



# Quantifying the role of the COVID-19 pandemic in the 2020 U.S. presidential elections

Pietro De Lellis<sup>1</sup>, Manuel Ruiz Marín<sup>2</sup> , and Maurizio Porfiri<sup>3,a</sup> 

<sup>1</sup> Department of Electrical Engineering and Information Technology, University of Naples Federico II, 80125 Naples, Italy

<sup>2</sup> Department of Quantitative Methods, Law and Modern Languages, Technical University of Cartagena, 30201 Cartagena, Murcia, Spain

<sup>3</sup> Department of Mechanical and Aerospace Engineering, Department of Biomedical Engineering, Center for Urban Science and Progress, New York University Tandon School of Engineering, Brooklyn, NY 11201, USA

Received 1 August 2021 / Accepted 4 October 2021

© The Author(s), under exclusive licence to EDP Sciences, Springer-Verlag GmbH Germany, part of Springer Nature 2021

**Abstract** In the media, a prevalent narrative is that the incumbent United States President Donald J. Trump lost the 2020 elections because of the way he handled the COVID-19 pandemic. Quantitative evidence to support this narrative is, however, limited. We put forward a spatial, information-theoretic approach to critically examine the link between voting behavior and COVID-19 incidence in the 2020 presidential elections. The approach overcomes classical limitations of traditional regression analysis, where it does not require an underlying mathematical model and it can capture nonlinear interactions. From the analysis of county-level data, we uncovered a robust association between voting behavior and prevalence of COVID-19 cases. Surprisingly, such an association points in the opposite direction from the accepted narrative: in counties that experienced less COVID-19 cases, the incumbent President lost more ground to his opponent, now President Joseph R. Biden Jr. A tenable explanation of this observation is the different attitude of liberal and conservative voters toward the pandemic, which led to more COVID-19 spreading in counties with a larger share of republican voters.

## 1 Introduction

The 2020 presidential elections in the United States (U.S.) have exposed, once again, a divided country with polarized opinions [1]. The response to the COVID-19 pandemic was a major topic in the debate between the two candidates (the incumbent President Donald J. Trump and the former Vice President Joseph R. Biden Jr.), who expressed radically different narratives on the epidemiological impact of the virus and policy agendas to halt the spread.

Trump was heavily criticized by the scientific community for his alleged role in spreading misinformation regarding the origin and the severity of the virus and his ambiguous stand on social distancing and mask-wearing [2]. With more than nine million people being infected and two hundred thousand losing their lives in cities and rural areas of the U.S. by November 2020 [3], authoritative voices in the press have identified the difference in the candidates' response to the pandemic as a critical factor that led to Trump's loss and Biden's victory [4–7]; for example, TIME wrote about Trump

that “his prospects for re-election were dragged down by... his reckless approach to a virus that landed him in the hospital at the peak of the campaign.”

Baccini et al. [8] have recently presented an insightful study on the relationship between the COVID-19 pandemic and the U.S. presidential elections, employing ordinary and two-stage least-squares models for the county-level effect of COVID-19 incidence (cumulative number of cases or deaths) on voting behavior (Trump's differential vote share from 2016 to 2020). Alongside these variables, the authors controlled for several county-level demographic and socioeconomic factors such as population, share of foreign-born population, share of the population with a college degree, share of non-Hispanic Black population, and social mobility index.

Based on findings from their analysis, Baccini et al. [8] proposed that COVID-19 cases were negatively associated with Trump's differential vote share from 2016 to 2020, that is, counties that experienced more COVID-19 cases reduced their preference for the incumbent President. Within a counterfactual analysis, it could be concluded that “Trump would have kept the presidency with 21% fewer cases.” The authors also determined that the election outcome had a weaker association with the death toll, whereby most of the analy-

Pietro De Lellis and Manuel Ruiz Marín contributed equally.

<sup>a</sup> e-mail: [mporfiri@nyu.edu](mailto:mporfiri@nyu.edu) (corresponding author)

sis failed to reach statistical significance when utilizing deaths rather than cases as the independent variable.

The results of Baccini et al. [8] have preceded other efforts which investigated the link between COVID-19 pandemic and the U.S. presidential elections. For example, Noland and Zhang [9] performed an equivalent regression analysis that confirmed the claims by Baccini et al. [8]. Specifically, they determined that Trump would have won the electoral vote and lost the popular vote, like in 2016, if the pandemic had not occurred or if it were mitigated by 30 percent. Interestingly, not all the evidences support the thesis that the handling of the crisis hurt Trump's re-election. The ordinary least-squares analysis by Lake and Nie [10] did not discover a positive association between Trump's defeat and COVID-19 incidence. Although failing to reach statistical significance, their correlations suggest the presence of a positive effect of COVID-19 on Trump's vote share.

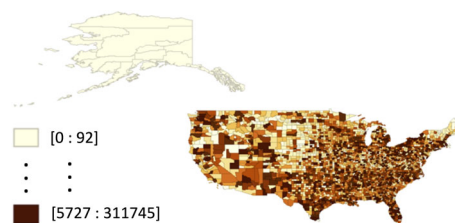
Here, we present a further analysis on the link between the COVID-19 pandemic and the U.S. presidential election within spatial information theory [11]. Spatial information theory offers a versatile basis for the study of spatial processes from sparse observational datasets, without the need of stringent model assumptions for hypothesis-testing. Inspired by our prior work on the inference of spatial associations using entropy-based measures [12–14], we investigated the link between the presidential election and the COVID-19 pandemic. In this vein, our work could offer insight into the extent to which the conclusions drawn by Baccini et al. [8] rely on the linearity and spatial independence of their calibrated models, albeit it should not be considered as a replication study, due to differences in the adopted dataset.

The main commodity of the analysis is the Shannon entropy of the spatial processes, which quantifies uncertainty from statistical distributions. By selectively computing the conditional mutual information between a target process (such as the voting behavior) and other source processes (such as COVID-19 incidence), it is possible to infer spatial links in the dataset. We systematically pursued such an approach to quantify the mechanisms underlying both the participation in the 2020 presidential election and vote difference between the candidates that led to the victory of Biden against Trump. We considered epidemiological and economic processes that might have influenced voting behavior, along with spatial interactions that encapsulate the social and political fabric of the country.

## 2 Data and methods

### 2.1 Data collection

County-level data on the COVID-19 pandemic were obtained from usafacts.org [15]. For each county, we recorded the total numbers of confirmed cases until November 3rd 2020 (Fig. 1).



**Fig. 1** Color map of the number of COVID-19 cases in each U.S. county until November 3rd 2020. Each color corresponds to one of ten bins with the same count, with darker colors identifying counties that experienced more cases or deaths

Unemployment data were collected from the U.S. Bureau of Labor Statistics [16] as a measure of local economic conditions throughout the country. Specifically, for each county we considered the average unemployment in the year preceding the 2020 U.S. presidential elections (September 2019–October 2020) and the unemployment in 2016 (January–December 2016), year of the previous U.S. presidential elections (Fig. 2).

Data on health insurance coverage were collected from the U.S. Census Bureau [17] to measure the expansion in insurance coverage that followed the Affordable Care Act (ACA) by President Barack H. Obama Jr., which became operational in January 2014. Specifically, for each county, we considered the share of the civilian, non-institutionalized population aged 19–64 that was covered by a health insurance in 2013 and 2018 (estimated from the five-year American Community Survey of 2013 and 2018, respectively; Fig. 3). These shares were selected according to Lake and Nie [10] to capture the effect of the ACA.

Data for the 2020 presidential elections were collected from politico.com [18] in all states, except New York, Virginia, Washington D.C., Minnesota, and South Dakota, whose electoral data were retrieved from nbcnews.com [19] (Fig. 4, top panels). Electoral votes from the 2016 elections were downloaded from dailykos.com [20], except for Connecticut, taken from politico.com [21], and New Mexico and Rhode Island, taken from cnn.com [22] (Fig. 4, bottom panels).

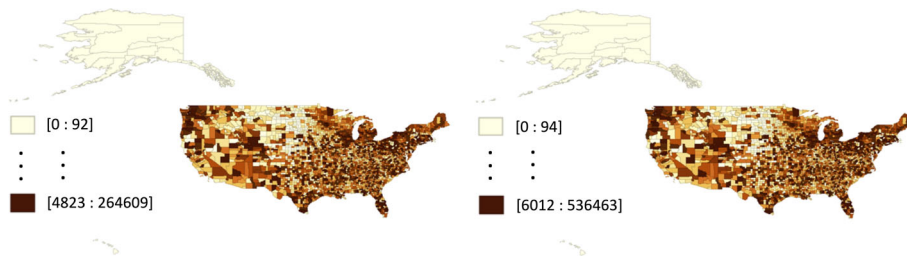
To normalize epidemiological, economic, and voting data, we collected population data in 2016 and 2020 from the U.S. Census Bureau [23] (Fig. 5).

### 2.2 Entropy-based measures

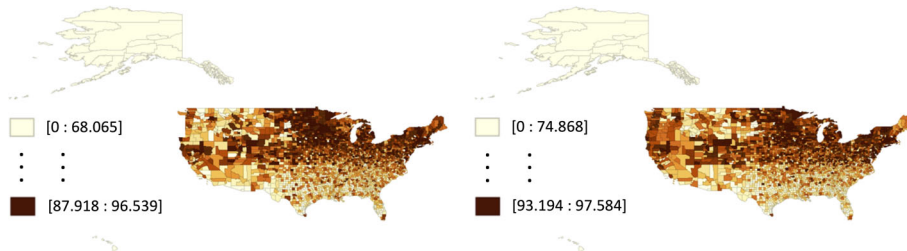
Let us consider a univariate or multivariate random variable  $X$ . The uncertainty encoded by  $X$  can be measured through its Shannon entropy  $H(X)$ , defined as

$$H(X) = - \sum_{x \in \mathcal{X}} P(X = x) \ln P(X = x), \quad (1)$$

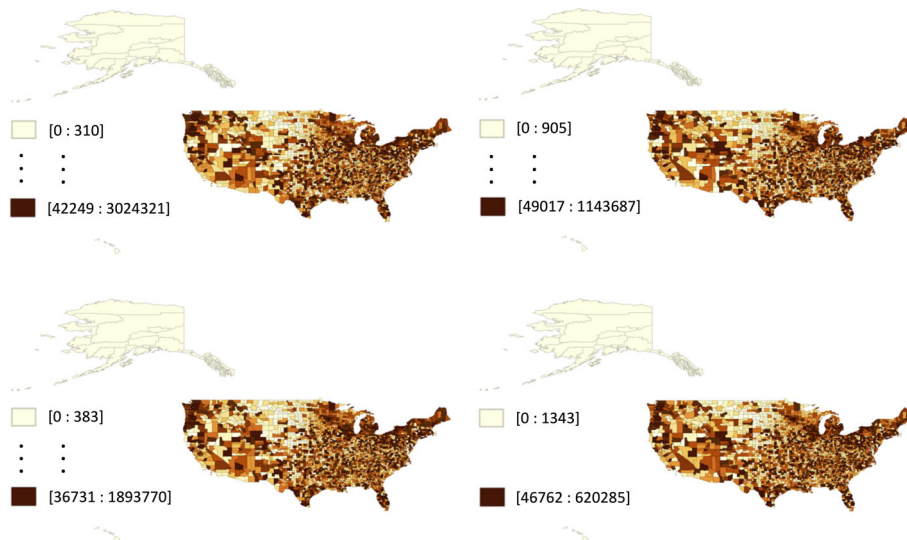
where  $\mathcal{X}$  is the set of all the realizations of  $X$ ,  $P(\cdot)$  indicates probability, and  $\ln(\cdot)$  is the natural logarithm so that entropy is measured in “nats.” For a determin-



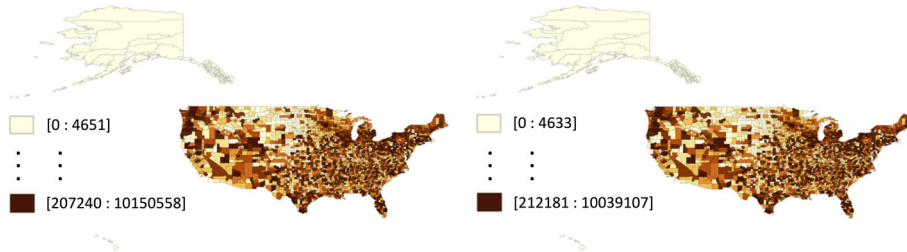
**Fig. 2** Color map of the average number of unemployed individuals in 2016 (left panel) and 2020 (right panel), in each U.S. county. Each color corresponds to one of ten bins with the same count, with darker colors identifying counties with more unemployed



**Fig. 3** Color map of the percent share of insured population in 2013 (left panel) and 2018 (right panel), in each U.S. county. Each color corresponds to one of ten bins with same count, with darker colors identifying counties with larger shares of insured



**Fig. 4** Color map of the number of votes in 2020 (top panels) and 2016 (bottom panels) U.S. presidential elections. The left and right panels report the votes for the democratic and republican candidate, respectively. Each color corresponds to one of ten bins with the same count



**Fig. 5** Color map of the 2016 (left panel) and 2020 (right panel) populations in each U.S. county. Each color corresponds to one of ten bins with same count, with darker colors identifying larger populations

istic variable which always attains the same value, the entropy is equal to zero.

Given a second random variable  $Y$  with realizations in  $\mathcal{Y}$ , conditional entropy is given by

$$\begin{aligned}
 H(X|Y) &= H(X, Y) - H(Y) \\
 &= - \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} P(X = x, Y = y) \ln \frac{P(X = x, Y = y)}{P(Y = y)} \\
 &= - \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} P(X = x, Y = y) \ln P(X = x|Y = y).
 \end{aligned}
 \tag{2}$$

This quantity should be understood as the amount of uncertainty encoded by the target variable  $X$ , given knowledge about all the sources comprising  $Y$ . A related quantity is the mutual information between  $X$  and  $Y$ , defined as

$$\begin{aligned}
 I(X; Y) &= H(X) - H(X|Y) \\
 &= \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} P(X = x, Y = y) \ln \frac{P(X = x, Y = y)}{P(X = x)P(Y = y)}.
 \end{aligned}
 \tag{3}$$

Mutual information is a non-negative quantity that is equal to zero if and only if  $X$  and  $Y$  are independent.

In our study of spatial associations between processes, we consider the variable  $X$  as our target process (that is, the voting behavior throughout the country) and we take  $Y = [Y_1, \dots, Y_n]$  to be a collection of source processes that could potentially explain  $X$  (for example, COVID-19 prevalence and unemployment variation). Under this premise, we are interested in exploring the dependence between  $X$  and any of the components of  $Y$ , given all the other components. Specifically, for any  $i = 1, \dots, n$ , we set ourselves to study the conditional mutual information

$$\delta_i := I(X; Y_i | Y^{(-i)}) = H(X | Y^{(-i)}) - H(X | Y),
 \tag{4}$$

where  $Y^{(-i)} = [Y_1, \dots, Y_{i-1}, Y_{i+1}, \dots, Y_n]$ . By construction, conditional mutual information is symmetric with respect to  $X$  and  $Y_i$ , as one would gather by recalling that  $Y$  is equivalent to  $(Y_i, Y^{(-i)})$  and expanding the conditional entropies into

$$\begin{aligned}
 \delta_i &= H(X, Y^{(-i)}) - H(Y^{(-i)}) \\
 &\quad - H(X, Y_i, Y^{(-i)}) + H(Y_i, Y^{(-i)}).
 \end{aligned}
 \tag{5}$$

Conditional independence of  $X$  and  $Y_i$  implies that

$$\delta_i = 0,
 \tag{6}$$

so that knowledge about  $Y_i$  does not help reduce the uncertainty about  $X$ . A positive value indicates that  $Y_i$  encodes useful information for the prediction of  $X$ , which we refer to as an association between the variables.

### 2.3 Statistical test

From the sample of the joint variable  $(X, Y)$ , we can perform a non-parametric one-sided test for the null " $H_0: \delta_i = 0$ " (that is, lack of an association between  $X$  and  $Y_i$ ). The test can be summarized in the following steps:

1. Compute the sample estimate  $\hat{\delta}_i$  of  $\delta_i$ . In our analysis, we used a simple plug-in estimation by binning the distributions of the target and source variables using  $b$  bins. To alleviate estimation errors due to low counts in some of the bins or disproportionate counts between multiple bins, we opted for equal-bin-count histograms [24]. The number of bins depends on the size of the dataset and the number of source processes: for a dataset like ours with about  $10^3$  samples,  $b$  should not exceed three to consider more than a few source processes.
2. Calculate  $B$  bootstrap realizations  $\hat{\delta}_i^1, \dots, \hat{\delta}_i^B$  of  $\delta_i$  by randomly shuffling the values of  $Y_i$ , while keeping unaltered those of the target and of any other source variables. For example, in the study of the association between COVID-19 prevalence and voting behavior, one should shuffle only COVID-19 prevalence data, without breaking patterns between voting behavior and economic indicators. This shuffling preserves the internal structure of the interaction between the target and all the other source variables than  $Y_i$ , including the spatial structure of underlying associations.
3. Compute a  $p$ -value from the distribution of the bootstrap realizations, corresponding to the quantile of  $\hat{\delta}_i$ :

$$p = \frac{1}{B} \sum_{j=1}^B \mathcal{I}(\hat{\delta}_i^j > \delta_i),
 \tag{7}$$

where  $\mathcal{I}(\cdot)$  is the indicator function which assign 1 to a true statement and 0 otherwise.

4. Reject the hypothesis of lack of association if  $p < \alpha$ , where  $\alpha$  is the chosen level of significance.

### 2.4 Selection of target and source processes

To study both the change in participation and preference of the U.S. population in the presidential election, we tracked the number of votes in favor of the democratic candidates in 2020 (Biden,  $v_{20}^{\text{dem}}$ ) and 2016 (Hillary D. R. Clinton,  $v_{16}^{\text{dem}}$ ), and the number of votes in favor of the republican candidate in 2020 (Trump,  $v_{20}^{\text{rep}}$ ) and 2016 (Trump,  $v_{16}^{\text{rep}}$ ) in each of the 3105 U.S. counties for which data were available.



We aggregated these county-level data on voting behavior into two target processes, considered separately, one by one. First, we studied the normalized percent variation in the total number of votes from 2016 to 2020, irrespective of whether they were cast in favor of Trump or Biden,

$$\Delta_v^{\text{tot}} = 100 \left( \frac{v_{20}^{\text{dem}} + v_{20}^{\text{rep}}}{p_{20}} - \frac{v_{16}^{\text{dem}} + v_{16}^{\text{rep}}}{p_{16}} \right), \quad (8)$$

with  $p_{16}$  and  $p_{20}$  being the U.S. population in 2016 and 2020, respectively. Second, we examined the normalized percent variation in the vote difference between the democratic (Clinton and Biden) and republican (Trump) candidates from 2016 to 2020,

$$\Delta_v^{\text{dif}} = 100 \left( \frac{v_{20}^{\text{dem}} - v_{20}^{\text{rep}}}{p_{20}} - \frac{v_{16}^{\text{dem}} - v_{16}^{\text{rep}}}{p_{16}} \right). \quad (9)$$

With respect to source processes, we considered the percent prevalence of COVID-19 confirmed cases until November 3rd 2020 ( $c$ ). In addition to this epidemiological indicator, we included the percent normalized unemployment variation

$$\Delta_u = \frac{u_{20}}{p_{20}} - \frac{u_{16}}{p_{16}}, \quad (10)$$

with  $u_{16}$  and  $u_{20}$  being the total number of unemployed in the year before the 2020 elections (September 2019–October 2020) and in 2016 (January–December 2016). As a further economic indicator, we considered the variation in the five-year health insurance coverage share

$$\Delta_h = h_{18} - h_{13}, \quad (11)$$

with  $h_{18}$  and  $h_{13}$  being the share of five-year health insurance coverage in 2018 and 2013, respectively. Finally, to acknowledge spatial interactions between neighboring counties that may underlie a common voting behavior, for each of the chosen target process, we included a spatial autoregression [25]. Specifically, when studying the county-level variation in the total number of votes (vote difference between the two candidates), we included the mean of the same quantity in the neighboring counties. Following standard practice in spatial statistics [25], we use the symbol  $W$  to identify spatial averaging on neighboring counties, so that, for example,  $W\Delta_v^{\text{dif}}$  is the vector of the averages of the normalized percent variation in the vote difference.

In all our computations, we used three bins ( $b = 3$ ) and 10,000 bootstrap realizations ( $B = 10,000$ ). Although a larger number of bins may be desirable, this is the largest value that could be reliably considered: increasing the number of bins may compromise the accuracy of the estimations of the probability mass functions. In addition, significance level was always set to  $\alpha = 0.050$ .

For completeness, in the Appendix we report a linear regression analysis utilizing the same variables considered as part of the spatial information-theoretic approach.

### 3 Results

Our information-theoretic analysis indicated that both the variations in the total number of votes from 2016 to 2020 ( $\Delta_v^{\text{tot}}$ ) and in the vote difference between the democratic and republican candidates from 2016 to 2020 ( $\Delta_v^{\text{dif}}$ ) were associated with COVID-19 incidence (Table 1 and 2). We uncovered an association between  $\Delta_v^{\text{tot}}$  and the prevalence of confirmed COVID-19 cases ( $p < 0.001$ ). Likewise, we determined that the variation in the vote difference between the democratic and republican candidates from 2016 to 2020 was associated with COVID-19 cases ( $p = 0.006$ ). The variation in health insurance coverage was associated with the total votes ( $p = 0.026$ ), but not with the variation in the vote difference between the two candidates ( $p = 0.160$ ). The variation in the rate of unemployment was associated with both variations in the total votes from 2016 to 2020 and in the vote difference between the two candidates ( $p < 0.001$ , for both processes).

The analysis also identified the presence of a spatial structure in both the target processes, variation in the total vote and in the vote difference between the republican and democratic candidates from 2016 to 2020 (Table 1 and 2). For both processes, we determined an association between county-level data and data in neighboring counties ( $p < 0.001$ , for both processes).

Conditional mutual information provides important insight into dependencies among spatial processes, but it does not describe the qualitative nature of the interaction among them. To understand whether the determined associations were positive or negative, we inspected the marginal probabilities underlying the computation of conditional mutual information in Eq. (4).

We determined that COVID-19 cases were negatively associated with the variation in the total vote count (Table 3), whereby it was more likely to register a large variation in vote count in counties that were less affected by the pandemic, and a small variation in those that suffered the most from COVID-19. Interestingly, these two effects did not exactly balance each other, whereby the increase in total votes in less affected counties was higher than the decrease in total votes in more affected counties. Variation in the total votes was negatively associated with the variation in the unemployment rate, whereby it was more likely to observe large increases in the electoral participation in counties that experienced more job losses; likewise, less participation was registered in counties that experienced smaller increases in unemployment rate (Table 3). The health insurance share did not seem to bear any effect either on the total vote count, whereby marginal probabilities were all close to chance (Table 3).

**Table 1** Conditional mutual information between each of the chosen source processes (confirmed cases of COVID-19,  $c$ ; unemployment variation from 2016 to 2020,  $\Delta_u$ ; variation in the health insurance coverage from 2013 to 2018,  $\Delta_h$ ; and spatial average of the variation in the total votes from 2016 to 2020,  $W\Delta_v^{\text{tot}}$ ) and the target processes (variation in the total votes from 2016 to 2020,  $\Delta_v^{\text{tot}}$ )

	Source			
	$c$	$\Delta_u$	$\Delta_h$	$W\Delta_v^{\text{tot}}$
Conditional mutual information $\delta_i$	0.0478 (0.0236)	0.0304 (0.0235)	0.0247 (0.0237)	0.0534 (0.0236)
$p$ -value	< <b>0.001</b>	< <b>0.001</b>	<b>0.026</b>	< <b>0.001</b>

In parenthesis, we report the threshold value for significance of conditional mutual information; bold  $p$ -values refer to significant associations

**Table 2** Conditional mutual information between each of the chosen source processes (confirmed cases of COVID-19,  $c$ ; unemployment variation from 2016 to 2020,  $\Delta_u$ ; variation in the health insurance coverage from 2013 to 2018,  $\Delta_h$ ; and spatial average of the variation in the vote difference between the democratic and republican candidates from 2016 to 2020,  $W\Delta_v^{\text{dif}}$ ) and the target processes (variation in the vote difference between the democratic and republican candidates from 2016 to 2020,  $\Delta_v^{\text{dif}}$ )

	Source			
	$c$	$\Delta_u$	$\Delta_h$	$W\Delta_v^{\text{dif}}$
Conditional mutual information $\delta_i$	0.0266 (0.0235)	0.0325 (0.0234)	0.0212 (0.0237)	0.0303 (0.0242)
$p$ -value	<b>0.006</b>	<b>0.001</b>	0.160	< <b>0.001</b>

In parenthesis, we report the threshold value for significance of conditional mutual information; bold  $p$ -values refer to significant associations

**Table 3** Distribution of the variation of the total votes from 2016 to 2020 ( $\Delta_v^{\text{tot}}$ ), conditioned on the confirmed cases of COVID-19 ( $c$ ), the unemployment variation from 2016 to 2020 ( $\Delta_u$ ), and the variation in the health insurance coverage from 2013 to 2018 ( $\Delta_h$ )

	$P(\Delta_v^{\text{tot}} c)$			$P(\Delta_v^{\text{tot}} \Delta_u)$			$P(\Delta_v^{\text{tot}} \Delta_h)$		
	$c \leq 2.03$	$2.03 < c \leq 3.38$	$c > 3.38$	$\Delta_u \leq 0.02$	$0.02 < \Delta_u \leq 0.45$	$\Delta_u > 0.45$	$\Delta_h \leq 5.03$	$5.03 < \Delta_h \leq 8.19$	$\Delta_h > 8.19$
$\Delta_v^{\text{tot}} \leq 4.29$	0.232	0.321	0.447	0.391	0.309	0.300	0.340	0.302	0.358
$4.29 < \Delta_v^{\text{tot}} \leq 7.41$	0.275	0.362	0.363	0.368	0.344	0.288	0.349	0.341	0.310
$\Delta_v^{\text{tot}} > 7.41$	0.494	0.317	0.189	0.241	0.347	0.412	0.311	0.357	0.332

The three bins have the same count, and all the quantities are normalized by the population and expressed in percent

An equivalent analysis of the vote difference between the democratic and republican candidate from 2016 to 2020 helped clarify the presence of a negative association with the incidence of COVID-19 cases (Table 4). Specifically, counties where the difference between the two candidates increased (that is, the margin of Biden with respect to Trump increased) were those that suffered the least from COVID-19. Counties where Biden registered the largest margin did not seem to be identified by COVID-19 prevalence. The study of marginal probabilities (Table 4) also suggested that the increase in Biden's margin was stronger in counties that experienced more job losses, and it was weaker in counties that registered less unemployment. We found no indication that either of the dependencies of the vote difference on COVID-19 cases and unemployment rate were linear, as one can gather from the fact that the probability of an intermediate margin variation between the candidates was maximized in counties that suffered the

most from COVID-19 and the least from unemployment. Similar to the total vote count, the health insurance share did not seem to bear any effect either on the vote difference between the candidates, since marginal probabilities were all close to chance (Table 4).

## 4 Discussion and conclusions

There is almost consensus in the public opinion and media press that former President Trump's response to the COVID-19 pandemic hurt him in the 2020 U.S. presidential elections, favoring his opponent, current President Biden. Such a consensus is not fully backed by the technical literature, which is divided in the assessment of the link between voting behavior and COVID-19 incidence.

**Table 4** Distribution of the variation of the vote difference between the democratic and republican candidate from 2016 to 2020 ( $\Delta_v^{\text{dif}}$ ), conditioned on the confirmed cases of COVID-19 ( $c$ ), the unemployment variation from 2016 to 2020 ( $\Delta_u$ ), and the variation in the health insurance coverage from 2013 to 2018 ( $\Delta_h$ )

	$P(\Delta_v^{\text{dif}} c)$			$P(\Delta_v^{\text{dif}} \Delta_u)$			$P(\Delta_v^{\text{dif}} \Delta_h)$		
	$c \leq 2.03$	$2.03 < c \leq 3.38$	$c > 3.38$	$\Delta_u \leq 0.02$	$0.02 < \Delta_u \leq 0.45$	$\Delta_u > 0.45$	$\Delta_h \leq 5.03$	$5.03 < \Delta_h \leq 8.19$	$\Delta_h > 8.19$
$\Delta_v^{\text{dif}} \leq -3.27$	0.328	0.337	0.335	0.361	0.365	0.273	0.336	0.310	0.354
$-3.27 < \Delta_v^{\text{dif}} \leq -0.34$	0.277	0.341	0.382	0.391	0.339	0.270	0.343	0.324	0.333
$\Delta_v^{\text{dif}} > -0.34$	0.395	0.322	0.283	0.247	0.296	0.457	0.321	0.366	0.313

The three bins have the same count, and all the quantities are normalized by the population and expressed in percent

Baccini et al. [8] determined that Trump’s differential vote was negatively associated with COVID-19 cases, proposing that a reduction in the case count by twenty percent would have been sufficient for Trump to be re-elected. A similar conclusion was reached by Noland and Zhang [9], who found that Trump would have won the presidency again if the pandemic did not happen or if it were substantially mitigated, by thirty percent. Lake and Nie [10], instead, did not discover a negative association between voting preference for Trump and COVID-19 incidence. Although not statistically significant, their claims point in the opposite direction to Baccini et al. [8] and Noland and Zhang [9], offering partial evidence that COVID-19 might have benefited Trump in the elections.

Our spatial information-theoretic analysis supports the claims by Baccini et al. [8] and Nolan and Zhang [9] regarding an effect of COVID-19 incidence on voting behavior. Specifically, we determined that the prevalence of COVID-19 cases was associated with both the voter turnout and the election outcome. However, different from Baccini et al. [8] and Noland and Zhang [9], we did not find evidence that COVID-19 incidence hurt Trump re-election. Quite the opposite, our results indicate that counties that suffered the least from COVID-19 were those where Trump was outperformed by his opponent.

This discrepancy can be attributed to differences in data collection and methodology between our work and the study by Baccini et al. [8]. Importantly, Baccini et al. [8] considered 2689 counties rather than the complete set of 3105, and Noland and Zhang [9] used Trump’s share of votes as the dependent variable, rather than the vote difference between the democratic and republican candidates. From a methodological point of view, our spatial information-theoretic analysis allows one to properly account for the expected spatial patterns in voting behavior; these patterns should be particularly important in assessing the effect of COVID-19 pandemic whose diffusion was supported by social interactions and physical mobility of people [26]. In addition, compared to a regression analysis, our approach does not assume a linear response among the variables, thereby facilitating accurate representation of the public’s appraisal of the consequences of the pandemic [27, 28].

We propose that the positive association between COVID-19 cases and Trump’s performance in the elections is related to different attitudes of liberal and conservative voters toward the pandemic. As surveyed by the Republican pollster Neil Newhouse [29], “Republican voters were not taking the kinds of precautions that other voters were taking with respect to protecting themselves from the spread of the virus.” Before the elections took place, Takagi [30] found that COVID-19 incidence was higher in states where votes for Trump in 2016 were more; tenably, this could have been partially due to the President’s “unscientific and irresponsible claims... [that might have] also potentially increase[d] the risk of COVID-19 transmission among his voters.” At the same time, this could be associated with different views between conservative and liberal voters regarding individual responsibilities in public health debates [31]. Right after the elections, Takagi [30] performed an analogous study that confirmed the 2016 pattern of association between voting preference and COVID-19 incidence, highlighting several instances in which states that suffered the most from COVID-19 increased their support for Trump.

While we favor the link between COVID-19 spreading and political identity within the republican party, we cannot exclude the speculation of Lake and Nie [10] that “voters perceived Trump as better at dealing with a COVID-ravaged economy.” As argued by Lake and Nie [10] on the basis of a poll by The New York Times [32] close to the elections, “despite a late-shift towards Biden, polls generally showed the economy as a clear issue advantage for Trump.” However, our analysis of the role of unemployment on the elections only offers partial evidence in this direction. Specifically, we found that voter turnout was associated with unemployment rate, so that counties that registered the largest drop in employment from 2016 to 2020 were those that registered the largest increase in the participation in the election. Such a dependence did not reverberate in a preference for Trump; on the contrary, we observed that Biden’s margin was stronger in counties that experienced more job losses.

Looking into the variation in the health insurance coverage did not provide the same insight that was offered by Lake and Nie [10], whereby we discovered that the health insurance coverage was associated with

voter turn-out, but not with voting preference. While it is tenable that voters who enjoyed the ACA might have abandoned Trump, we do not find statistical evidence supporting this claim. Perhaps, this is due to the limited number of source processes that are included in our analysis, compared to the richness of the control variables that are part of the study of Lake and Nie [10]. This is a key limitation of our approach, which is not designed to account for many source processes due to challenges in estimating probability mass functions without an underlying mathematical model. With data on 3105 counties, we can reliably include not more than four source processes in the analysis (for three bins, this yields  $3^{(4+1)} = 243$  probabilities to be estimated from 3105 data points). As such, the approach cannot resolve the several control variables that are readily accounted for in the standard machinery of linear analysis. Another limitation is the lack of a granular means to account for the size of electoral votes and differentiate between urban and rural areas [33], which have likely reacted differently to Trump’s handling of the pandemic.

The main element of novelty of our study with respect to other studies on the link between the U.S. presidential elections and COVID-19 pandemic lies in the application of an information-theoretic approach in lieu of a regression analysis, which would fail to capture several of the identified associations (as discussed in the Appendix). An information-theoretic framework should be preferred when there is limited knowledge about the type of interactions among the spatial variables, whereby the approach does not postulate a-priori relationships among the variables. There are a number of applications in public health and economics that have benefited from information-theoretic approaches, such as the study of firearm violence [34], epidemic spreading [35,36], urban agglomerations [37], and sustainable developments [38].

Our work contributes to the field by demonstrating the value of spatial information-theoretic tools in the study of the mechanisms underlying government elections. Understanding these mechanisms is critical to support decision-making processes in urban sciences, which define the future of our cities as they face dramatic changes due to environmental and sociotechnical stressors, such as those posed by climate change and social justice.

**Acknowledgements** P. De Lellis was supported by the program “STAR 2018” of the University of Naples Federico II and Compagnia di San Paolo, Istituto Banco di Napoli—Fondazione, project ACROSS. M. Ruiz Marín was supported by Ministerio de Ciencia, Innovación y Universidades under grant number PID2019-107800GB-I00/AEI/10.13039/501100011033. M. Porfiri was partially supported by the National Science Foundation under grant numbers CMMI-1953135 and CMMI-2027990. This study was also part of the collaborative activities carried out under the programs of the region of Murcia (Spain): ‘Groups of Excellence of the region of Murcia, the Fundación Séneca, Science and Technology Agency’ project 19884/GERM/15.

### Appendix

To identify linear interactions among source and target variables, we considered two linear spatial autoregressive models of the form

$$\Delta_v^{\text{tot}} = b_1^{\text{tot}} c + b_2^{\text{tot}} \Delta_u + b_3^{\text{tot}} \Delta_h + b_4^{\text{tot}} W \Delta_v^{\text{tot}} + e^{\text{tot}}, \tag{12}$$

$$\Delta_v^{\text{dif}} = b_1^{\text{dif}} c + b_2^{\text{dif}} \Delta_u + b_3^{\text{dif}} \Delta_h + b_4^{\text{dif}} W \Delta_v^{\text{dif}} + e^{\text{dif}}, \tag{13}$$

where  $e^{\text{tot}} = [e_1^{\text{dif}}, \dots; e_{3105}^{\text{dif}}]$  and  $e^{\text{dif}} = [e_1^{\text{dif}}, \dots; e_{3105}^{\text{dif}}]$  are multivariate normal variables, with zero mean and covariance matrix  $\sigma^{\text{dif,tot}^2} I$  ( $I$  is an identity matrix of appropriate size, and  $\sigma^{\text{dif,tot}^2}$  the common variance of each scalar variable  $e_i^{\text{dif,tot}}$ ); coefficients  $b^{\text{tot}} = [b_1^{\text{tot}}; \dots; b_4^{\text{tot}}]$  and  $b^{\text{dif}} = [b_1^{\text{dif}}; \dots; b_4^{\text{dif}}]$  of regressions (12) and (13) were obtained using maximum likelihood estimation [25] and are reported in Tables 5 and 6, together with the corresponding  $p$ -values. Compared with the entropy analysis, a linear model fails to capture the effect of the pandemic on the variation in the total votes from 2016 to 2020 (Table 5). Furthermore, a linear analysis does not find a significant effect of unemployment variation on the variation of the vote difference  $\Delta_v^{\text{dif}}$  between the democratic and republican candidates from 2016 to 2020, and it does not allow to identify a spatial structure in  $\Delta_v^{\text{dif}}$  (Table 6). These results further support the necessity of accounting for nonlinear interactions between the target and source variables as made possible by our spatial information-theoretic approach.

**Table 5** Maximum likelihood estimation of the coefficients  $b^{\text{tot}}$  of the spatial autoregressive model (12) linking each of the chosen source processes (confirmed cases of COVID-19,  $c$ ; unemployment variation from 2016 to 2020,  $\Delta_u$ ; variation in the health insurance coverage from 2013 to 2018,  $\Delta_h$ ; and spatial average of the variation in the total votes from 2016 to 2020,  $W \Delta_v^{\text{tot}}$ ) and the target process (variation in the total votes from 2016 to 2020,  $\Delta_v^{\text{tot}}$ ); bold  $p$ -values refer to significant associations

	Source			
	$c$	$\Delta_u$	$\Delta_h$	$W \Delta_v^{\text{tot}}$
Coefficients $b^{\text{tot}}$	0.0558	1.2476	0.1690	0.2363
$p$ -value	0.486	< <b>0.001</b>	< <b>0.001</b>	< <b>0.001</b>



**Table 6** Maximum likelihood estimation of the coefficients  $b^{\text{dif}}$  of the spatial autoregressive model (13) linking each of the chosen source processes (confirmed cases of COVID-19,  $c$ ; unemployment variation from 2016 to 2020,  $\Delta_u$ ; variation in the health insurance coverage from 2013 to 2018,  $\Delta_h$ ; and spatial average of the variation in the vote difference between the democratic and republican candidates from 2016 to 2020,  $W\Delta_v^{\text{dif}}$ ) and the target process (variation in the vote difference between the democratic and republican candidates from 2016 to 2020,  $\Delta_v^{\text{dif}}$ ); bold  $p$ -values refer to significant associations

	Source			
	$c$	$\Delta_u$	$\Delta_h$	$W\Delta_v^{\text{dif}}$
Coefficients $b^{\text{dif}}$	– 0.2016	0.5931	0.0112	0.0526
$p$ -value	<b>0.008</b>	0.060	0.725	0.263

## References

- D. French, Its clear that America is deeply polarized. No election can overcome that (2020). <https://time.com/5907318/polarization-2020-election/>. Accessed 05 Jul 2021
- G. Viglione, Nature (2020). <https://www.nature.com/articles/d41586-020-03035-4>
- Trends in number of COVID-19 cases and deaths in the US reported to CDC, by state/territory (2021). [https://covid.cdc.gov/covid-data-tracker/#trends\\_dailytrendscases](https://covid.cdc.gov/covid-data-tracker/#trends_dailytrendscases). Accessed 05 Jul 2021
- J. Dawsey, M. Scherer, A. Parker, M. Viser, How Trump's erratic behavior and failure on coronavirus doomed his reelection (2020). <https://time.com/5907318/polarization-2020-election/>. Accessed 05 Jul 2021
- Trump pollster says COVID-19, not voter fraud, to blame for reelection loss (2021). <https://edition.cnn.com/2021/02/02/politics/trump-2020-reelection-loss-covid/index.html>. Accessed 05 Jul 2021
- B. Bennett, T. Berenson, How Donald Trump lost the election (2020). <https://time.com/5907973/donald-trump-loses-2020-election/>. Accessed 05 Jul 2021
- N. Bryant, US election 2020: Why Donald Trump lost (2020). <https://www.bbc.com/news/election-us-2020-54788636>. Accessed 05 Jul 2021
- L. Baccini, A. Brodeur, S. Weymouth, J. Popul. Econ. pp. 1–29 (2021)
- M. Noland, Y.E. Zhang, Peterson institute for international economics working paper (21-3) (2021)
- J. Lake, J. Nie, CESifo working paper (2021)
- A. Golan, E. Maasoumi, Econ. Rev. **27**(4–6), 317 (2008)
- M. Herrera, J. Mur, M. Ruiz, Pap. Reg. Sci. **95**(3), 577 (2016)
- F. López, M. Matilla-García, J. Mur, M. Ruiz Marín, Reg. Sci. Urban Econ. **40**(2–3), 106 (2010)
- M. Porfiri, M. Ruiz Marín, Proc. R. Soc. A **476**(2242), 20200113 (2020)
- US COVID-19 cases and deaths by state (2020). <https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/>. Accessed 05 Jul 2021
- Local area unemployment statistics (2020). <https://www.bls.gov/lau/>. Accessed 05 Jul 2021
- Health insurance coverage status—table s2701 (2021). <https://data.census.gov/cedsci/>. Accessed 13 Jul 2021
- Full presidential results (2020). <https://www.politico.com/2020-election/results/>. Accessed 05 Jul 2021
- U.S. presidential election results 2020: Biden wins (2020). <https://www.nbcnews.com/politics/2020-elections/president-results>. Accessed 05 Jul 2021
- The ultimate Daily Kos Elections guide to all of our data sets (2021). <https://www.dailykos.com/stories/2018/2/21/1742660/-The-ultimate-Daily-Kos-Elections-guide-to-all-of-our-data-sets>. Accessed 05 Jul 2021
- 2016 Connecticut presidential election results (2016). <https://www.politico.com/2016-election/results/map/president/connecticut/>. Accessed 05 Jul 2021
- 2016 Election results (2016). <https://edition.cnn.com/election/2016/results>. Accessed 05 Jul 2021
- United States Census Bureau. <https://www.census.gov/programs-surveys/popest/data/data-sets.html>. Accessed 08 Jul 2021
- P. Sulewski, J. Appl. Stat. pp. 1–20 (2020)
- J. LeSage, R.K. Pace, *Introduction to Spatial Econometrics* (Chapman and Hall/CRC, London, 2009)
- A. Scala, A. Flori, A. Spelta, E. Brugnoli, M. Cinelli, W. Quattrocchi, F. Pammolli, Sci. Rep. **10**(1), 1 (2020)
- C.A. Harper, L.P. Satchell, D. Fido, R.D. Latzman, Int. J. Mental Health Addict. (2020). <https://doi.org/10.1007/s11469-020-00281-5>
- B.J. Kuper-Smith, L.M. Doppelhofer, Y. Oganian, G. Rosenblau, C. Korn, (2020). <https://psyarxiv.com/epcyb/>
- Cnn transcripts (2004). <http://transcripts.cnn.com/TRANSCRIPTS/2004/08/cnr.08.html>. Accessed 13 Jul 2021
- H. Takagi, J. Med. Virol. **93**(7), 4071 (2021)
- H. Lundell, J. Niederdeppe, C. Clarke, J. Health Commun. **18**(9), 1116 (2013)
- A. Burns, M. Martin. Voters prefer Biden over Trump on almost all major issues. <https://rebrand.ly/p4r9k>. Accessed 05 Jul 2021
- J. Molinaro, S. Spjeldnes, The electoral college and the rural-urban divide (2021). <https://www.aspeninstitute.org/blog-posts/the-electoral-college-and-the-rural-urban-divide/>. Accessed 05 Jul 2021
- M. Porfiri, R.R. Sattanapalle, S. Nakayama, J. Macinko, R. Sipahi, Nat. Hum. Behav. **3**(9), 913 (2019)
- E.Y. Erten, J.T. Lizier, M. Piraveenan, M. Prokopenko, Entropy **19**(5), 194 (2017)
- S.M. Kissler, C. Viboud, B.T. Grenfell, J.R. Gog, J. R. Soc. Interface **17**(164), 20190628 (2020)
- R. Murcio, R. Morphet, C. Gershenson, M. Batty, PLoS One **10**(7), e0133780 (2015)
- X. Liang, D. Si, X. Zhang, Int. J. Environ. Res. Public Health **14**(10), 1219 (2017)