

U.S. States

– Supplementary Information

Jiachen Sun, Peter Gloor

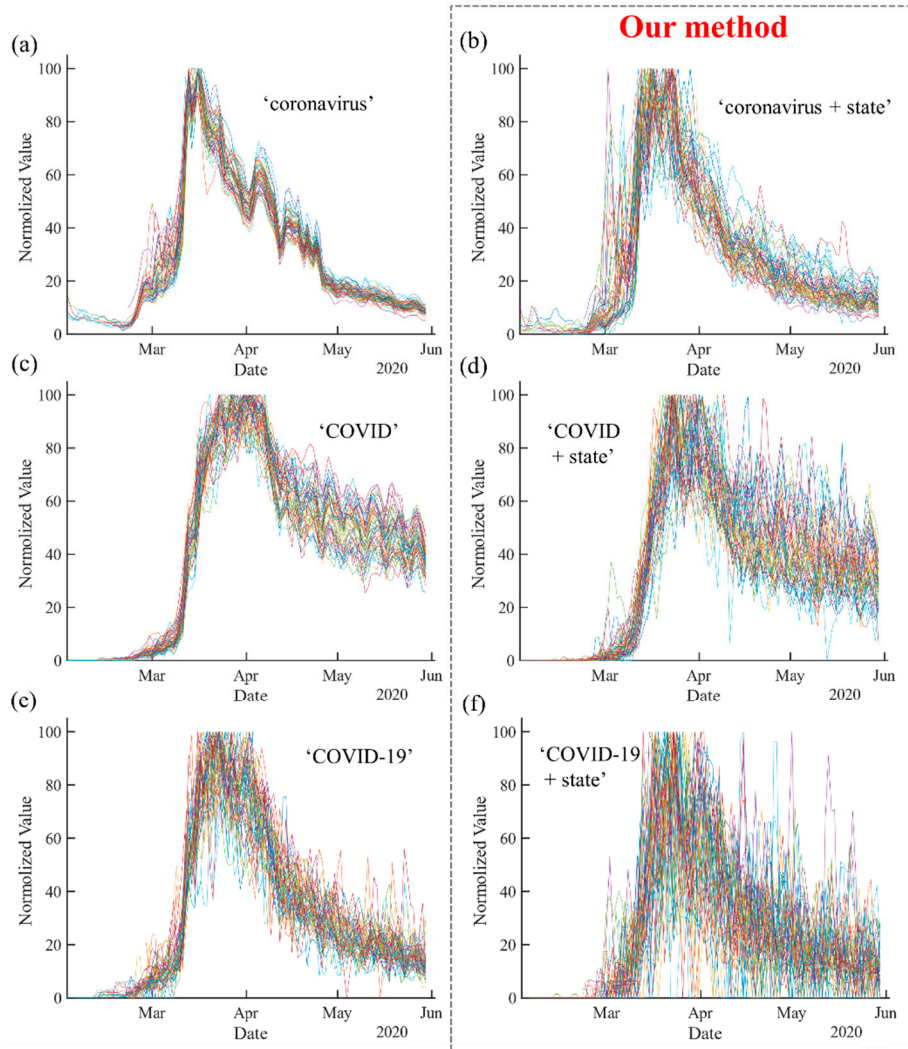


Figure S1. Comparison of Google Trends index with and without the state's full name in 50 U.S. states. Each curve corresponds to an individual state, with the values smoothed by a 3-day-average. (a-b) Using keyword 'coronavirus'. (c-d) Using keyword 'COVID'. (e-f) Using keyword 'COVID-19'.

Table S1. Spearman correlation coefficient between c^* and l^* ($N = 50$) from Twitter data and the Google search data. The significance level is denoted by stars in red: *** $p < 0.01$.

Twitter	Google Trend (coronavirus)	Google Trend (COVID)	Google Trend (COVID-19)
0.564 ***	-0.247	0.063	0.162

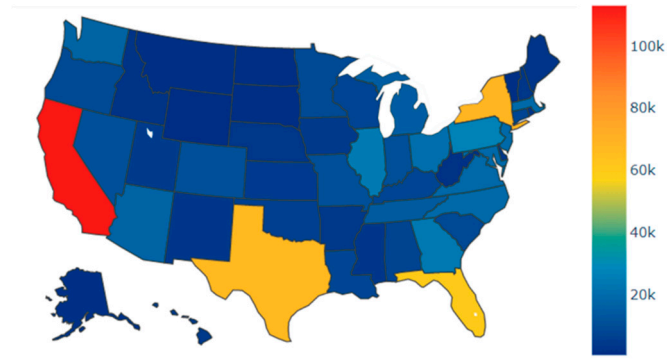


Figure S2. The number of COVID-19 related tweets in 50 U.S. states available in the used dataset from January 27, 2020 to May 30, 2020.

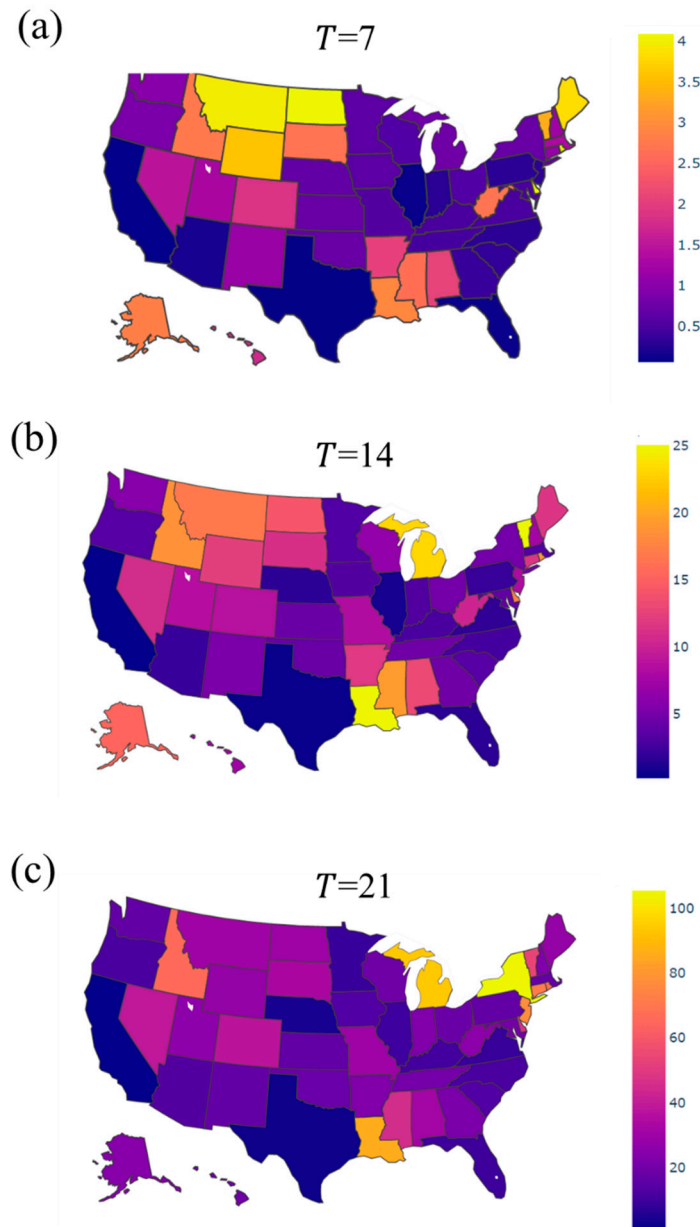


Figure S3. The state-level *early* infection rate calculated as the proportion of residents (per 100,000) being infected in the earliest T days.

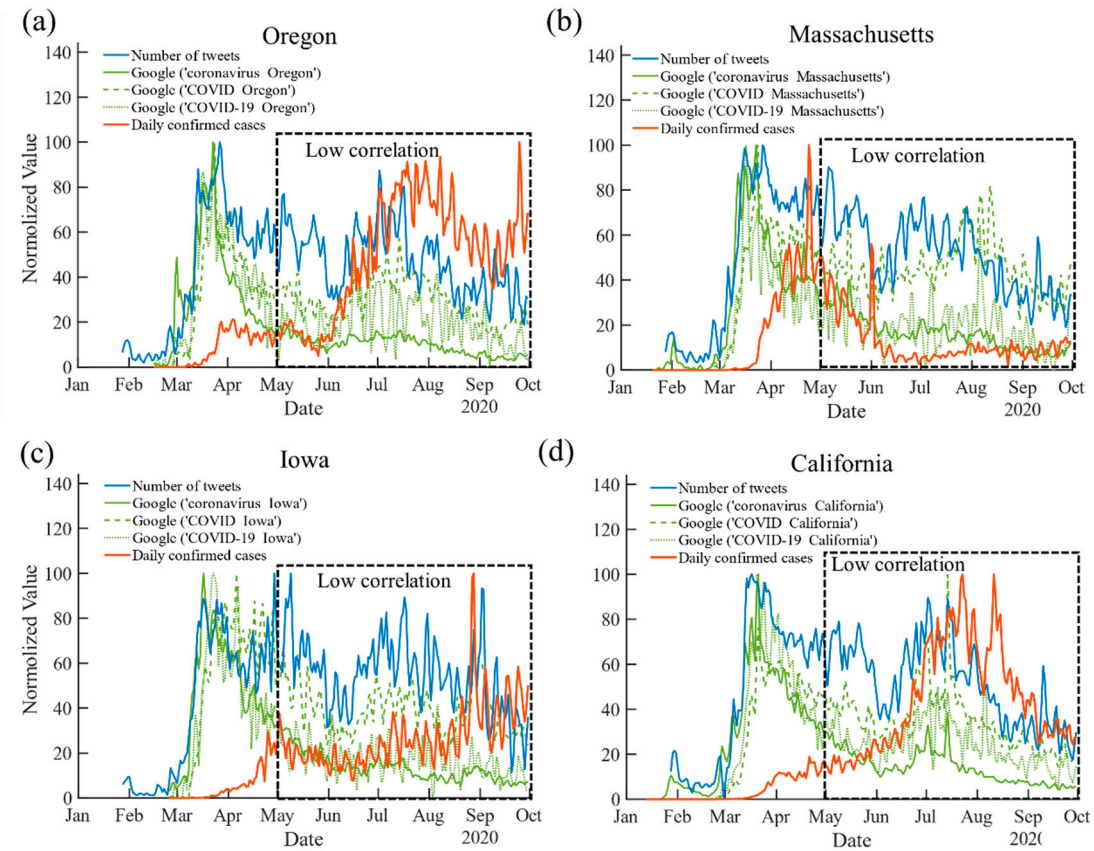


Figure S4. Number of COVID-19 related tweets, Google Trends index using different COVID-19 keywords (integrated with the state's full name) and daily infection numbers till September 30, 2020. The values of each curve are first smoothed by a 3-day-average to reduce the noise and are then normalized to $[0, 100]$ for comparison.

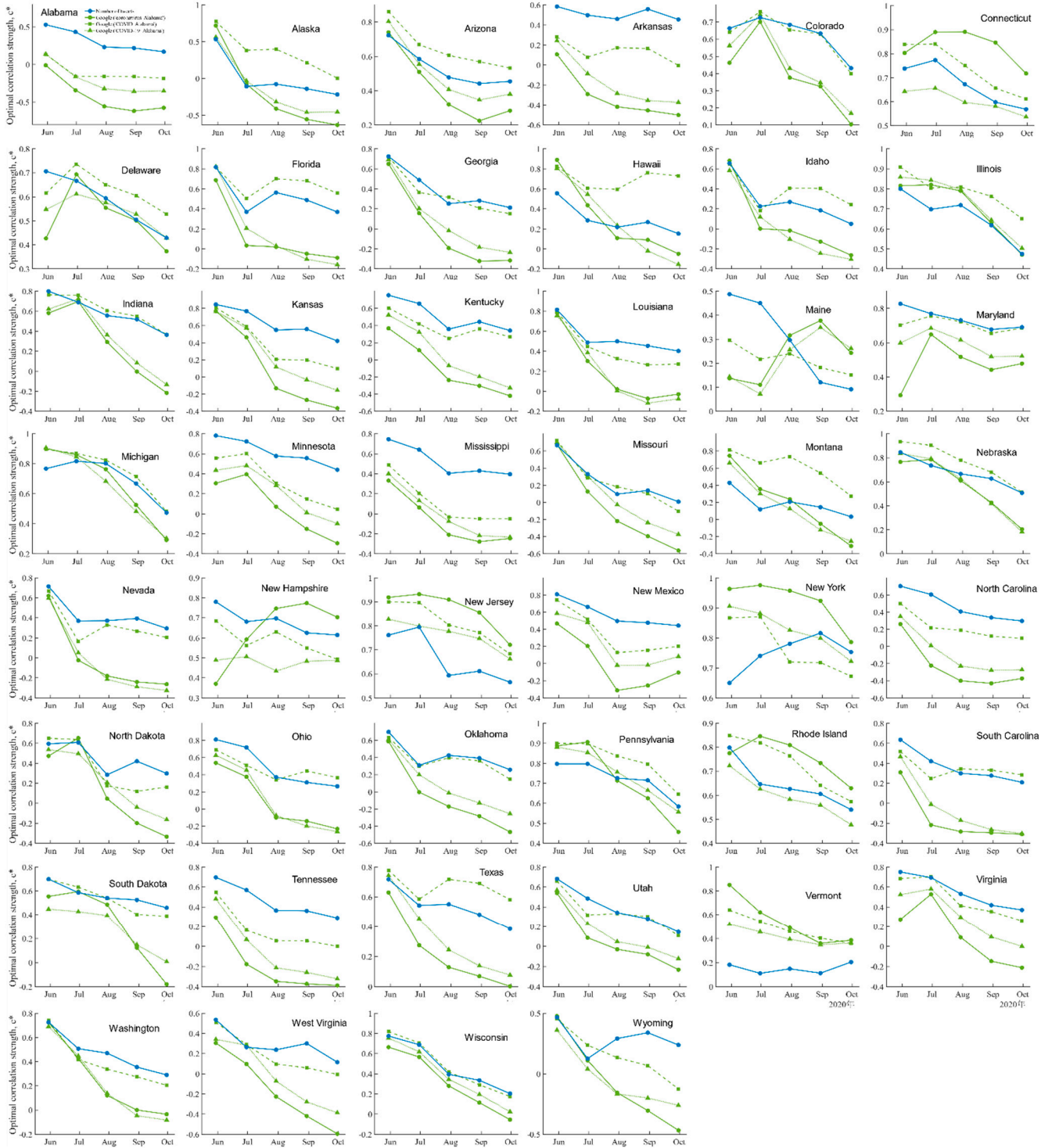


Figure S5. The highest correlation strength c^* between the COVID-19 daily infection and the Internet data vs. the studied period in 46 remaining states. The x-axis denotes the end date of the studied period for a given state, while the start date is chosen as the date when the first case was reported in this state. Each point of c^* is obtained by applying the lagged Spearman correlation analysis

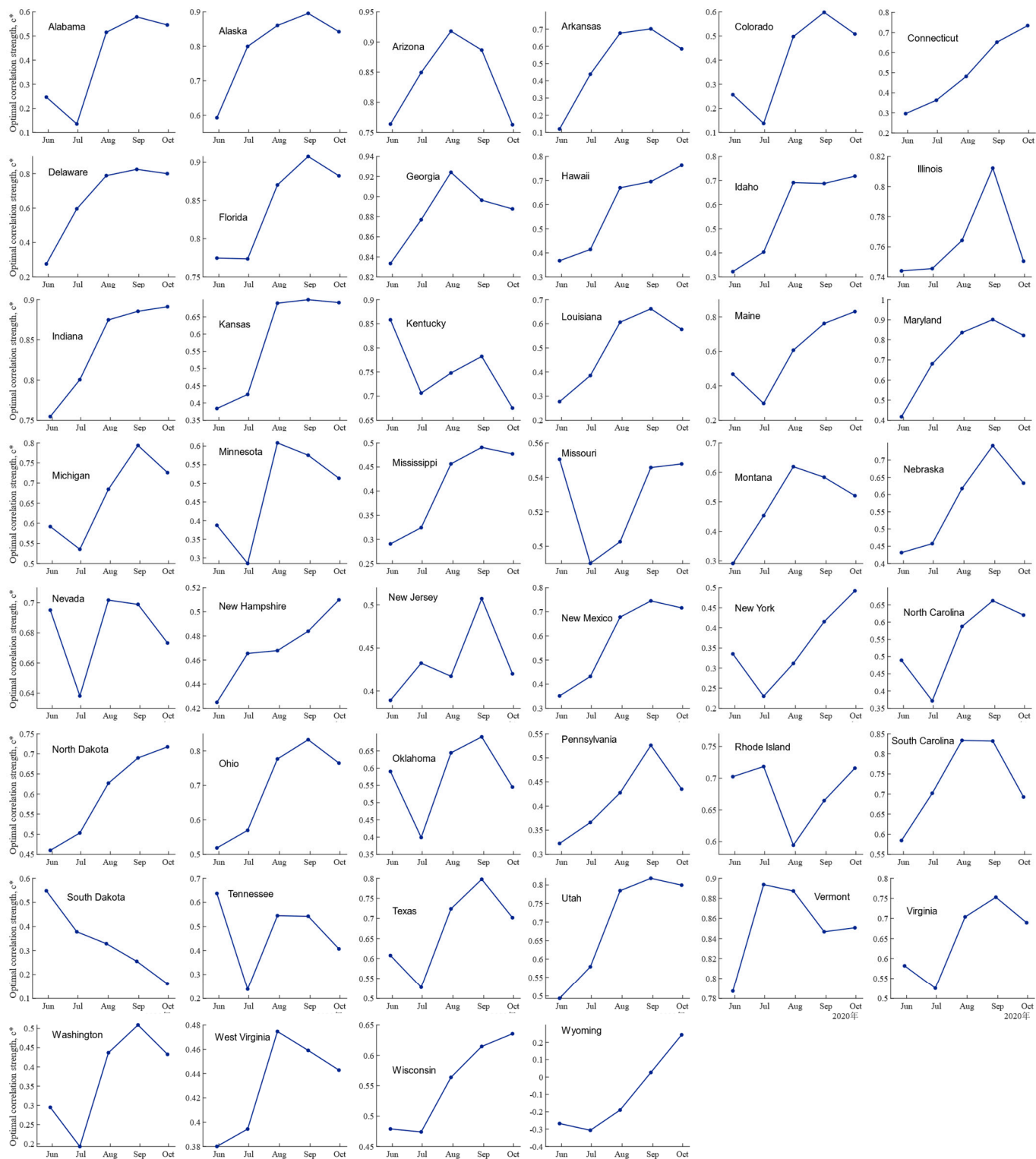


Figure S6. Value of the highest correlation strength c^* vs. the studied period for the COVID-19 daily testing number and the Google index using ‘COVID testing + state’s name’ in 46 states.