SCIENCE CHINA
Information Sciences

• MOOP •

June 2020, Vol. 63 164102:1–164102:3
https://doi.org/10.1007/s11432-020-2828-1

Visualization of COVID-19 spread based on spread and extinction indexes

Song-Hai ZHANG[1,2*], Yun CAI[1] & Jian LI[1]

[1] *Department of Computer Science and Technology, Tsinghua University, Beijing* 100084*, China;*
[2] *Beijing National Research Center for Information Science and Technology* (*BNRist*)*, Beijing* 100084*, China*

Received 29 February 2020/Accepted 12 March 2020/Published online 13 May 2020

Citation Zhang S-H, Cai Y, Li J. Visualization of COVID-19 spread based on spread and extinction indexes. Sci China Inf Sci, 2020, 63(6): 164102, https://doi.org/10.1007/s11432-020-2828-1

From December 2019, Coronavirus disease 2019 (COVID-19) rapidly spread from the city of Wuhan (Hubei Province) to all of China, which was exacerbated by mass population movements associated with the Chinese Spring Festival. To date, more than 80000 people have been infected, and more than 3000 have died. In addition, COVID-19 has already spread to 142 countries, such as Korea, Iran, and Italy. We have created a system to analyze and visualize the disease data in Hubei Province and the rest of China, to help decision makers and the public make rational decisions without requiring data analysis expertise.

The influence of COVID-19 has already exceeded that of severe acute respiratory syndrome (SARS) in 2003. Its spread is rapid and extremely unbalanced, spanning four orders of magnitude in the number of infections in different cities. Classical epidemiological modeling is absolutely important [1], however, the spread of COVID-19 influenced by several factors such models do not necessarily consider, such as travel from Wuhan to other cities for the Spring Festival, airborne transmission without contact, and government blockade policies. These influences make it quite difficult to predict the spread of COVID-19 because a single chance event can result in significant consequences.

The proposed system is primarily based on official daily pandemic data released by the National Health Commission of China and the Health Commission of Hubei. These data include the total number of and daily increase in confirmed cases, the number of cured patients, and the number of fatalities. We also consider basic statistics from the National Bureau of Statistics and the Hubei Provincial Bureau of Statistics, including resident population, and the rural and urban population density of each province in China and each city in Hubei Province. We have published an daily analysis and visualization report on the official account of Tencent_UR[1)] since February 7 with over 20 million page views in total by the end of February. These reports play an important role in helping the public rationally understand the pandemic situation and persuading the public to actively cooperate with the government in the fight against COVID-19.

However, it is non-trivial to design an intuitive visualization system that allows unskilled individuals quickly understand the key concepts. Professional visualization methods for high-dimensional data, e.g., parallel coordinate plots and calendar based approaches [2], are not intuitive for the inexperienced. Basic statistical charts and visualization methods with up to four visual channels are preferable, and the need for user interaction should be avoided. Therefore, we define spread index and extinction index that explicitly reveal the pandemic situation from the raw data, which will be introduced later. Beyond that, the system must also cope with high dynamic range data varying over four orders of magnitude.

[*] Corresponding author (email: shz@tsinghua.edu.cn)
[1)] https://mp.weixin.qq.com/s/Hfjl4wkWKIlLdgIvgKXHkg.

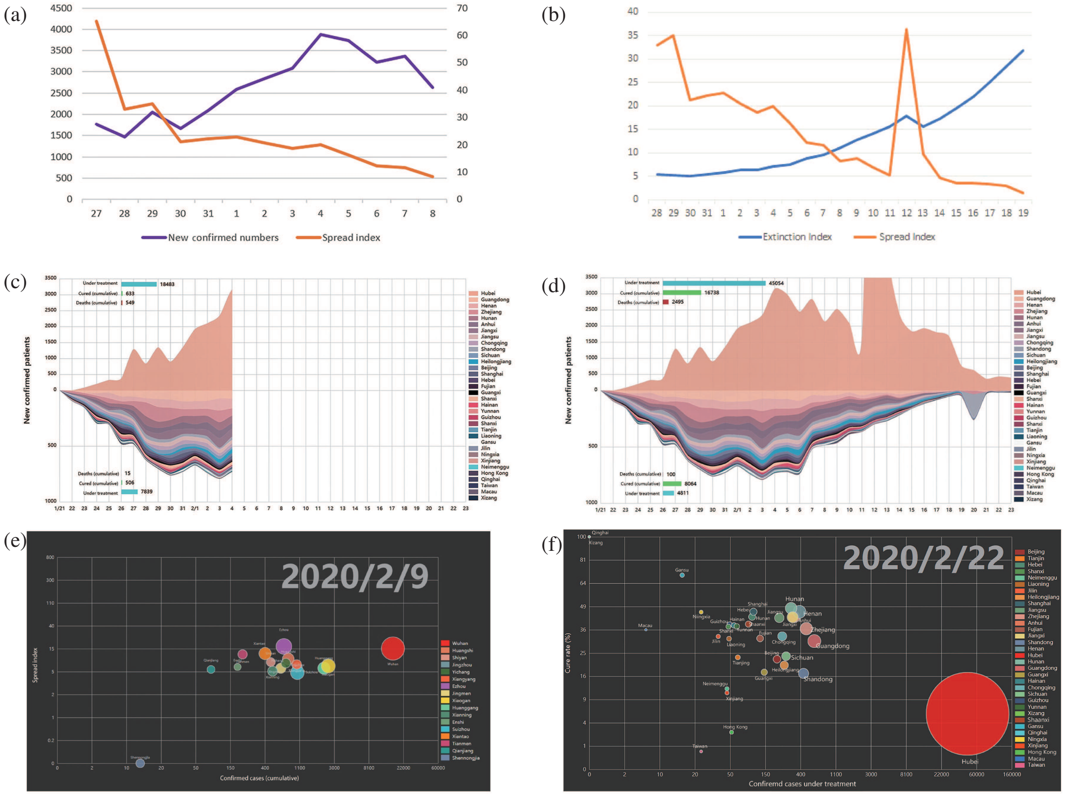

Figure 1 (Color online) (a) Line chart showing the number of confirmed new infections each day, and the spread index; (b) line chart for spread index and extinction index; (c) and (d) ThemeRivers for new confirmed infections by February 4 and 23, 2020, respectively; (e) bubble chart showing the pandemic situation for all cities in Hubei; (f) bubble chart showing the pandemic situation for all provinces in China.

Spread index and extinction index. Here, we define several quantities related to the pandemic data. In the early part of the spread of an pandemic, a model [1] of the number of people infected $I(t)$ follows exponential growth. We express the model as $I(t) = ba^t$ in a local domain, and then derive the spread index $F_s = I'/I = \ln(a)$ which expresses the spreading speed of the pandemic. Here, $I_S(i)$ denotes the total number of confirmed cases, $I_t(i)$ denotes the number of confirmed cases under treatment, $I(i)$ denotes the number of new confirmed cases, and $C(i)$ and $D(i)$ denote the increases in the number of people cured and the fatalities on day i, respectively. The spread index $F_s(i)$ can be computed easily as $F_s(i) = I(i)/I_t(i-1)$. We can also investigate when the pandemic declines, which occurs when the daily increase in confirmed cases becomes less than the number of people cured and the number of fatalities. There, we can define the extinction index $F_e(i)$ for day i as $F_e(i) = (C(i) + D(i))/I_t(i-1)$. Figure 1(b) shows the spread and extinction indexes.

Visualization method. To help non-experts understand the pandemic situation quickly without requiring user interactions, in addition to the daily line chart, we use bubble chart and ThemeRiver animations to visualize how the pandemic is changing over time [3,4].

(1) ThemeRiver. As mentioned previously, the distribution of COVID-19 is extremely unbalanced, with 83% of infections in China being in Hubei Province, and Wuhan city itself accounting for 60% of all infections in China by February 23. Note that ThemeRiver [5] cannot handle such data well, thus, we have redesigned ThemeRiver by plotting Hubei and all other provinces in opposite directions along a horizon line. Here, the horizontal axis represents time, and the vertical axis represents the number of new confirmed patients. In the process, the stacking order is important for ThemeRiver. From the definition of $I_s(i)$, we derive $I_s(i) = \sum_{j=1}^{i} I(j)$. Thus, the area of each branch in the ThemeRiver is the total number of confirmed cases $I_s(i)$. The curve on the bottom of each branch k is the sum of the heights of branches above it, and its own height. Here, let $I^k(i)$ denote the number of new confirmed patients for the kth branch in the ThemeRiver stacking order. Thus, the curve position $P^k(i)$ of each branch k on day i of ThemeRiver can be computed as $P^k(i) = P^{k-1}(i) + I^k(i) = \sum_{j=1}^{k} I^j(i)$. In

an interaction-less animation, we sort all branches in descending order by their total number of confirmed cases; thus, users can focus on the current situation and readily observe changes in the most severely affected provinces.

A bar chart of the number of confirmed cases in treatment, cured and fatalities is also presented and synchronized with ThemeRiver. Figures 1(c) and (d) show the ThemeRiver and its corresponding bar chart on February 4 and 22, respectively.

(2) Bubble chart. To reduce user cognitive burden, we employ bubble chart animations with only four visual channels, i.e., the horizontal axis, vertical axis, color, and bubble area. Note that the color always encodes labels. To handle high dynamic range data of up to four orders of magnitude, logarithmic coordinates are used for the two axes and bubble area. Figures 1(e) and (f) show bubble charts of the situation for all cities in Hubei Province and all provinces in China respectively. To facilitate analysis of the situation and find rules and abnormal cases, additional factors are calculated based on the raw situational data and combined into visualizations. The infection rate $F_r(i)$ is a very important factor for pandemic spread. It is defined as $F_r(i) = I_s(i)/P$, where P is the resident population of an area. We can also define the cured rate as $F_c(i) = \sum_{j=1}^{i} C(j)/I_s(i)$. Bubble charts are flexible, and there are different compositional approaches for the three visual channels to represent various indexes, which emphasize different aspects of analysis, comparing Figures 1(e) (X: F_s; Y: I; AREA: F_r) and (f) (X: I_t; Y: F_c; AREA: I_s).

Case study. Using this system, several important developments are identified. These cases demonstrate the usefulness of the proposed system.

(1) New confirmed cases increasing while spread index is decreasing. Prior to February 5, the number of new confirmed cases across China was increasing rapidly, which caused public anxiety. This situation is shown in Figure 1(a). However, the spread index curve was gradually decreasing, which demonstrates the pandemic was spreading at decreasing rates, despite the daily increase in confirmed cases. This report with the above analysis gained over 1 million page views the day it was published. Note that the daily increase in the number of confirmed cases has been decline since February 5.

(2) Balancing point of spread index and extinction index. As shown in Figure 1(b), the spread index curve F_s deceases, and the extinction index curve F_e is increasing over time. Note that the balancing point of $F_s = F_e$ was reached on February 7. Since February 7, the difference $F_e - F_s$ has become ever greater, with the expection of February 12. We know that $I_t(i) = I_t(i-1) + I(i) - C(i) - D(i) = I_t(i-1)(1+F_s(i)-F_e(i)) < I_t(i-1)$. In addition, since February 7, the number of confirmed cases in treatment has been decreasing.

(3) High risk of the city of E'zhou (Hubei Province) from the bubble chart. Figure 1(e) shows the situation for all cities in Hubei Province on February 10. Here, the bubble area represents the infection rate. This bubble chart clearly shows the high risk of pandemic in E'zhou with highest infection rate (except for Wuhan) and highest spread index, although the total number of confirmed cases was still low. Thus, we delivered an early warning for E'zhou in our report on February 11.

Summary. We have introduced a system to analyze and visualize the COVID-19 pandemic situation. Using this system, we have published daily reports with over 20 million page views in total. The comments from the public show that they played a positive role in helping the public understand the pandemic situation rationally, stabilizing public mood, and promoting the public to actively cooperate with the government in the fight against COVID-19.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 61521002, 61772298) and Tsinghua-Tencent Joint Laboratory for Internet Innovation Technology. We would like to thank Professor Shi-Min HU for his valuable suggestions, and thank Shao-Kui ZHANG and Wen-Yang ZHOU for their help in system development.

Supporting information Videos and other supplemental documents. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Zhou T, Liu Q, Yang Z, et al. Preliminary prediction of the basic reproduction number of the Wuhan novel coronavirus 2019-nCoV. 2020. ArXiv: 2001.10530

2 Wang D Q, Guo D H, Zhang H. Spatial temporal data visualization in emergency management: a view from data-driven decision. In: Proceedings of the 3rd ACM SIGSPATIAL Workshop on Emergency Management, 2017. 1–7

3 Dimara E, Perin C. What is interaction for data visualization? IEEE Trans Visual Comput Graph, 2020, 26: 119–129

4 Robertson G, Fernandez R, Fisher D, et al. Effectiveness of animation in trend visualization. IEEE Trans Visual Comput Graph, 2008, 14: 1325–1332

5 Kosara R. Presentation-oriented visualization techniques. IEEE Comput Grap Appl, 2016, 36: 80–85