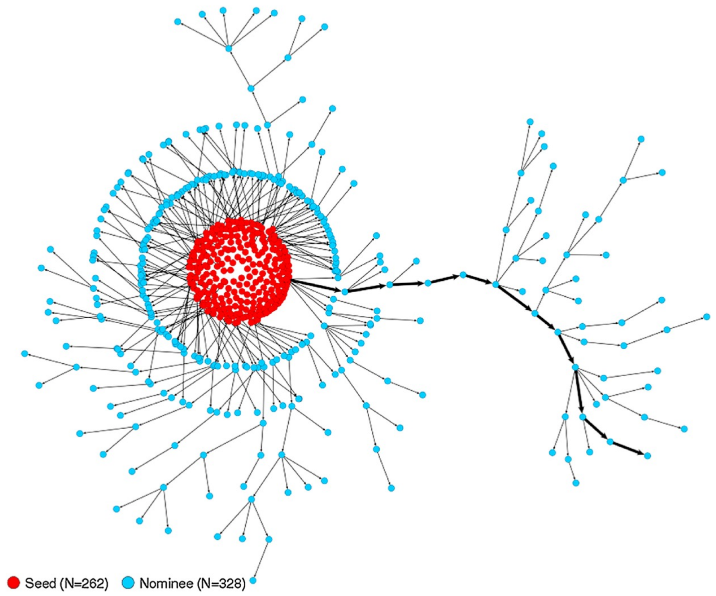
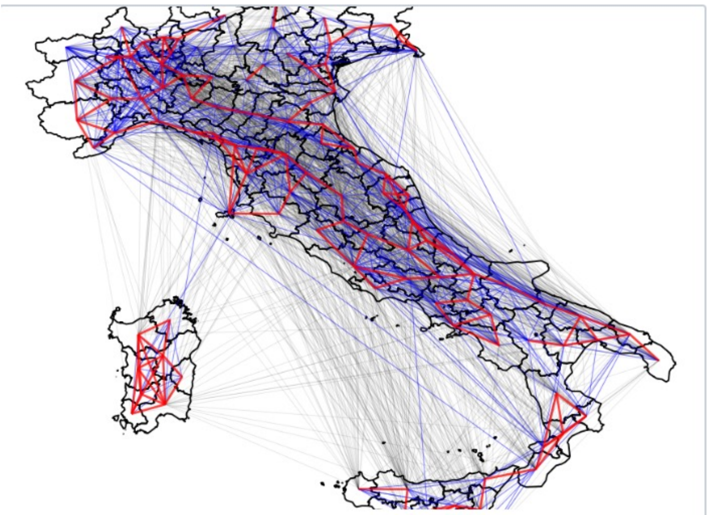
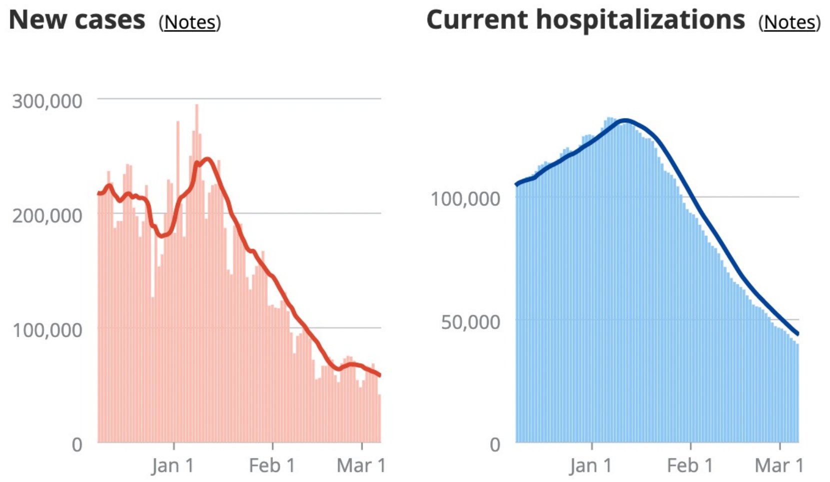
Emerging Data on Epidemics and Networks

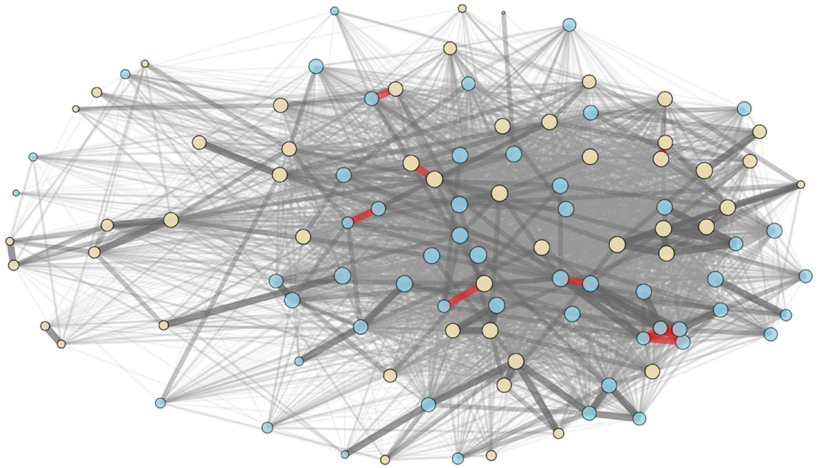
Fan Bu, UCLA
Reading group, Oct 3







eX-FLU: epidemic data + individual-level contact tracing



Very rich data set:

- Medium-scale epi study on university campus
- ~600 participants
- Weekly flu survey + viral test to confirm
- 5min-interval Bluetooth mobile contact tracing on 103 participants

But still missing:

- Exact infection & recovery times
- Possible external links & infections

A. E. Aiello, A. M. Simanek, M. C. Eisenberg, A. R. Walsh, B. Davis, E. Volz, C. Cheng, J. J. Rainey, A. Uzicanin, H. Gao, et al.
Design and methods of a social network isolation study for reducing respiratory infection transmission: The eX-FLU cluster randomized trial.
Epidemics, 15:38–55, 2016

Facebook GPS-based mobility data

https://dataforgood.facebook.com/dfg/tools

- Using GPS tracking of FB users to track locations & movements
- 5min intervals, down to 600m-by-600m tiles
- Mobility & travel pattern changes during COVID-19 (due to lockdown etc.)
- Inferred "co-location" of users as proxy of contacts
- (NOT at individual level, and provides estimate of contact rates/density)
- UK co-location case study example at https://cmmid.github.io/colocation_dashboard_cmmid/

Twitter API V2: information diffusion on social networks

- Can query live or historical Twitter threads for specific topics/hashtags/key words/users
- Lots of research on misinformation, hate speech, popularization, etc.
- Online tutorial available: https://github.com/twitterdev/getting-started-with-the-twitter-api-v2-for-academic-research

Example study on popularization: can exposure to users with opposite views help?

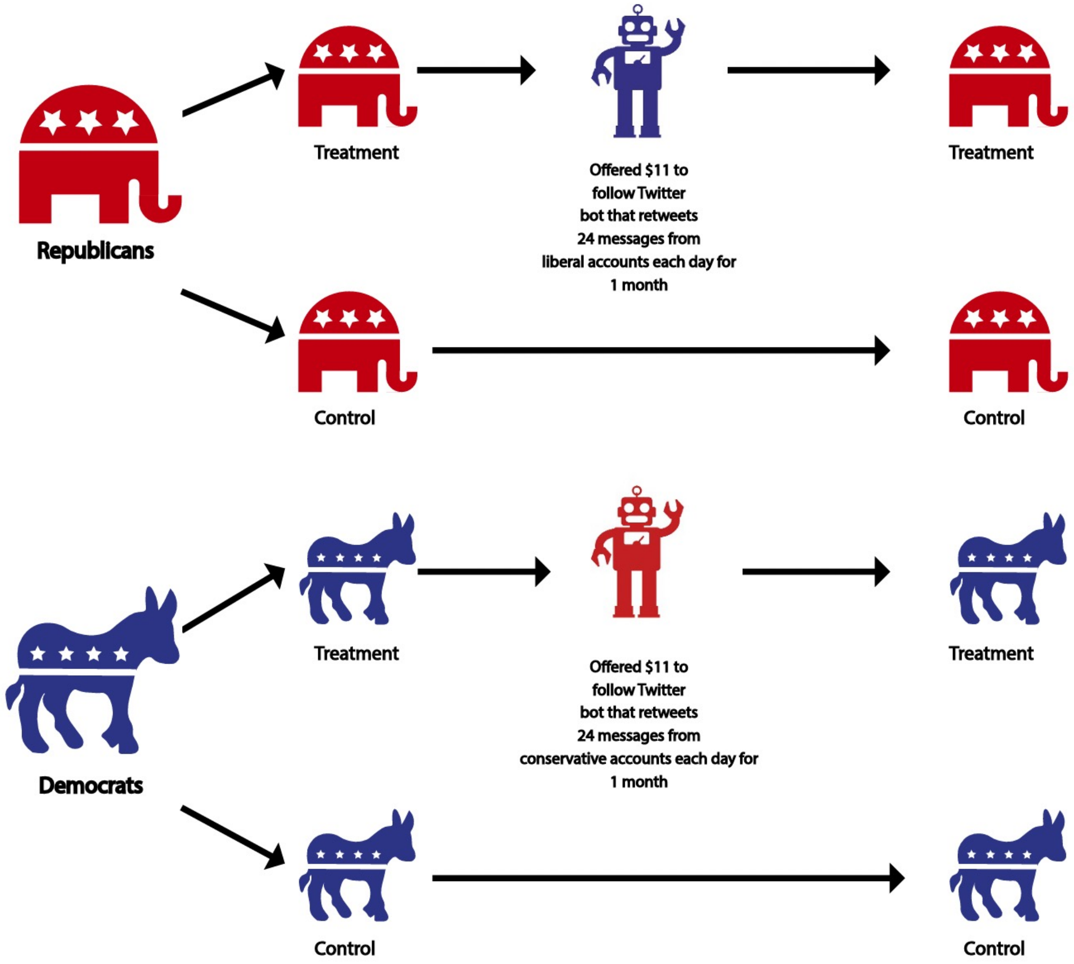


Fig. 1. Overview of research design.

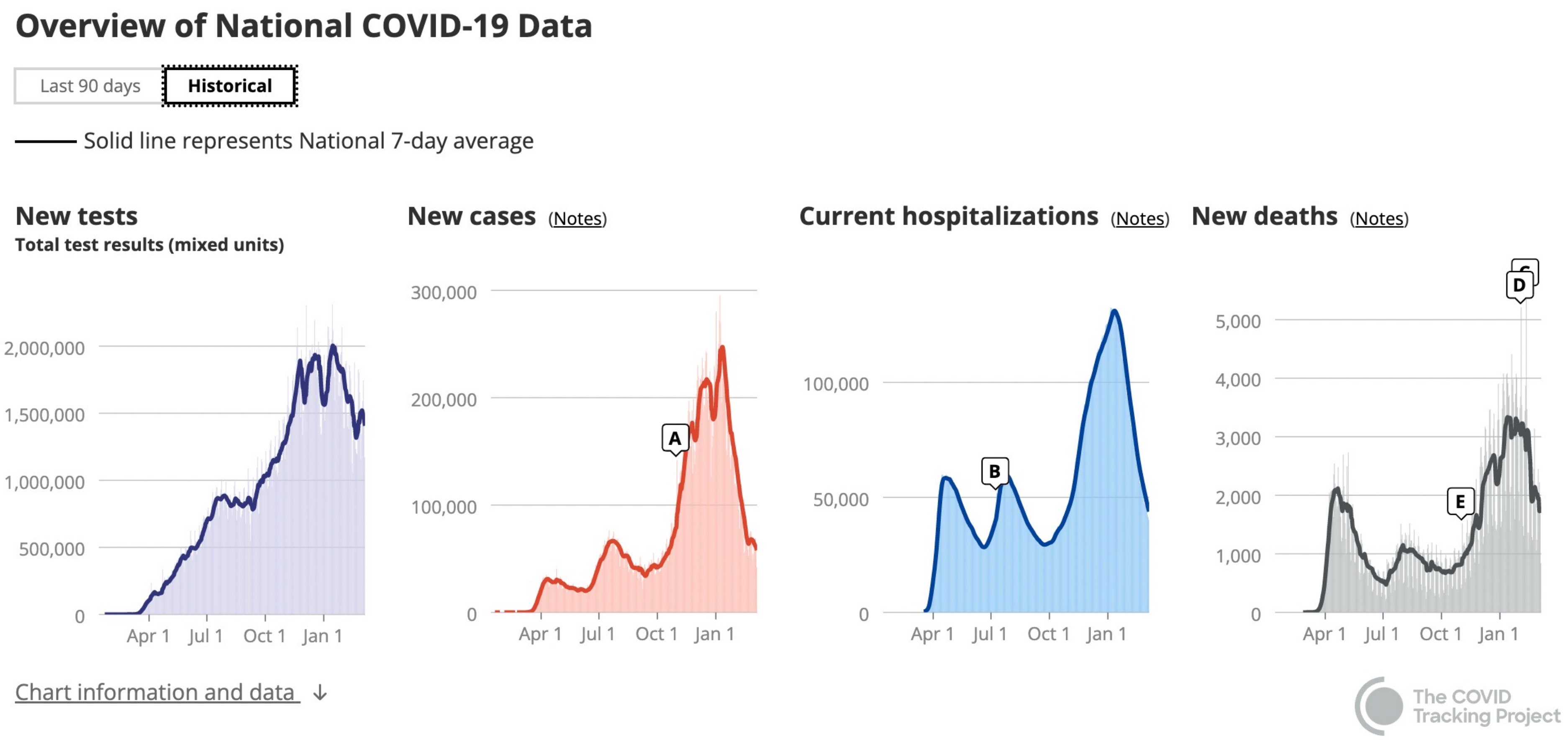
Researchers mined participants' Twitter behaviors and pushed opposite-view tweets to their feeds.

Bail et al. **Exposure to opposing views on social media can increase political polarization.** PNAS (2018).

Epidemic data without network information

- US COVID Tracking Project
  - Multi-level aggregated counts data at national, state, county level
- Israel COVID-19 data
  - Public data only have aggregated case counts
  - Contact tracing data (still not at individual level) are proprietary

Example: US COVID Tracking Project



Summary

- “Network” data are hard to collect, even harder to scale
  - Wearable devices are recent, data collection is expensive
  - Also, data privacy concern
- Epidemic data are much more accessible, but often aggregated
  - Hard to jointly model with networks without “who infected who” information

Data-driven analysis challenges

- There is always data missingness
  - Unknown infection & recovery times
  - Disease latency
  - Unobserved contact links
- Epidemic data are multi-scale and multi-level
- Model validation/goodness-of-fit