



OPEN

The impact of social ties and SARS memory on the public awareness of 2019 novel coronavirus (SARS-CoV-2) outbreak

Haohui Chen^{1,2✉}, Cecile Paris¹ & Andrew Reeson¹

This study examines publicly available online search data in China to investigate the spread of public awareness of the 2019 novel coronavirus (SARS-CoV-2) outbreak. We found that cities that had previously suffered from SARS (in 2003–04) and have greater migration ties to Wuhan had earlier, stronger and more durable public awareness of the outbreak. Our data indicate that 48 such cities developed awareness up to 19 days earlier than 255 comparable cities, giving them an opportunity to better prepare. This study suggests that it is important to consider memory of prior catastrophic events as they will influence the public response to emerging threats.

Public awareness is important in managing the spread of infectious diseases. Here, public awareness is defined as knowledge and understanding among the population about the risk of infectious diseases. Individual actions, such as increased attention to hygiene and avoiding crowds, can reduce disease spread. Awareness also supports rapid identification and treatment of new cases and facilitates collective responses, such as closures of schools or transit systems¹. In the modern world, diseases can move faster than ever due to the growing movement of people between cities, regions and countries. However, digital technology means information can move even faster, providing an opportunity for individuals and communities to protect themselves ahead of the disease itself arriving². This study considers the spread, and persistence, of public awareness of the novel coronavirus which emerged in late 2019. During the first few weeks of this outbreak, there was little coverage from mainstream media outlets, providing an unusual opportunity to study the spread of awareness of an emerging disease via other channels.

Previous studies show that the spread of awareness is strongly related to the physical locations of individuals in a social network in relation to the unfolding events^{2–4}, termed the *social ties* effect. In online social networks, people with more connections tend to receive earlier warnings of catastrophic events. For example, during Hurricane Sandy in the USA, Twitter users with more followers had an awareness lead-time of up to 26 h than less connected users⁴. Moreover, the magnitude of awareness increases over decreasing distances to the epidemic centers. For example, public awareness in Weibo, a Chinese social media platform, was two orders of magnitude stronger for the H7N9 influenza outbreak that occurred in China than the Middle East Respiratory Syndrome Coronavirus (MERS-CoV) outbreak that occurred elsewhere⁵.

Experience of similar events, such as outbreaks of H5N1 influenza in 2001, SARS (Severe Acute Respiratory Syndrome) in 2003, H1N1 influenza in 2009 and Ebola in 2014 is also likely to influence awareness. In China, the outbreak of SARS between 2003 and 2004 caused a total of 7,429 reported cases and 685 deaths⁶, and had a lasting traumatic impact on survivors and communities^{7,8}. In this work, we set out to test whether public awareness of the new disease outbreak is related to social ties distance from the place impacted by the epidemic and past experience of the SARS epidemic in 2003. The SARS outbreak was 17 years ago, but its horror might still condition public awareness of lethal infectious diseases. To the best of our knowledge, few studies have been carried out to understand how past severe outbreaks affect public awareness when a new outbreak occurs. This study estimates the post-SARS effect, called *SARS memory* effect, on the current outbreak.

We use the continuing coronavirus outbreak as our case study to estimate the effects of *social ties* and *SARS memory* on the spread of public awareness. Chinese authorities officially announced human-to-human transmission of the 2019 novel coronavirus (SARS-CoV-2) on January 20th, 2020⁹. The outbreak originated in Wuhan City, a major transportation hub in central China long known as the “Nine Provinces” thoroughfare transiting more than 120 million passengers every year¹⁰. The massive numbers of transits provided a perfect opportunity for the virus to spread. Another feature is the timing of the outbreak, close to the Spring Festival travel season,

¹Data61, Commonwealth Scientific and Industrial Research Organization, Canberra, Australia. ²Monash University, Melbourne, Australia. ✉email: Caronhaohui.chen@data61.csiro.au

Chunyun, which started on January 10th, 2020. Accordingly, this study focuses on the time period between December 15th, 2019 and January 23rd, 2020.

Data and methods

Public awareness measurement. Seeking epidemic-related information online can provide an indicator of public awareness of this new disease. In this study, we use the Baidu Search Index (BSI), available publicly at <https://index.baidu.com>, to measure public awareness over time and locations (e.g., city). Currently, Baidu releases online search information across 364 Chinese cities, covering all four direct-administered municipalities, two special administrative regions (Macau and Hong Kong), 293 prefectural-level cities, Taiwan (released as a city in Baidu) and additional 64 country-level cities. The total number of internet users using the Baidu search engine reached 649 million in 2014, accounting for 47.9% of China's population¹¹. BSI has been used to predict epidemic outbreak¹², HIV/AIDS incidence¹³ and tourism flows¹⁴, suggesting BSI can provide a representative proxy for public awareness. BSI provides a weighted index for each search term. In this study, we used a search term in Chinese that combines “Wuhan” and “pneumonia” (called “pneumonia” hereafter), as Chinese internet users widely used it during the time period of this study. We also tried another search term in Chinese “novel coronavirus”, but it did not exhibit a search surge on Baidu. Due to the privacy concern, Baidu masks daily readings that are below 57 as zero. Therefore, we used the maximum BSI value of the search term “common cold” (in Chinese) between Dec 10th and 31st, 2019, to control the size effect for each city. We use the Ljung-Box test¹⁵ to estimate whether or not the daily readings of “common cold” are stationary. As a result, the daily readings of 18 out of 364 cities were found to be non-stationary. In this study, that means the “common cold” readings in those 18 cities exhibit seasonality or a trend, so they were excluded from this study. The magnitude of public awareness of the Wuhan outbreak over time t and city $i \in \{1, \dots, 346\}$ can be represented as $O_{t,i}^{\text{COVID-19}}$, as defined below. $S_{t,i}^{\text{COVID-19}}$ and $S_{t,i}^{\text{cold}}$ represent the BSI values of the search terms “pneumonia” and “common cold” respectively. The BSI raw data is provided in S1 of the supplementary materials (SM).

$$O_{t,i}^{\text{COVID-19}} = \frac{S_{t,i}^{\text{COVID-19}}}{\text{MAX}\left(\left\{S_{t_{\text{Dec 15th, 2019}, i}}^{\text{cold}}, \dots, S_{t_{\text{Dec 31st, 2019}, i}}^{\text{cold}}\right\}\right)} \quad (1)$$

The earliest day the magnitude of public awareness exceeds the arbitrary thresholds $C \in \{1.5, 2, 3, 4\}$ is defined as the earliest warning day, $t_{\text{warning}(i)}$, for city i . We also define the starting day of *Chunyun* as t_{chunyun} , indicating the onset day when it is likely the virus would reach all cities. As *Chunyun* transited approximately 3 billion passengers in 40 days in 2019¹⁶, crowded transport hubs create perfect opportunities for the virus to spread. Therefore, the earlier the lead-time awareness, the better for infection control. The lead-time of awareness for city i is thus defined as:

$$\Delta t_i = t_{\text{warning}(i)} - t_{\text{chunyun}} \quad (2)$$

Awareness typically follows a cyclical process, called the unaware-aware-unaware (UAU) process, as time passes. Keeping the public at a high level of awareness could help mitigate the virus transmission process. Therefore, we also measure the awareness retention rate as the average of magnitude from the next day of $t_{\text{warning}(i)}$ to the day of *Chunyun* over the magnitude at $t_{\text{warning}(i)}$.

$$\Delta O_i = \frac{\text{mean}\left(\sum_{1+t_{\text{warning}(i)}}^{t_{\text{chunyun}}} O_{t,i}^{\text{COVID-19}}\right)}{O_{t_{\text{warning}(i)}}^{\text{COVID-19}}} \quad (3)$$

Measuring social ties. Microblogs (e.g., Weibo) and private social media (e.g., WeChat) are the primary communication tools used by most Chinese people¹¹. While information flows cannot be observed directly, empirical studies show that social networks are influenced by long distance travel¹⁷. We therefore use migration flows as a proxy for long-distance information flows. To be more specific, if workers born and raised in city A now work in city B, they are likely to relate information about an epidemic in city B back to friends and family in city A. This is particularly relevant in the Chinese context, where migrant workers account for more than one-third of the working population¹⁸.

We use the migration flows extracted from the Baidu Migration Matrix (BMM) to build a migration network (Fig. 1). We then compute the shortest steps between any city to Wuhan, deriving the variable *social ties* for city i as $D_i \in (1, 8)$. Wuhan and the cities located in Hubei province have $D_i = 1$, while cities located far away from Wuhan tend to have larger values, e.g., $D_{i=\text{Lhasa}} = 6$.

Moreover, we added $D_{i=\text{HongKong}} = 1$, based on the rationale that Hong Kong has more airline traffic flows to Wuhan than Shanghai¹⁹, which has $D_{i=\text{Shanghai}} = 1$. Macau and Taiwan both have frequent traffic flows to Hong Kong¹⁹, so we added $D_{i=\text{Macau}} = 2$ and $D_{i=\text{Taiwan}} = 2$.

Measuring SARS memory. We collected all reported SARS cases in mainland China, Hong Kong, Taiwan and Macau, and assign the numbers of cases to each city as $\text{SARS}_i, i \in \{1, \dots, 346\}$. The values range between zero (no reported cases) and 2,521 ($i = \text{Beijing}$), with an average of 70.8 and a median of 4 cases in cities with at least one case reported. Given the long-tailed distribution of reported SARS cases, we use the logarithm. As most cities reported zero cases of SARS, we use the equation below to transform the variable SARS_i :

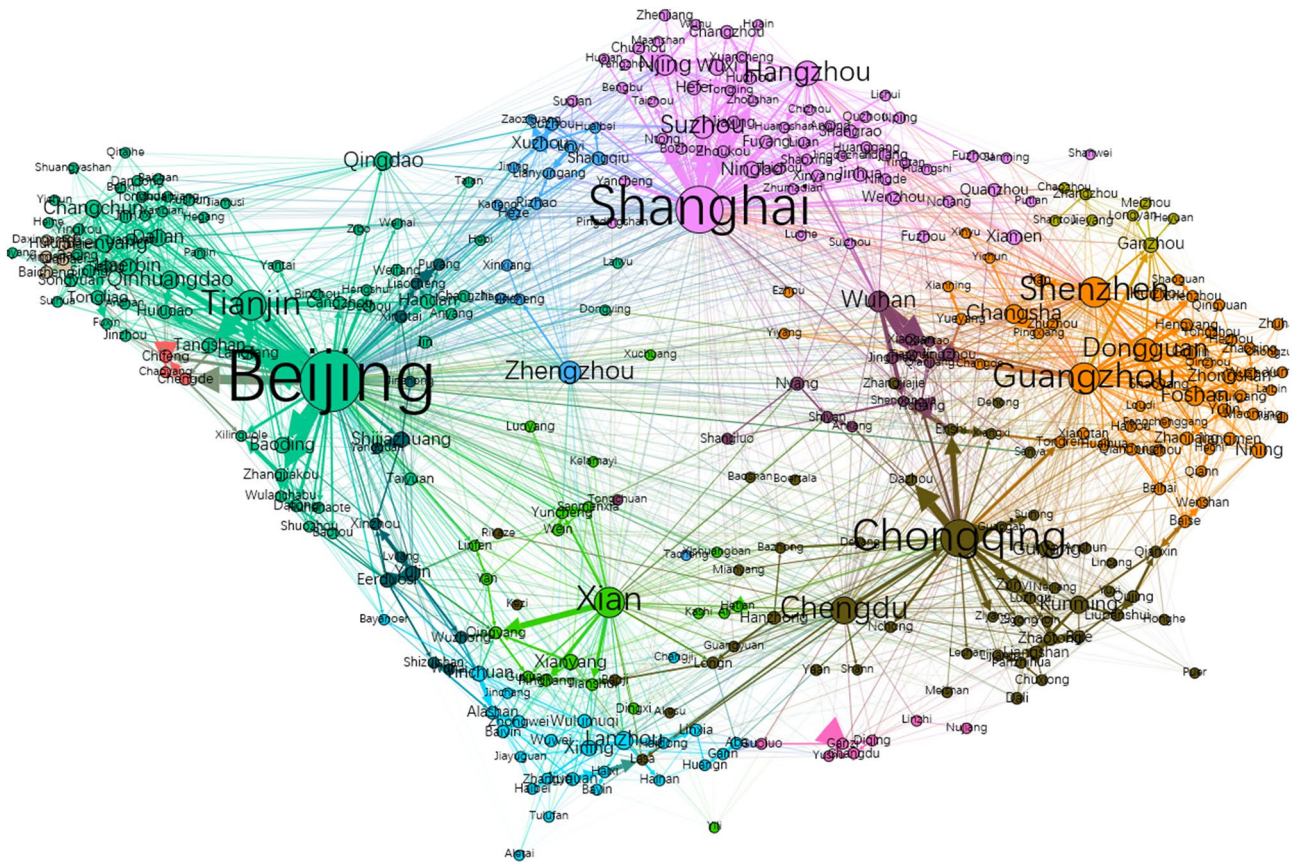


Figure 1. The migration network based on Baidu Migration Matrix. The nodes represent 361 Chinese cities whose migration flows are available to us, and the length and thickness of the edges indicate the travel frequency between the corresponding pair of cities. The node size is proportional to the numbers of transit passengers at the corresponding city. The network does not include Hong Kong, Macau and Taiwan, as the Baidu Migration Matrix does not include them. The graph is produced in Gephi v0.9.2, available at <https://gephi.org/>.

$$SARS_i = \log(SARS_i + 1)$$

The log-transformation would alter the distributions of the variable, but it is still appropriate for testing the assumptions, i.e., cities with more reported SARS cases would exhibit stronger, earlier and more durable public awareness.

Estimating social ties and SARS memory effects. We build three groups of regression models to estimate the effects of *social ties* and *SARS memory* on public awareness measures Δt_i , $O_{t_{warming(i)}}^{COVID-19}$ and ΔO_i respectively. $GDP_per_capita_i$ represents the gross domestic product (GDP) per capita for city i . $SubProvincial_i$ indicates whether or not city i has sub-provincial or greater administrative power. Sub-provincial cities are mostly capitals of the provinces in which they are located, or important cities designated by the central government. Four cities, including Beijing, Shanghai, Tianjin and Chongqing, which are under direct control of the central government are also labelled as $SubProvincial_i = 1$. Those sub-provincial and above cities have much better facilities and expertise for infection control than other cities²⁰, so we assume residents could be more alert. $SubProvincial_i$ is used to control the effects of *administrative level*. We also introduce Euclidean distances as a control variable, denoted as d_i , which presents the straight-line distance between Wuhan to each city. The introduction of d_i is to control for the fact that cities physically closer to Wuhan are likely to get more news from there regardless of the strength of social ties.

$$\Delta t_i = GDP_per_capita_i + SubProvincial_i + D_i + d_i + SARS_i + D_i * SARS_i \tag{4}$$

$$O_{t_{warming(i)}}^{COVID-19} = GDP_per_capita_i + SubProvincial_i + D_i + d_i + SARS_i + D_i * SARS_i \tag{5}$$

$$\Delta O_i = GDP_per_capita_i + SubProvincial_i + D_i + d_i + SARS_i + D_i * SARS_i \tag{6}$$

All data necessary to replicate the analysis is attached as S2 of the SM.

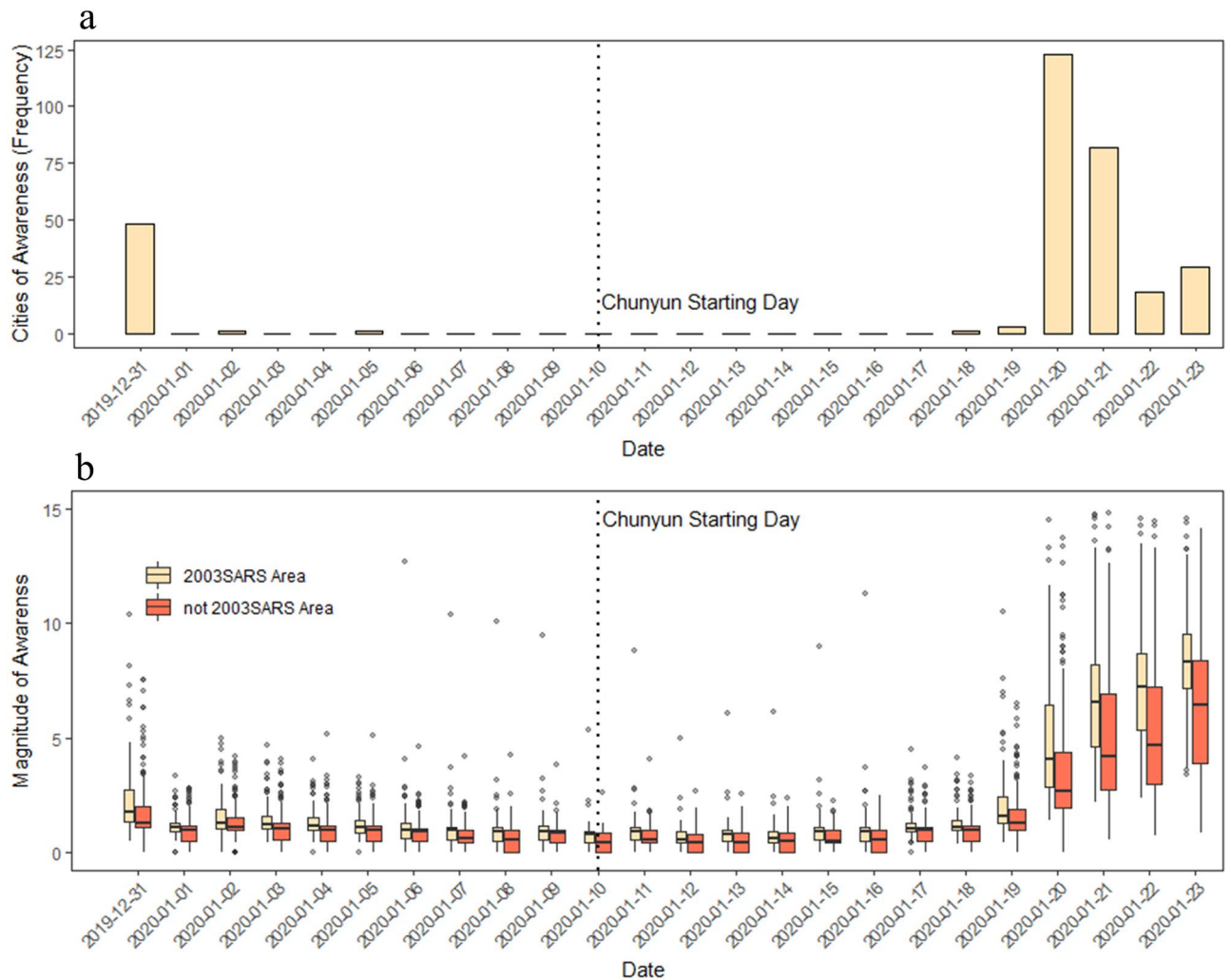


Figure 2. Public awareness over time. **(a)** The frequency distributions of cities that exhibit the first significant signal of awareness over time. The number of cities for which searches for the combined term “Wuhan” and “pneumonia” exceed $C = 3$ times the search term “common cold” is reported every day. **(b)** Public awareness on the topic of “pneumonia” over time. All 346 cities exhibit at least some searches of the term “pneumonia” during the initial outbreak period. Of these, 326 cities recorded searches about it as early as Dec 31st, 2019. Cities are divided into two groups according to whether or not they had reported SARS cases in 2003–04. The mean values of awareness magnitude were computed on a daily basis for two groups of cities respectively. Accordingly, a paired t-test was performed on those two time-series, and we found the cities that had reported SARS cases had greater of awareness (t-statistic: 3.56; degrees of freedom: 23; $p < 0.005$).

Results

The early warnings of the outbreak.

As early as Dec 31st, 2019, when Wuhan Municipal Health Commission first informed the public about the emerging pneumonia cases²¹, most of the cities (326 out of 346) exhibited at least some awareness of the emerging SARS-CoV-2 outbreak (Fig. 2b). However, awareness then decreased until Jan 19th, 2020, one day before the Chinese Centre for Disease Control and Prevention confirmed human-to-human transmissions of the novel coronavirus⁹. Since Jan 20th, 2020, overall awareness increased by a magnitude of at least five, demonstrating significant awareness across all cities (Fig. 2b). Awareness remained low as the epidemic spread, falling close to its lowest point on the starting day of *Chunyun* (Jan 10th, 2020). Considering cities that showed initial novel coronavirus awareness levels at least 1.5 times that of the search term “common cold”, we found a total of 166 alert cities as early as Dec 31st, 2019 (48 cities at a tighter threshold of $C = 3.0$ times, illustrated in Fig. 2a). However, awareness decreased significantly during *Chunyun*.

The evolution of public awareness over time followed an unusual pattern. In a typical UAU process, people are unaware of emerging catastrophic events until they are told by their social contacts. They remain aware during the event, and awareness then fades subsequently^{2,22}. However, during the Wuhan outbreak, the public experienced a process as aware-unaware-aware, with public awareness declining during the early phase of the outbreak.

Dividing cities into two groups according to whether or not they had reported SARS cases in 2003–04, we found the cities that had been struck by SARS to be more alert during onset (Fig. 2b). Therefore, we believe the *SARS memory* still conditions public awareness. We provide evidence of its effects at the end of this section.

	Dependent variable						
	Lead-time Δt_i ($C = 3.0$)						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
SARS _i	1.916*** (0.345)		2.566*** (0.664)		4.644 (4.531)	1.590*** (0.593)	-2.276 (4.016)
Social ties D_i		-3.222*** (0.271)	-2.753*** (0.286)		-2.310*** (0.372)	-1.476*** (0.315)	-1.367*** (0.334)
$D_i * SARS_i$			-0.617*** (0.238)		-0.660*** (0.243)	-0.412** (0.206)	-0.453** (0.211)
log(d_i)				-3.567*** (0.527)	-1.125 (0.720)	-1.744*** (0.490)	-2.139*** (0.637)
log(d_i) * SARS _i					-0.267 (0.679)		0.583 (0.599)
SubProvincial _i						8.516*** (1.135)	8.763*** (1.163)
GDP_per_capita _i						2.818*** (0.599)	2.762*** (0.602)
Constant	-8.571*** (0.470)	3.766*** (1.007)	1.610 (1.104)	16.287*** (3.523)	7.441* (4.121)	-22.169*** (7.068)	-19.291** (7.662)
Observations	306	306	306	306	306	305	305
R ²	0.092	0.318	0.357	0.131	0.368	0.535	0.537
Adjusted R ²	0.089	0.316	0.351	0.128	0.358	0.526	0.526
Residual SE	7.356 (df=304)	6.376 (df=304)	6.211 (df=302)	7.196 (df=304)	6.176 (df=300)	5.312 (df=298)	5.312 (df=297)
F statistic	30.820*** (df=1; 304)	141.645*** (df=1; 304)	55.885*** (df=3; 302)	45.859*** (df=1; 304)	34.999*** (df=5; 300)	57.218*** (df=6; 298)	49.170*** (df=7; 297)
AIC	2093.70	2006.20	1992.24	2080.25	1990.93	1893.62	1894.77

Table 1. Estimate social ties and SARS memory effects on lead-time advantage. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Awareness advantage. Three features define the awareness advantage of *alert* cities, including early awareness, strong magnitude and high retention of awareness.

Lead-time advantage. During onset, between Dec 31st, 2019 and Jan 23rd, 2020, 266 cities exhibited significant public awareness (using a threshold at $C = 3.0$; 210 cities at $C = 4.0$; 314 cities at $C = 2.0$; and 322 cities at $C = 1.5$). The lead-time advantage Δt_i ranges between -13 and 10 ($C = 3.0$), with an average of -7.4 days and a median of -10 days. Forty-eight cities emerge with early signals of public awareness, as early as Dec 31st, 2019, while for most others (255 cities), awareness is as late as Jan 20th, 2020 (Fig. 2a).

Magnitude of awareness. In terms of the magnitude of awareness, all 346 cities exhibit at least some awareness during the onset. The values of $O_{i,warning(i)}^{COVID-19}$ range between 0.50 and 61.51, with an average of 2.18 and a median of 1.49. Wuhan undoubtedly ranked the first, while Wuzhong in Ningxia ranked last. Cities in Hubei province exhibit much greater awareness, with an average of 7.83 and a median of 3.46. Shenzhen, Shanghai and Beijing also had high scores at 10.40, 10.65, and 9.78 respectively. Those three cities are both close to Wuhan in terms of social ties D_i and struck by SARS.

Retention of awareness. Even though most of the cities exhibit at least some awareness as early as Dec 31st, 2019, only a few retain it over the following weeks as the virus began to spread. The retention rates, ΔO_i , range between zero and 137%, with an average of 54% and a median of 55%. Eight cities lost awareness before *Chunyun*, while four cities developed greater awareness. Xilingol League in Inner Mongolia ranked 4th, with a retention rate at 103%. Xilingol is far away from Wuhan in terms of social ties distance, but it was struck by SARS. It is worth noting that a confirmed case of plague was reported in Xilingol on Nov 16th, 2019, only 45 days before the Wuhan authority confirmed the emerging pneumonia cases²¹.

Estimation of the social ties and SARS memory effects. The effects of *social ties* and *SARS memory* on the lead-time advantage are estimated according to Eq. 4, controlled by Euclidean distances, GDP per capita and the city's *administrative level* (Table 1). We found that, in model (3) in Table 1, SARS_i exhibits positive effects, while D_i shows a negative association with awareness. That means cities of strong *SARS memory* and which are closer to Wuhan in terms of *Social ties* develop early awareness. Moreover, the interaction term $D_i * SARS_i$ exhibits negative effects, indicating that the *SARS memory* effect becomes stronger where cities are closer to Wuhan in terms of *social ties distance*.

	Dependent variable					
	Magnitude of awareness at the earliest day of awareness: $O_{i,t}^{19-nCoV}$					
	(1)	(2)	(3)	(4)	(5)	(6)
$SARS_i$	0.550*** (0.163)		1.308*** (0.350)		25.629*** (1.539)	23.752*** (1.432)
Social ties D_i		-0.791*** (0.126)	-0.619*** (0.133)		-0.100 (0.116)	0.164 (0.110)
$D_i * SARS_i$			-0.412*** (0.125)		-0.284*** (0.082)	-0.218*** (0.075)
$\log(d_i)$				-2.534*** (0.204)	-0.730*** (0.240)	-0.988*** (0.222)
$\log(d_i) * SARS_i$					-3.577*** (0.231)	-3.354*** (0.214)
$SubProvincial_i$						2.438*** (0.415)
$GDP_per_capita_i$						0.961*** (0.206)
Constant	1.874*** (0.211)	5.071*** (0.495)	4.342*** (0.546)	19.208*** (1.381)	6.933*** (1.370)	-2.702 (2.643)
Observations	338	338	338	338	338	336
R^2	0.033	0.106	0.141	0.315	0.653	0.720
Adjusted R^2	0.030	0.103	0.134	0.313	0.648	0.714
Residual SE	3.518 (df=336)	3.382 (df=336)	3.324 (df=334)	2.961 (df=336)	2.120 (df=332)	1.915 (df=328)
F statistic	11.384*** (df=1; 336)	39.693*** (df=1; 336)	18.349*** (df=3; 334)	154.271*** (df=1; 336)	124.843*** (df=5; 332)	120.568*** (df=7; 328)
AIC	1813.52	1787.04	1777.33	1697.07	1475.5	1400.51

Table 2. Estimate social ties and SARS memory effects on the Magnitude of awareness at the earliest day of awareness. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

While controlling the model with Euclidean distances (model (5) in Table 1), we found that SARS memory effect becomes non-significant, but social ties and its interaction with SARS memory hold. Meanwhile, Euclidean distances are non-significant, even though it exhibits a negative effect on its own in model (4) in Table 1.

We further control the model with GDP per capita and administrative level (models (6) & (7) in Table 1). Using Akaike information criterion (AIC) to select the best model²³, we found the performance of model (6) and (7) are very similar. However, because model (6) achieves a slightly lower AIC score²⁴ at 1893.62 with fewer degrees of freedom (df=8) than model (7) (AIC=1894.77, df=9), model (6) is [very slightly] preferred. For more information about the model selection for Eq. 4, 5 and 6, see File S3 in the SM. In model (6) in Table 1, we found that both social ties and Euclidean distances exhibit negative effects, but the social ties effects decrease almost half compared to model (5) in Table 1. The SARS memory effects hold. Also, the interaction term $D_i * SARS_i$ is still significant, which means cities with stronger SARS memory will develop more lead-time advantage, particularly when they are closer to Wuhan. For example, Changchun in Jilin province with 34 SARS cases and far away from Wuhan still achieved a ten days lead-time advantage. The cities that did not exhibit awareness, such as Qaramay and Heihe, are mainly located far away from Wuhan and did not suffer from the SARS outbreak. GDP per capita and the binary variable $SubProvincial_i$ exhibit significant positive effects on the lead-time advantage.

The effects of social ties and SARS memory on the magnitude of awareness are estimated according to Eq. 5 (Table 2). Similar to the findings in Table 1, $SARS_i$ memory positively affects public awareness in all models. Social ties D_i show a significant negative effect only in the models without controlling variables (model (2) & (3) in Table 2). However, the interaction term between social ties with SARS memory show a significant negative effect. Using AIC-based model selection method, we found that model (6), which control by Euclidean distances, GDP per capita and the administrative level, is the best model. The effects of administrative level and development level both exhibit positive effects on the magnitude of awareness. We hypothesize that residents with better education (proxied by GDP per capita) better understand the danger of deadly infectious diseases and, accordingly, tend to seek up-to-date information online.

The effects of social ties and SARS memory on retention of awareness are estimated according to Eq. 6. Model (6) in Table 3 is the best model based on the model selection using AIC. Unlike the results in Tables 1 and 2, we observe no effects from SARS memory (model (6) in Table 3). When we control Euclidean distances, development level and administrative level, the explanatory power of the model is still relatively weak (Adj. $R^2 = 0.104$). It seems the decreasing awareness is a collective behavior that occurred simultaneously. Interestingly, social ties have a significant effect while the Euclidean distances do not. Development level exhibits positive effects, which suggests residents of better educated cities could be more alert during the epidemic onset. However, administrative

	Dependent variable					
	Awareness retention rate: ΔO_i					
	(1)	(2)	(3)	(4)	(5)	(6)
SARS _i	0.003 (0.009)		-0.055*** (0.019)		-0.268** (0.131)	-0.106 (0.129)
Social ties D _i		-0.011 (0.007)	-0.017** (0.007)		-0.017* (0.010)	-0.026*** (0.010)
D _i * SARS _i			0.022*** (0.007)		0.020*** (0.007)	0.019*** (0.007)
log (d _i)				-0.008 (0.013)	-0.007 (0.020)	0.016 (0.020)
log (d _i) * SARS _i					0.032 (0.020)	0.009 (0.019)
SubProvincial _i						-0.191*** (0.037)
GDP_per_capita _i						0.059*** (0.019)
Constant	0.536*** (0.011)	0.577*** (0.027)	0.597*** (0.030)	0.591*** (0.086)	0.643*** (0.116)	-0.096 (0.237)
Observations	336	336	336	336	336	334
R ²	0.0003	0.008	0.037	0.001	0.046	0.123
Adjusted R ²	-0.003	0.005	0.028	-0.002	0.032	0.104
Residual SE	0.183 (df= 334)	0.183 (df= 334)	0.180 (df= 332)	0.183 (df= 334)	0.180 (df= 330)	0.172 (df= 326)
F statistic	0.097 (df= 1; 334)	2.528 (df= 1; 334)	4.230*** (df= 3; 332)	0.401 (df= 1; 334)	3.214*** (df= 5; 330)	6.511*** (df= 7; 326)
AIC	-182.53	-184.96	-190.92	-182.83	-190.14	-215.96

Table 3. Estimate social ties and SARS memory effects on the awareness retention rate. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

level shows a negative effect. It seems residents living in important cities (in terms of administrative power) lost interest in the disease before *Chunyun*.

Discussion

The novel coronavirus outbreak is still spreading, with a growing death toll around the world. From this study we found that the spread of public awareness varied markedly across Chinese cities. Through controlling for development, administrative levels, and Euclidean distances, we observe cities that were struck by SARS and have more migration to Wuhan, showed earlier, stronger and more durable public awareness of the outbreak. These cities will have been better prepared to respond to the virus if and when it arrived. Specifically, 48 cities had developed public awareness as early as Dec 31st, 2019, with up to 19 days of lead-time advantage, compared to some other 255 cities. The study suggests that memory of previous events, as well as social links to an emerging threat, may influence public behaviour.

Greater awareness could help slow the spread of a disease, for example through increased attention to hygiene, mask-wearing and reduced interpersonal contact. It might also facilitate collective responses such as enforced quarantine measures. However, in some circumstances enhanced awareness could have negative impacts, such as unnecessary panic or ostracism of groups perceived as being at greater risk of infection.

Due to the lack of infection statistics, we cannot yet statistically estimate the effect of public awareness on the subsequent seriousness of the outbreak. We note that Xilingol League in Inner Mongolia, which had relatively stronger and more durable public awareness, had fewer cases (two cases as reported at Feb 7th, 2020²⁵) than other cities in the same province (totally 50 cases, with an average of 4.55 cases per city²⁵).

To the best of our knowledge, this study is the first to investigate how memory of previous catastrophic events, e.g., SARS, and *social ties* could affect the spread of public awareness. Further studies will be needed to understand whether this holds in other contexts, beyond the unusual circumstances of the novel coronavirus in Wuhan. As much of the world has subsequently suffered from this virus, it at least gives cause to hope that we will be more aware, and so respond more rapidly to, a future pandemic.

Data availability

All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

Received: 15 March 2020; Accepted: 28 September 2020

Published online: 26 October 2020

References

1. Bootsma, M. C. J. & Ferguson, N. M. The effect of public health measures on the 1918 influenza pandemic in U.S. cities. *Proc. Natl. Acad. Sci. USA* **104**, 7588–7593 (2007).
2. Granell, C., Gómez, S. & Arenas, A. Dynamical interplay between awareness and epidemic spreading in multiplex networks. *Phys. Rev. Lett.* **111**, 128701 (2013).
3. Kryvasheyeu, Y. *et al.* Rapid assessment of disaster damage using social media activity. *Sci. Adv.* **2**, e1500779 (2016).
4. Kryvasheyeu, Y., Chen, H., Moro, E., Van Hentenryck, P. & Cebrian, M. Performance of social network sensors during Hurricane Sandy. *PLoS ONE* **10**, e0117288 (2015).
5. Fung, I.C.-H. *et al.* Chinese social media reaction to the MERS-CoV and avian influenza A(H7N9) outbreaks. *Infect. Dis. Poverty* **2**, 31 (2013).
6. Smith, R. D. Responding to global infectious disease outbreaks: Lessons from SARS on the role of risk perception, communication and management. *Soc. Sci. Med.* **63**, 3113–3123 (2006).
7. Gardner, P. J. & Moallem, P. Psychological impact on SARS survivors: Critical review of the English language literature. *Can. Psychol.* **56**, 123–135 (2015).
8. Füller, H. Pandemic cities: biopolitical effects of changing infection control in post-SARS Hong Kong. *Geogr. J.* **182**, 342–352 (2016).
9. High-level expert group of the National Health Commission answers reporters' questions about the novel coronavirus pneumonia. CCTV (2020). Available at: <https://m.news.cctv.com/2020/01/20/ARTI4F17LEu8TRQIsnde93B200120.shtml>. (Accessed: 10th May 2020)
10. Wuhan Urban and Rural Construction Bureau - Wuhan transits ranked No. 1, becoming the largest transit station in China. *Wuhan Urban and Rural Construction Bureau* (2014). Available at: https://cjw.wuhan.gov.cn/content/2014-06/12/content_325895.htm. (Accessed: 1st February 2020)
11. China Internet Network Information Center The Thirty-Sixth Statistics Report of China Internet Development. CNNIC (2015).
12. Liu, K. *et al.* Using Baidu search index to predict dengue outbreak in China. *Sci. Rep.* **6**, 38040 (2016).
13. He, G. *et al.* Using the Baidu search index to predict the incidence of HIV/AIDS in China. *Sci. Rep.* **8**, 9038 (2018).
14. Huang, X., Zhang, L. & Ding, Y. The Baidu Index: uses in predicting tourism flows—a case study of the Forbidden City. *Tour. Manag.* **58**, 301–306 (2017).
15. Ljung, G. M. & Box, G. E. P. On a measure of lack of fit in time series models. *Biometrika* **65**, 297–303 (1978).
16. Approximately 3 billion passengers transited in the Chunyun of 2019. *Xinhuanet* (2019). Available at: https://www.xinhuanet.com/politics/2019-03/01/c_1124181637.htm. (Accessed: 2nd February 2020)
17. Cho, E., Myers, S. A. & Leskovec, J. Friendship and mobility: User movement in location-based social networks. in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 1082–1090 (2011). doi:<https://doi.org/10.1145/2020408.2020579>
18. Migrant workers and their children. *China Labour Bulletin* (2020). Available at: <https://clb.org.hk/content/migrant-workers-and-their-children>. (Accessed: 1st February 2020)
19. EpiRisk. *GLEAM* (2020). Available at: <https://epirisk.net/>. (Accessed: 9th February 2020)
20. Wang, Y. *et al.* The basic situation and expertise development of infection control facilities in sub-provincial cities. *Chin. J. Public Health Manag.* **28**, 741–743 (2012).
21. Wuhan Municipal Health Commission on the current situation of pneumonia in our city. (2020). Available at: https://wjw.hubei.gov.cn/fbjd/dtyw/201912/t20191231_1822343.shtml. (Accessed: 9th May 2020)
22. Granell, C., Gómez, S. & Arenas, A. Competing spreading processes on multiplex networks: Awareness and epidemics. *Phys. Rev. E Stat. Nonlinear Soft Matter Phys.* **90**, 012808 (2014).
23. Venables, W. N. & Ripley, B. D. *Modern Applied Statistics with S-PLUS* (Springer, Berlin, 2013) <https://doi.org/10.2307/2685660>.
24. Sakamoto, Y., Ishiguro, M. & Kitagawa, G. *Akaike Information Criterion Statistics* (Springer, Netherlands, 1986).
25. Real-time tracking: novel coronavirus statistics. *Tencent* (2020). Available at: <https://news.qq.com/zt2020/page/feiyan.htm>. (Accessed: 1st February 2020)

Acknowledgments

We thank Dr. Ross Sparks from Data61, CSIRO for helpful suggestions and advice on the methodologies. We also thank the anonymous reviewers for their valuable suggestions.

Author contributions

H.C., C.P. and A.R. conceived of the research question. H.C. processed the Baidu Search Index data. H.C. collected the other meta data including SARS cases and GDP per capita. H.C., A.R. and C.P. interpreted the result. H.C., A.R. and C.P. drafted the manuscript and compiled supplementary information. All authors edited the manuscript and supplementary information and aided in concept development.

Funding

H.C., C.P. and A.R. were supported by the CSIRO, an Australian federal government agency responsible for scientific research.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-75318-9>.

Correspondence and requests for materials should be addressed to H.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© Crown 2020