Title: Methodology details appendix for “Chinese state media Facebook ads are linked to changes in news coverage of China worldwide”
Authors: Arjun Tambe (1), Toni Friedman (2)
Date: January 14th, 2022
Note: The material contained herein is supplementary to the article named in the title and published in the Harvard Kennedy School (HKS) Misinformation Review.

Appendix C: Methodology details

Treatment variation plot

Figure 1 shows a treatment variation plot (Imai et al., 2018). Each row corresponds to a country, and 2-letter ISO codes are used to label each country. Each column represents one day. Red marks indicate a country was treated on that day (impressions above the “high” threshold), and blue marks indicate a country was untreated on that day.

Figure 1. Treatment variation plot.

Covariate data

Data sources for our covariates are described in the Methods section of the main paper. All covariates were filtered to include only bilateral data between China and each other country in the dataset. For the Events dataset, we apply a "one-a-day" filter (Schrodt, 2015) to include only one event of a given type (e.g., one CAMEO code) between each pair of countries on a given day, and considered only the Goldstein intensity score of the event.

To ensure that the data set was not missing relevant events, human coders determined whether each ad in our dataset mentioned a bilateral event or not. Some advertisements referred to multiple events (e.g., aid provision to several countries), while some ads all referred to a singular event. An ad must have referred to a discrete event, rather than an ongoing process to be included. For example, an ad highlighting an ongoing construction project does not count as an event, whereas an ad highlighting the completion of a project does. We found the corresponding ICEWS event for each event, if any corresponding ICEWS event exists. In total, we found 100 events from the advertisements. We found that 93 of the 100 events were in the ICEWS database. Five of the remaining seven advertisements concern multilateral events with many countries (e.g., praise for China from WHO), which were unlikely to confound our results because units in the matched set were also exposed to the same event.

Further, even when an event was missing, there were typically "redundant" events leading up to each relevant event. For instance, "Provide aid" events were typically preceded by "Intent to provide aid" events. In the ICEWS dataset, 97% of all aid provision events, had another positive event on or within ten days before that event in the same country. This ensured the matched set was still similar to the treated observation with respect to real-world bilateral events, even when events were missing from our dataset.

Procedure for refining matched sets

To be considered in the estimation of the average treatment effect, a treated unit must have been untreated (i.e., impressions below the "high" threshold) on all 10 days preceding the day it became treated. Cases where treatment status flips back and forth in the preceding ten days were excluded because it made matching treatment histories difficult in our case, where an "untreated" status included a wide range of impressions. Finally, instances in which treatment status changes within the first ten or last seven days of the study period were excluded because lagged covariate data or outcome data would have been missing for these units.

We included the day of treatment within the lags; that is, if t denotes the day on which a unit was treated, then treated and control observations were matched on the eleven days of observations of covariates from $t-10$ up to and including t. This ensured the matched set accounted for real-world events occurring simultaneously with the treatment. This is important because there were a handful of advertisements delivered on the same day as a real-world event, which would otherwise not have been included in the set of observations used to refine the matched set.

The formula for the average Mahalanobis distance for a "control" unit i' in the matched set, given a treated observation of unit i at time t, is given in Equation 1, and follows the PanelMatch framework (Imai et al., 2018). Here, $V_{i,t}$ represents a vector of covariate observations for unit i and time t, and Σ_{it} is the sample covariance matrix of $V_{i,t}$.

$$S_{i,t}(i') = \frac{1}{L}\sum_{l=0}^{10}\sqrt{(V_{i,t-l} - V_{i\prime,t-l})^{\top}\Sigma_{i,t-l}^{-1}(V_{i,t-l} - V_{i\prime,t-l})}$$

The five units with the lowest Mahalanobis scores were weighted equally. The “untreated” outcome in the difference-in-differences estimator is the weighted sum of the outcome variable at the relevant time periods, as in Imai et al. (2018). Treated observations that lacked any eligible matching “control” observations were not used to estimate the treatment effect, but there were no instances of this in our dataset.