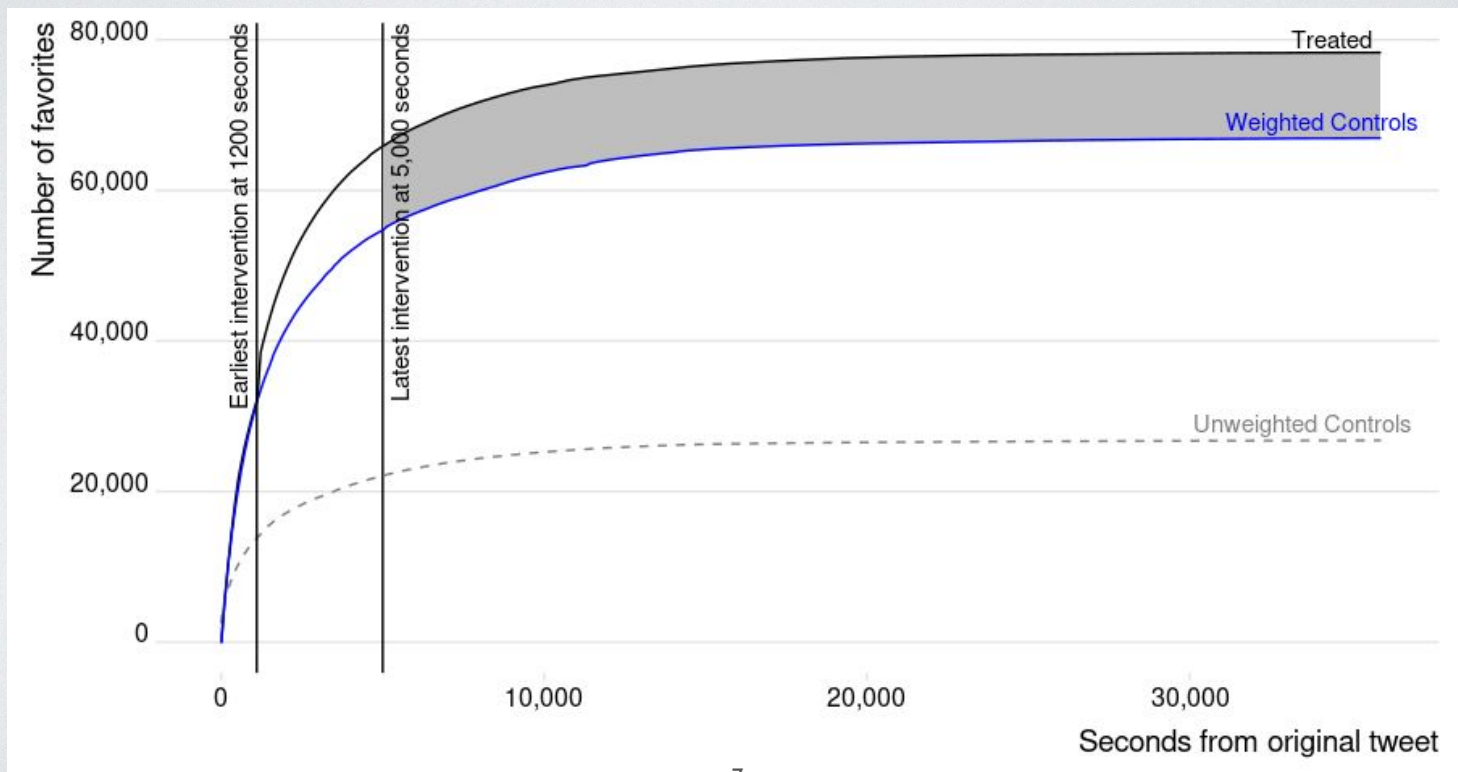
HOW TO GET TO CAUSAL ID

- Exploit the time-series richness of the data
- Match the entire “trajectory” of the tweets pre-intervention
- Account for confounding by balancing on the outcome variable – really powerful!

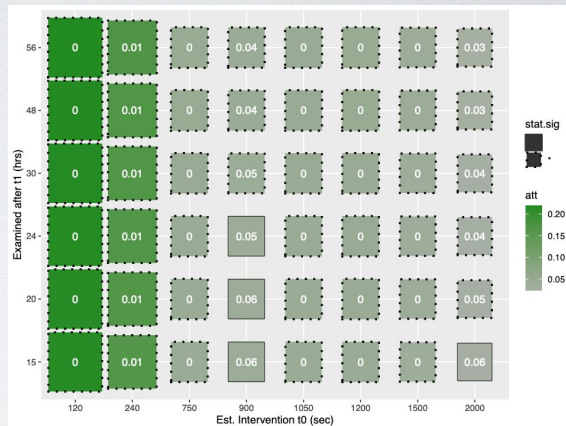
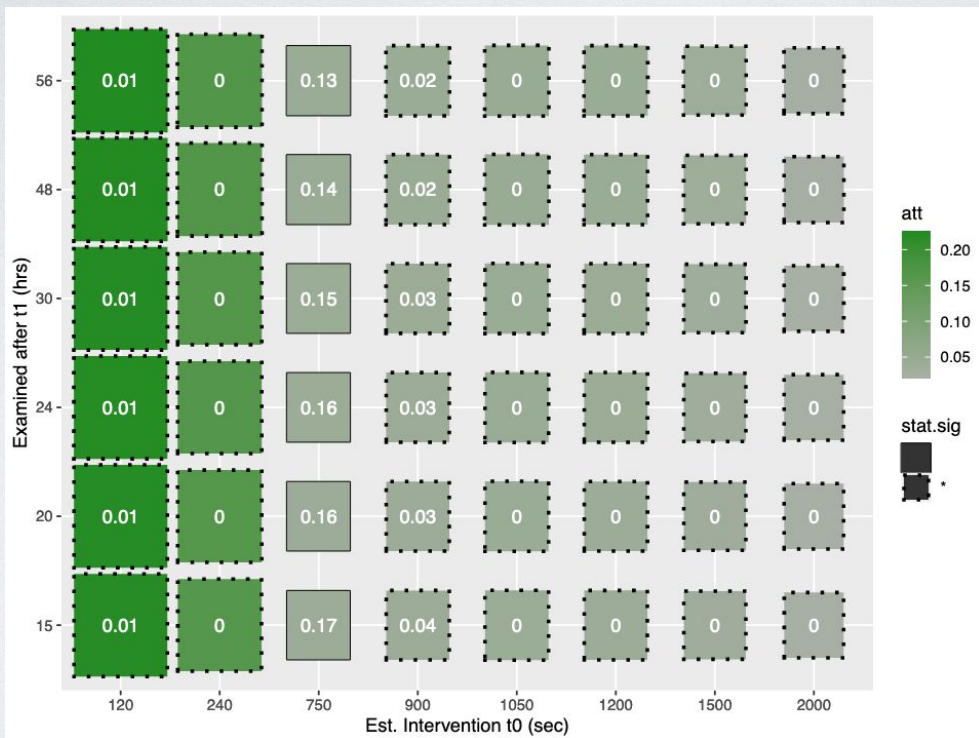
Trajectory Balancing – Hazlett and Xu, 2018

- No balance
- **Mean balancing:** match period-wise means of pre-treatment covariates
- **Kernel balancing:** match higher-order moments of pre-treatment covariates

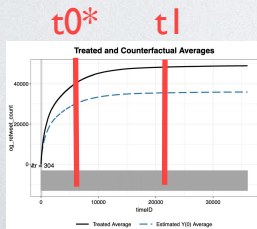
TRAJECTORY BALANCING



Causal Effects of Soft Interventions

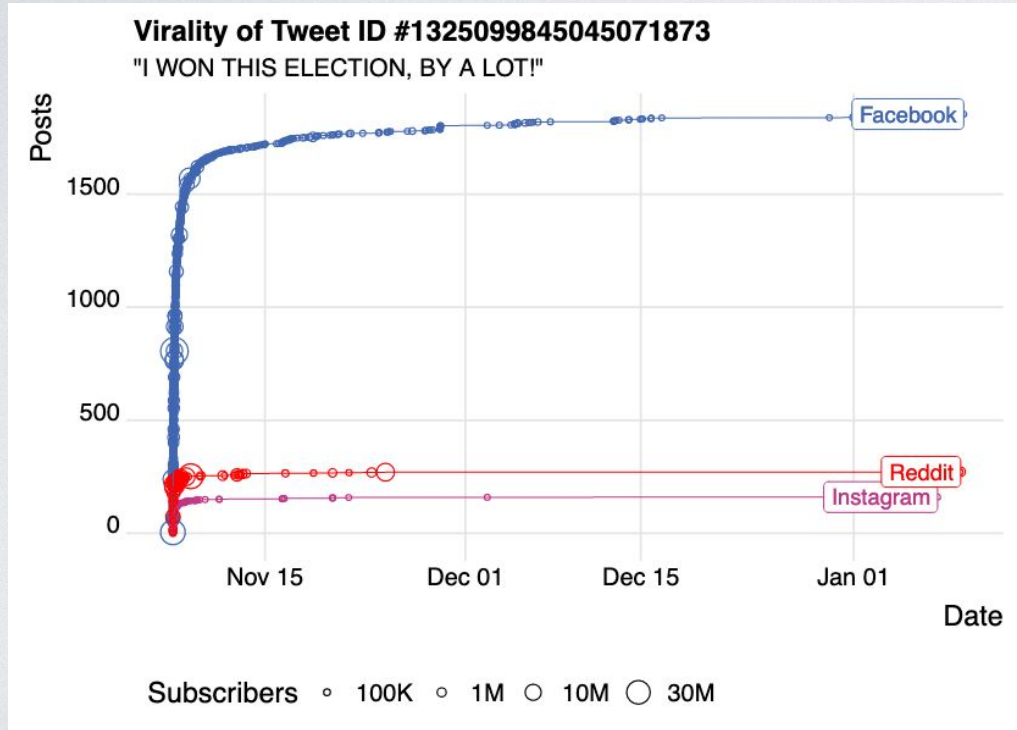


Soft Interventions cause a milder increase in Favorites!



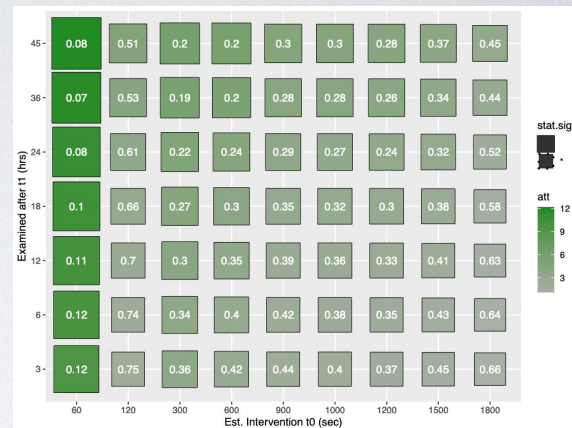
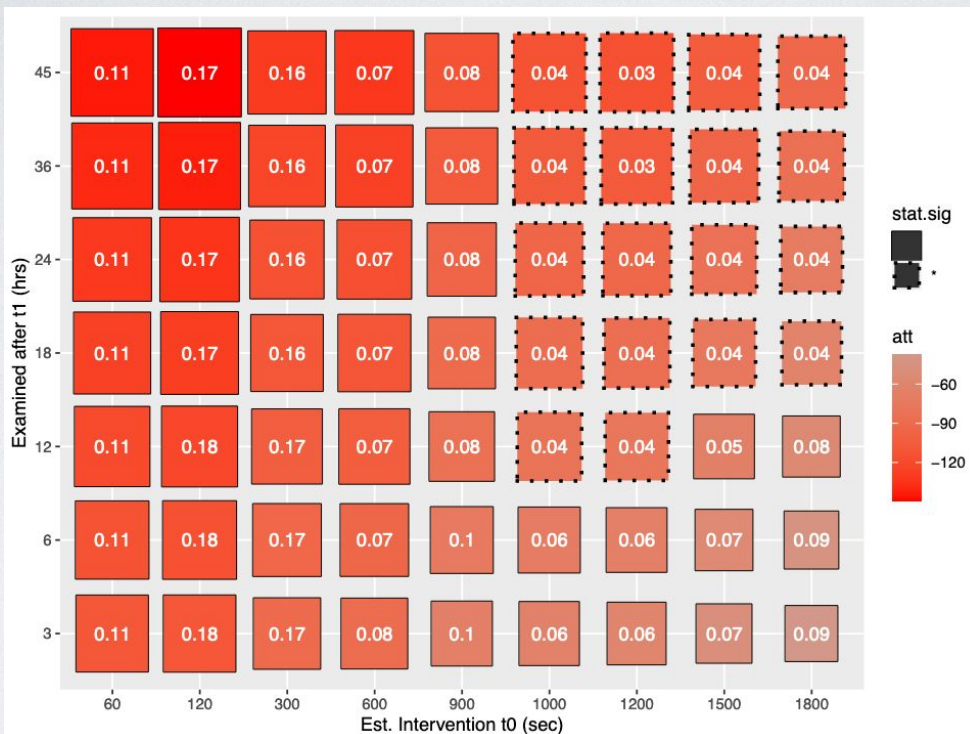
Soft Interventions cause a milder increase in Retweets!

Effects on Reddit, FB, Instagram?



- Tweets are shared beyond just platform of origin, Twitter
- What are the effects of interventions across different platforms?
- **Use CrowdTangle to collect cross-platform data re: tweets**
- **Can track “hard” interventions!**
- **Twitter’s Interventions are effective on FB, Instagram!**

Causal Effects of Hard Interventions



Hard Interventions cause an increase in Reddit posts (with mostly negative sentiment)!

Hard Interventions cause a decrease in FB posts!

