medRxiv preprint doi: https://doi.org/10.1101/2020.02.16.20023820; this version posted February 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Fractal kinetics of COVID-19 pandemic

Anna L. Ziff* Robert M. Ziff†

February 17, 2020

Abstract

The novel coronavirus (COVID-19) continues to grow rapidly in China and is spreading in other parts of the world. The classic epidemiological approach in studying this growth is to quantify a reproduction number and infection time, and this is the approach followed by many studies on the epidemiology of this disease. However, this assumption leads to exponential growth, and while the growth rate is high, it is not following exponential behavior. One approach that is being used is to simply keep adjusting the reproduction number to match the dynamics. Other approaches use rate equations such as the SEIR and logistical models. Here we show that the current growth closely follows power-law kinetics, indicative of an underlying fractal or small-world network of connections between susceptible and infected individuals. Positive deviations from this growth law might indicate either a failure of the current containment efforts while negative deviations might indicate the beginnings of the end of the pandemic. We cannot predict the ultimate extent of the pandemic but can get an estimate of the growth of the disease.

The number of cases of novel coronavirus 2019 n-CoV (which has recently been renamed as COVID-19) that started in China less than three months ago has been growing rapidly, pointing to a high infectivity. The mortality rate seems to fall somewhere between the common flu (0.1%) and the SARs coronavirus (10%). In Figs. 1a and 1b, we plot the total number of confirmed cases and total number of reported deaths, as a function of the number of days beginning with January 21, 2020, when the WHO first started reporting on this pandemic in its Situation Reports [World Health Organization, 2020]. Both cases and deaths continue to grow at an alarming rate.

In standard epidemiological analysis, one assumes that the number of cases in diseases like this one grows exponentially, based upon the idea of a fixed reproduction rate. If each person infects n other people where $n > 1$ (the reproduction number), then the total number of cases should grow as $n^{t/\tau} = e^{at}$, where τ is the incubation time, which depends on the characteristics of the particular disease. Several studies of COVID-19 have taken this approach to model the number of cases [e.g., Zhao et al., 2020]. This assumes there is no inhibition due to the interaction with already infected people, quarantine, or other prophylactic measures.

While the data display large growth, they do not in fact follow exponential behavior. In the inset in Fig. 2 we show a log-normal plot of the mortality data, and the behavior is clearly not linear as it would be

*Department of Economics, Duke University, Durham, NC 27708-0097 USA

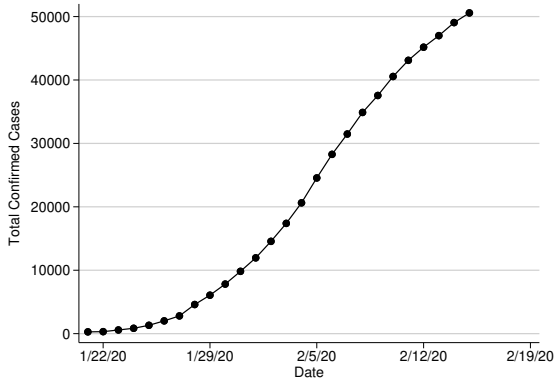
†Center for the Study of Complex Systems and Department of Chemical Engineering, University of Michigan, Ann Arbor, MI 48109-2136 USA

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

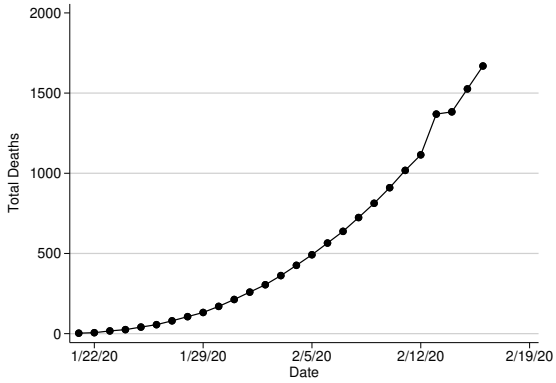
medRxiv preprint doi: https://doi.org/10.1101/2020.02.16.20023820; this version posted February 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Figure 1: Raw Trends in COVID-19 Cases and Deaths

(a) Total Confirmed Cases



(b) Total Deaths



Note: Data from World Health Organization [2020].

for exponential growth. In contrast, we find that the data are very well fit by assuming a power-law behavior with an exponent somewhat greater than two, as shown in the log-log plot in Fig. 2. The non-integer value of this power-law suggests a fractal type of behavior of the susceptibility mechanism, as we discuss below.

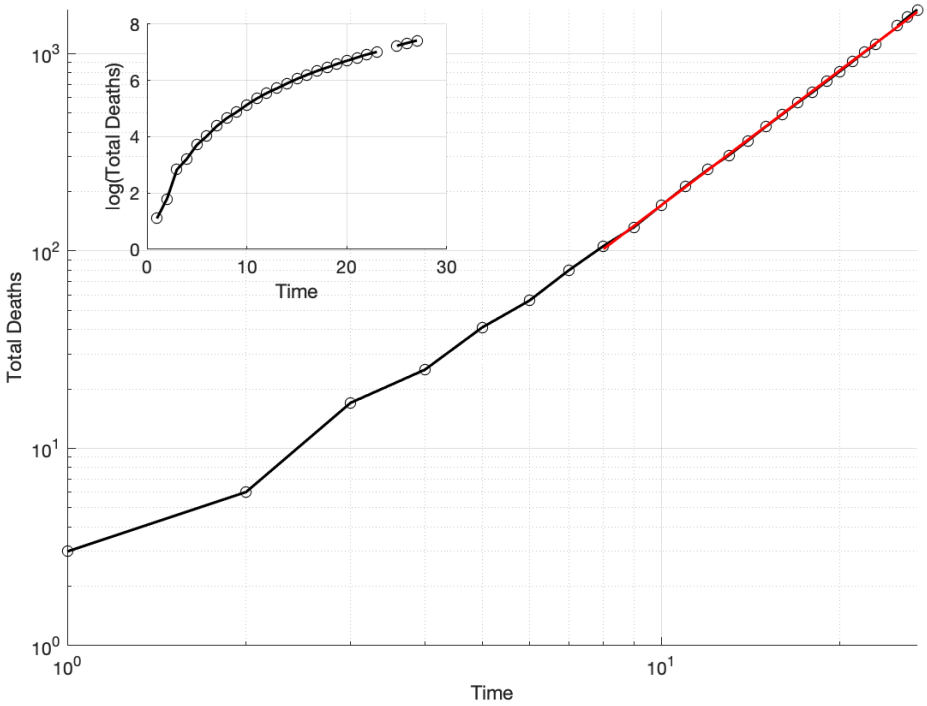
First we notice that the total number of deaths in Fig. 1b appear to be growing in a roughly quadratic form (see Table 1 in Appendix A for an estimate of a quadratic model), while the number of confirmed cases shown in Fig. 1a appears to follow similar behavior but is slowing down in the last few data points. There could be measurement error due to multiple factors.[1] Starting on February 7, those individuals who test positive for COVID-19 but show no symptoms are no longer being reported as confirmed cases [Cohen, 2020]. We focus our analysis on deaths to circumvent some of the ambiguities of reporting confirmed cases, assuming that deaths are a more reliable indication of the extent of the disease. (However, there are reports that some deaths might be reported as due to other causes, such as severe pneumonia, and not attributed to this virus, so these numbers may also be inaccurate. Still, one set of data should give a good indication of the general trend of the pandemic.)

Growing as t^2 is what one would expect if a disease were growing in a population at the periphery of a compact region of infection. We believe that this is more likely than an exponential growth, because the number of susceptible individuals around an infected individual decays with time. Individuals already infected might face increased immunity and there are other individuals who might have had a mild infection that imparts immunity but without developing symptoms warranting testing. These various effects would inhibit the exponential growth of the virus.

[1]Some of these factors includes misrepresentation of the number of confirmed cases, and varying the criteria used for the identification of the disease. These have been discussed in news reporting [Wang, 2020, BBC, 2020]. This analysis does not aim to quantify any intentional bias in the reported data.

medRxiv preprint doi: https://doi.org/10.1101/2020.02.16.20023820; this version posted February 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Figure 2: Total Deaths vs. Time (days since 1/21/2020) on a log-log plot (main plot) and log-normal plot (inset)



Note: The red line in the large plot corresponds to the power-law fit for periods 8–27, omitting 24, i.e., 1/28/2020 to 2/15/2020, omitting spurious data for 2/13/2020. Data from World Health Organization [2020].

medRxiv preprint doi: https://doi.org/10.1101/2020.02.16.20023820; this version posted February 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

We find that a power-law provides a better fit to the data than a simple quadratic, as postulated by Brandenburg [2020]. In Fig. 2 we plot the number of total deaths vs. time on a log-log plot. As Fig. 3 shows, considering all of the time periods (excluding report #24) from January 21 to February 15 (1–27 on the horizontal axis), the estimated coefficient is statistically indistinguishable from 2. However, removing just the first period period (2–27 on the horizontal axis), the estimated coefficient is statistically larger than 2. As more periods are removed, the coefficient moves farther away from 2. Even at this preliminary stage with few data points, this illustration suggests that a power-law fit is more appropriate than a quadratic one.

For illustration purposes, consider periods 8–27. The estimated coefficient is approximately 2.27 (the bootstrapped standard error with 5,000 replications is 0.014). This would imply that as the number of days of the pandemic increases by a factor of $2^{1/2.27} = 1.357$, the number of cases will double. Thus, starting from day 27 (Feb. 16) with 1,669 deaths, the number of deaths will double to approximately 3,340 on day 37 (Feb. 22), and then double again to 6,680 on day 50 (March 10), and double again to 13,350 on day 67 (March 27). This is a large number of fatalities, but not nearly as large as if the growth were exponential. On the other hand, it is larger than some epidemiological models have been predicting.

The appearance of a fractional exponent and power-law behavior suggests an underlying fractal process. Based upon a kind of small-world interaction network where individuals have many local neighbors and occasional long-range connections (such as caused by people traveling on trains, boats and planes). The results imply that this network has an effective minimum path fractal dimension of 2.25. The implication is that the disease produces flares in which the growth is momentarily exponential, but then slows down until another flare-up. The averaged effect of this growth apparently yields a power-law.

Wu et al. [2020] suggest that if the pandemic spreads into new cities, exponential growth like what happened in Wuhan in the early stages would occur. Hopefully, if this were the case, it would be quickly stopped and slower behavior would ensue. But our prediction of approximately 13,350 deaths by March 27 relies on the assumption that a new major outbreak will not happen. Hermanowicz [2020] and Roosa et al. [2020] have recently modeled the disease using a logistical and other growth models. These models are able to predict the ultimate number of cases, and find results in the range of 20,000 to 60,000. We believe the numbers will be much greater, and expect a number of *deaths* in this range, if the fractal growth continues as it has for a few more months.

We have followed our prediction for several days and have found no marked change in behavior, except for the data from WHO Situation Report #24, which gave an outsized jump of 254 cases to a total of 1,369 deaths. Reports stated that the data from China were incorrect that day because of double-counting of the results, but so far the WHO has not revised report #24. Looking at the large deviation from our prediction in fact led us to suspect that that data point was an outlier at the time.

medRxiv preprint doi: https://doi.org/10.1101/2020.02.16.20023820; this version posted February 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

In recent days there has been a drop-off in the growth of new confirmed cases from China, suggesting an attenuation of the pandemic. We do not see that drop-off in the death rate. There are two possible explanations for this discrepancy: (1) the number of confirmed cases may be showing the variability as it has in the past, and the current numbers are not capturing the actual situation, or (2) the pandemic is indeed slowing down but it will take a while (perhaps a week) for that to show up in the fatality data. Of course, we hope the latter is the case, but are not entirely confident considering the history of the reporting problems concerning this pandemic.

Deviations above this power-law behavior might indicate that the pandemic is expanding from the current level of control, while deviations below might indicate that the disease is starting to fade. This analysis is early in the outbreak of COVID-19 and cannot predict the length of the outbreak nor the final fatality. However, it suggests that under current conditions, power-law growth better approximates the number of deaths than exponential growth, and may provide a tool to monitor whether a fundamental change is occurring. Hopefully, continued vigilance, timely testing, and careful care will end this scourge on humanity.

The authors report no funding related to this research and have no conflicting financial interests.

medRxiv preprint doi: https://doi.org/10.1101/2020.02.16.20023820; this version posted February 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

References

BBC. Coronavirus: "Way too early" to predict end of outbreak, WHO says, February 2020. URL https://www.bbc.com/news/world-asia-china-51482994.

Axel Brandenburg. Quadratic growth during the 2019 novel coronavirus epidemic. *arXiv:2002.03638 [q-bio]*, February 2020. URL http://arxiv.org/abs/2002.03638. arXiv: 2002.03638.

Jon Cohen. Scientists are racing to model the next moves of a coronavirus that's still hard to predict, February 2020. URL https://www.sciencemag.org/news/2020/02/scientists-are-racing-model-next-moves-coronavirus-thats-still-hard-predict.

Bradley Efron and Robert J. Tibshirani. *An Introduction to the Bootstrap*. CRC Press, 1994.

Slav W Hermanowicz. Forecasting the Wuhan coronavirus (2019-nCoV) epidemics using a simple (simplistic) model - update (Feb. 8, 2020). *Epidemiology*, In Press, February 2020. doi: 10.1101/2020.02.04.20020461. URL http://medrxiv.org/lookup/doi/10.1101/2020.02.04.20020461.

Joel L. Horowitz. The Bootstrap. In *Handbook of Econometrics*, volume 5, pages 3159–3228. Elsevier, 2001. ISBN 978-0-444-82340-3. doi: 10.1016/S1573-4412(01)05005-X. URL https://linkinghub.elsevier.com/retrieve/pii/S157344120105005X.

K. Roosa, Y. Lee, R. Luo, A. Kirpich, R. Rothenberg, J.M. Hyman, P. Yan, and G. Chowell. Real-time forecasts of the 2019-nCoV epidemic in China from February 5th to February 24th, 2020. *Infectious Disease Modelling*, February 2020. ISSN 24680427. doi: 10.1016/j.idm.2020.02.002. URL https://linkinghub.elsevier.com/retrieve/pii/S2468042720300051.

Vivan Wang. How Many Coronavirus Cases in China? Officials Tweak the Answer, February 2020. URL https://www.nytimes.com/2020/02/12/world/asia/china-coronavirus-cases.html.

World Health Organization. Novel Coronavirus (2019-nCoV) situation reports. Technical Report 1-24, WHO, January 2020. URL https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports.

Joseph T. Wu, Kathy Leung, and Gabriel M Leung. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *The Lancet*, In Press, Jan 2020. ISSN 01406736. doi: 10.1016/S0140-6736(20)30260-9. URL https://linkinghub.elsevier.com/retrieve/pii/S0140673620302609.

Shi Zhao, Qianyin Lin, Jinjun Ran, Salihu S Musa, Guangpu Yang, Weiming Wang, Yijun Lou, Daozhou Gao, Lin Yang, Daihai He, and Maggie H Wang. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. *International Journal of Infectious Diseases*, In Press, January 2020. ISSN 1201-9712. doi: 10.1016/j.ijid.2020.01.050. URL http://www.sciencedirect.com/science/article/pii/S1201971220300539.

A Statistical Testing

Table 1 lists the estimated coefficients from the regression Deaths $= \beta_0 + \beta_1 \text{time} + \beta_2 time^2 + \varepsilon$, where ε is assumed to be time-independent and Normally distributed. We also report bootstrapped standard errors in brackets, which are more conservative in this small-sample context. See Efron and Tibshirani [1994] and Horowitz [2001] for introductions to the bootstrap method.

In Fig. 3, the confidence intervals are calculated as follows. First, robust standard errors are calculated to account for heteroskedasticity. These standard errors are more conservative, which is advantageous considering our small sample size. Next, we calculate the critical value (the 95th quantile) from a t-distribution with $n-2$ degrees of freedom. The value n changes with each point along the x-axis. As fewer periods are

medRxiv preprint doi: https://doi.org/10.1101/2020.02.16.20023820; this version posted February 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

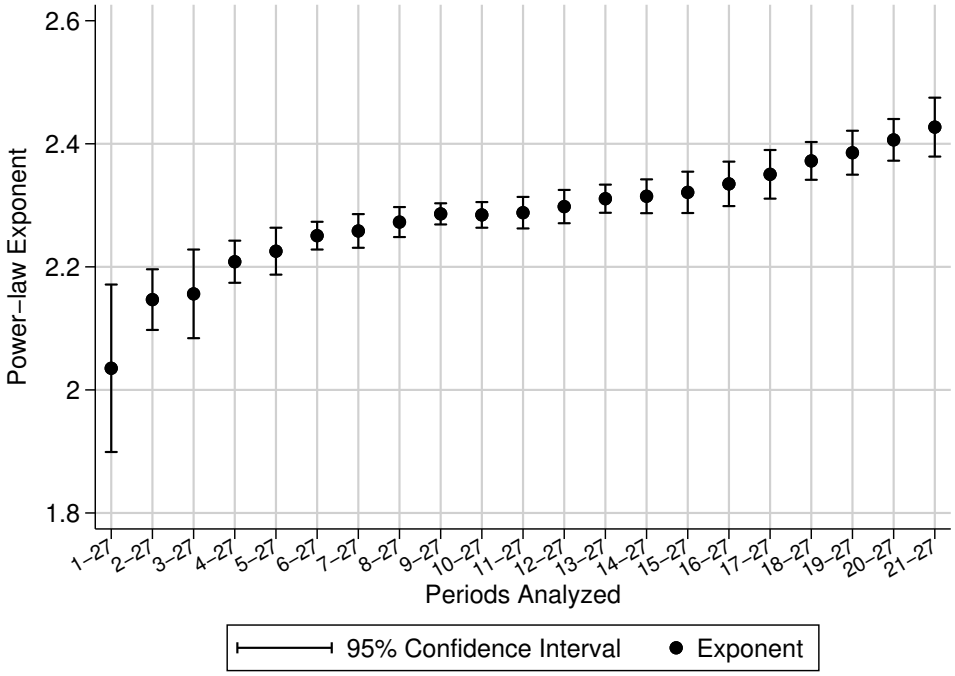
Table 1: Estimates of a Quadratic Model of Deaths

Time	-16.64***
	(2.060)
	[2.621]
Time^2	2.799***
	(0.0765)
	[0.0996]
Constant	44.48**
	(12.32)
	[15.640]
Observations	26
R^2	0.999

Note: Parametric standard errors are in parentheses, bootstrapped standard errors are in square brackets. We perform 5,000 replications. Data from World Health Organization [2020]. *($p < 0.05$),** ($p < 0.01$),*** ($p < 0.001$).

considered, the critical value increases. The degrees of freedom is a function of this n, but subtracting two to account for the fact that two parameters are estimated (β_0, β_1). The standard errors are quite small for these calculations. Considering periods 2-27 produces a bootstrapped standard error of 0.033, using 5,000 bootstrap replications.

Figure 3: Power-law Coefficients, Dropping Earlier Periods

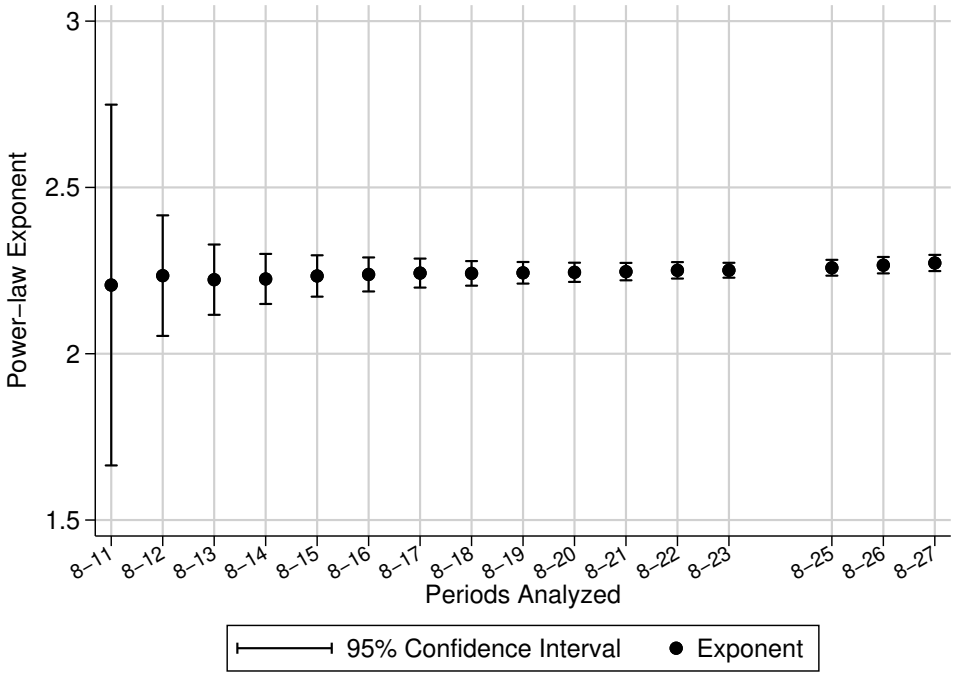


Note: The coefficients are the ordinary least squares estimates of β_1 in the regression $\log(\text{Deaths}) = \beta_0 + \beta_1 \log(\text{time}) + \varepsilon$ where ε is unobserved and assumed to be time-independent and distributed according to a t distribution with $n - 2$ degrees of freedom. Robust standard errors are calculated for suggestive inference, considering the small sample size. Data from World Health Organization [2020].

medRxiv preprint doi: https://doi.org/10.1101/2020.02.16.20023820; this version posted February 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Fig. 4 shows the estimated power-law exponents starting at period 8 and adding more periods forward. This mirrors our process of testing this model as more data were published on the number of deaths. The coefficient is stable. The confidence intervals are calculated analogously to those in Fig. 3. Note that data at period 24 are omitted due to probable inaccuracy, as we discuss above.

Figure 4: Power-law Coefficients, Dropping Later Periods



Note: The coefficients are the ordinary least squares estimates of β_1 in the regression $\log(\text{Deaths}) = \beta_0 + \beta_1 \log(\text{time}) + \varepsilon$ where ε is unobserved and assumed to be time-independent and distributed according to a t distribution with $n-2$ degrees of freedom. Robust standard errors are calculated for suggestive inference, considering the small sample size. Data from World Health Organization [2020].

The replication files include a `.csv` file of the compiled reports from World Health Organization [2020]. These data are publicly available, but we condense the information used in this report. A `.do` file contains the STATA code to clean the data, perform the statistical analysis, and produce the plots. A machine-specific environment variable needs to be set with the accompanying file structure. Once this is set, the code will run seamlessly on any machine with STATA installed. The Matlab file is only used to create a graph.