



Short Communication

Predicting local COVID-19 outbreaks and infectious disease epidemics based on landscape network entropy

Rui Liu ^{a,b,1}, Jiayuan Zhong ^{a,1}, Renhao Hong ^a, Ely Chen ^c, Kazuyuki Aihara ^d, Pei Chen ^{a,*}, Luonan Chen ^{e,f,g,h,*}

^a School of Mathematics, South China University of Technology, Guangzhou 510640, China

^b Pazhou Lab, Guangzhou 510330, China

^c Stanford University, Stanford 94305, USA

^d International Research Center for Neurointelligence, The University of Tokyo, Tokyo 113-8654, Japan

^e State Key Laboratory of Cell Biology, Shanghai Institute of Biochemistry and Cell Biology, Center for Excellence in Molecular Cell Science, Chinese Academy of Sciences, Shanghai 200031, China

^f Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming 650223, China

^g Key Laboratory of Systems Biology, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Hangzhou 310024, China

^h School of Life Science and Technology, ShanghaiTech University, Shanghai 201210, China

ARTICLE INFO

Article history:

Received 19 August 2020

Received in revised form 14 December 2020

Accepted 15 March 2021

Available online 26 March 2021

© 2021 Science China Press. Published by Elsevier B.V. and Science China Press. All rights reserved.

As a highly infectious disease, the reproduction number of coronavirus disease 2019 (COVID-19) was estimated to be as high as 6.47 in the early stage [1]. Many studies have suggested that early interventions, such as the use of masks, social distancing, self-isolation, quarantine and even lockdown of entire regions and communities, are effective in containing or at least, mitigating the spread of the virus [2]. It is thus crucial to detect the early-warning signal of the COVID-19 outbreak so that a timely public health strategy can be carried out to reduce the magnitude and spread of the pandemic. However, the complex characteristics of both biological and social systems lead to the challenge of achieving the real-time prediction of infectious disease outbreaks. Furthermore, a surveillance system for such detection can be costly, thereby resulting in the failure to detect the potential progression of epidemics in many countries that lack adequate public health infrastructure [3,4]. Machine learning methods have been developed in the field of forecasting [5] but generally fail to predict outbreaks of infectious diseases when there are only limited samples. Statistical and mathematical models have been proposed to describe the transmission of the current COVID-19 epidemic [6], which helps clinicians understand its spread. However, unlike time series data prediction, an outbreak is a typical nonlinear event with characteristics that develop from gradual change to drastic transition, thus making the prediction of a COVID-19 outbreak rather difficult. Therefore, it is of great importance to develop an effective

model-free method for directly predicting such nonlinear events or for detecting early-warning signals of infectious disease outbreaks based on real-time data such as daily new cases.

The spread of COVID-19 is usually regarded as a time-dependent nonlinear dynamical system with a tipping point at which the system falls into a large-scale disease outbreak state through a critical transition [7]. There is a general consensus that the dynamic progression of the infectious disease can be described as three stages (Fig. S1 online) [8,9], i.e., a normal stage, a pre-outbreak stage and an outbreak stage, where the pre-outbreak stage is viewed as the tipping point just before the critical transition that leads to a large and possibly uncontrollable epidemic outbreak. At the pre-outbreak stage with only limited sporadic cases, disease transmission could be controlled with appropriate measures. However, if the disease spreads continuously with no or few containment measures, the pervasiveness of the virus and the massive number of cases pose insurmountable difficulties for governments, resulting in an irreversible local epidemic or even the global pandemic. Thus, it is crucial to detect early-warning signals of epidemic disease outbreaks at the pre-outbreak stage. Recently, we proposed the dynamic network marker (DNM) or dynamic network biomarker (DNB) concept [10,11], which is an extension of the critical slowing down theory to high-dimensional systems [12]. When the dynamical system approaches the critical point, the DNM theory suggests that a small group of the observed variables (DNM elements) will intensively fluctuate with high correlations and thus convey early-warning signals for the impending critical transition by several statistical indices; that is, the correlations between variables/elements in the DNM group

* Corresponding authors.

E-mail addresses: chenpei@scut.edu.cn (P. Chen), lnchen@sibs.ac.cn (L. Chen).

¹ These authors contributed equally to this work.

drastically increase; and the deviations of all the DNM elements rapidly increase in a strongly collective manner [13] (Supplementary materials Note S1 online).

In this study, on the basis of the DNM concept, a computational method, landscape network entropy (LNE), is proposed to predict epidemic disease outbreaks. Specifically, by exploring the dynamic information based on a combination of a district network and high-dimensional daily new case data, the proposed LNE method can quantitatively characterize the spread of an infectious disease in a district network and thus detect the early-warning signals of COVID-19 outbreaks. Unlike the time-series prediction or traditional detection of the outbreak stage, the LNE method aims to identify the pre-outbreak stage or critical stage that generally has no clear abnormalities but with high potential of critical transition into the catastrophic/nonlinear event in the near future. In particular, in contrast to traditional methods, a major feature of this work is to exploit high-dimensional dynamic information in the network collectively to provide reliable prediction.

LNE is a model-free method based only on the topological structure of the district network and daily new case data. In a country/region, the geographic distribution of districts and their adjacent information are modeled as a network, in which each node represents a district. Specifically, based on geographical and transportation information, there is an edge between two neighboring districts in such a network, representing their adjacency and interactive relations. Practically, we construct the district network for a region according to the main transportation form and the specific geographical proximity among districts (Fig. S2 online). For example, in the Kanto region of Japan, the main transportation of travellers is by intercity railways such as Shinkansen, Japan Railways (JR) lines and metro systems, which serve as the main public transportation system in Kanto with approximately 3.16 billion annual passenger rides and a daily ridership of 6.31 million people. Therefore, the district network of Kanto is constructed based on geographical proximity, with edges representing the main railways that connect two adjacent districts. The main mode of transportation in the other 4 regions (i.e., Hubei Province of China, South Korea, mainland Italy, and Western Europe) is similar, implying that geographically adjacent networks are suitable. However, among the 17 US states with severe COVID-19 epidemics, the main interstate transportation is by air. A 17-node complete graph/fully-connected network in which each pair of states is connected by an edge/airline is more suitable for modeling the main mode of interstate transportation (Fig. S2 online). A given district network is partitioned into many local networks. Each local network is composed of a central node/district and all its first-order neighbors. Thus, for a local network N^k with $(L + 1)$ members, i.e., a central node k and L first-order neighbors k_i ($i = 1, 2, \dots, L$) (Fig. S3 online), its local LNE at time point t can be calculated in terms of the product of entropy and deviation as follows:

$$H_k(t) = -\frac{|\text{SD}(\vec{\mathbf{k}}(t)) - \text{SD}(\vec{\mathbf{k}}(t-1))|}{\log(L)} \sum_{i=1}^L p_i(t) \log(p_i(t)), \quad (1)$$

with

$$p_i(t) = \frac{|\text{PCC}(\vec{\mathbf{k}}_i(t), \vec{\mathbf{k}}(t))|}{\sum_{j=1}^L |\text{PCC}(\vec{\mathbf{k}}_j(t), \vec{\mathbf{k}}(t))|}, \quad (2)$$

where vector $\vec{\mathbf{k}}(t) = (c^k(t-M+1), \dots, c^k(t-1), c^k(t))$ with $c^k(t)$ representing the number of daily new cases reported in district k at time point t , and M representing the length of the sliding window. $\text{PCC}(\vec{\mathbf{k}}_i(t), \vec{\mathbf{k}}(t))$ is the Pearson's correlation coefficient (PCC) between the central node k and its first-order neighbor k_i at sampling point t , and $\text{SD}(\vec{\mathbf{k}}(t))$ is the standard deviation of the number

of daily new cases for central node k at sampling point t . The specific choice of M is provided in Table S1 (online). On each day, the mean LNE of all the local networks was taken to quantitatively detect the early-warning signal of the critical transition from the normal stage to the outbreak stage in a region; that is, if the LNE index (the reciprocal of significance P -value of LNE) at time point t is above a significance threshold value, then time point t is considered to be in the pre-outbreak stage. The calculation details of LNE, and the specific choice of the threshold value are presented in Supplementary materials Note S2 (online).

As a model-free method of nonlinear event prediction, the LNE method was applied to the data of daily new cases in six regions, i.e., Hubei Province of China, the Kanto region of Japan, Western Europe, mainland Italy, South Korea, and 17 US states (Table S2 online). LNE detected the early warning signals of all six regional COVID-19 outbreaks from both global and local perspectives. First, the data of daily new COVID-19 cases were collected from March 2, 2020 to November 21, 2020 in ten districts around the Kanto region of Japan. A ten-node network was built based on the geographic adjacency information of the ten districts (Fig. 1a). As presented in Fig. 1b, the first warning point, marked by a fuchsia circle, appeared on March 21, indicating an upcoming critical transition into the outbreak of COVID-19 thereafter, and, in fact, there was an abrupt increase in daily new cases from March 26 (the blue curve in Fig. 1b). The signaling point was also backed up by the announcement that the government of Japan officially declared a nationwide state of emergency on April 7 in response to the surge of new COVID-19 cases. There was a signal on April 28, indicating a possible tipping point in the opposite direction that the first wave of the epidemic would soon be under control. After approximately two months of fighting against the epidemic, an end to Japan's COVID-19 state of emergency was formally declared by the government of Japan. However, 12 days later, another warning signal appeared on June 8, prior to the second wave of the COVID-19 epidemic, and prior to the emergency reports from the government, i.e., "Tokyo revised its COVID-19 monitoring system guidelines on June 30" and "Tokyo raises virus alert to the highest level on July 15 as infections resurge". Furthermore, after the government announced the reopening of borders on September 25 and resumed bilateral business travel from October 7, a third warning signal was detected on October 31, which was then validated by the third-wave surge of daily new cases. On November 19, the COVID-19 alert level in Tokyo was raised to the maximum again. In addition, the early-warning signals of the COVID-19 outbreak were also detected by the district-specific LNE index for the six districts in the Kanto region (Fig. 1c). For these districts, the LNE indices signal imminent disease outbreaks. For instance, in Tokyo, the first and second signaling points (the fuchsia circles) appeared on March 20 and June 8 earlier than those of the Kanto region, after which the daily new cases sharply increased in this city. During the epidemic, the dynamic change of the district network mapped with local LNE is presented in Fig. 1d. The district-specific LNE indices for all ten districts are shown in Fig. S4 (online).

For the 17 US states with severe epidemics, as shown in Fig. 1e, the first warning signal (the fuchsia circle mark) provided by LNE appeared on March 8, indicating the emergence of an upcoming COVID-19 outbreak thereafter. The state of emergency over COVID-19 was declared on March 13. The second warning signal was detected on May 28, before the second-wave of a significant increase in daily new cases. On June 22, a series of restrictions were issued on immigration to the United States, suspending most H-1B, H-2B, and H-4 visas. The third and fourth warning signals occurred on July 3 and October 16, which were before the events that Washington, D.C. issued a new mask mandate as cases rose sharply from July 22, and the US government issued a border-control agreement to help curb the spread of COVID-19 on October 19. The network

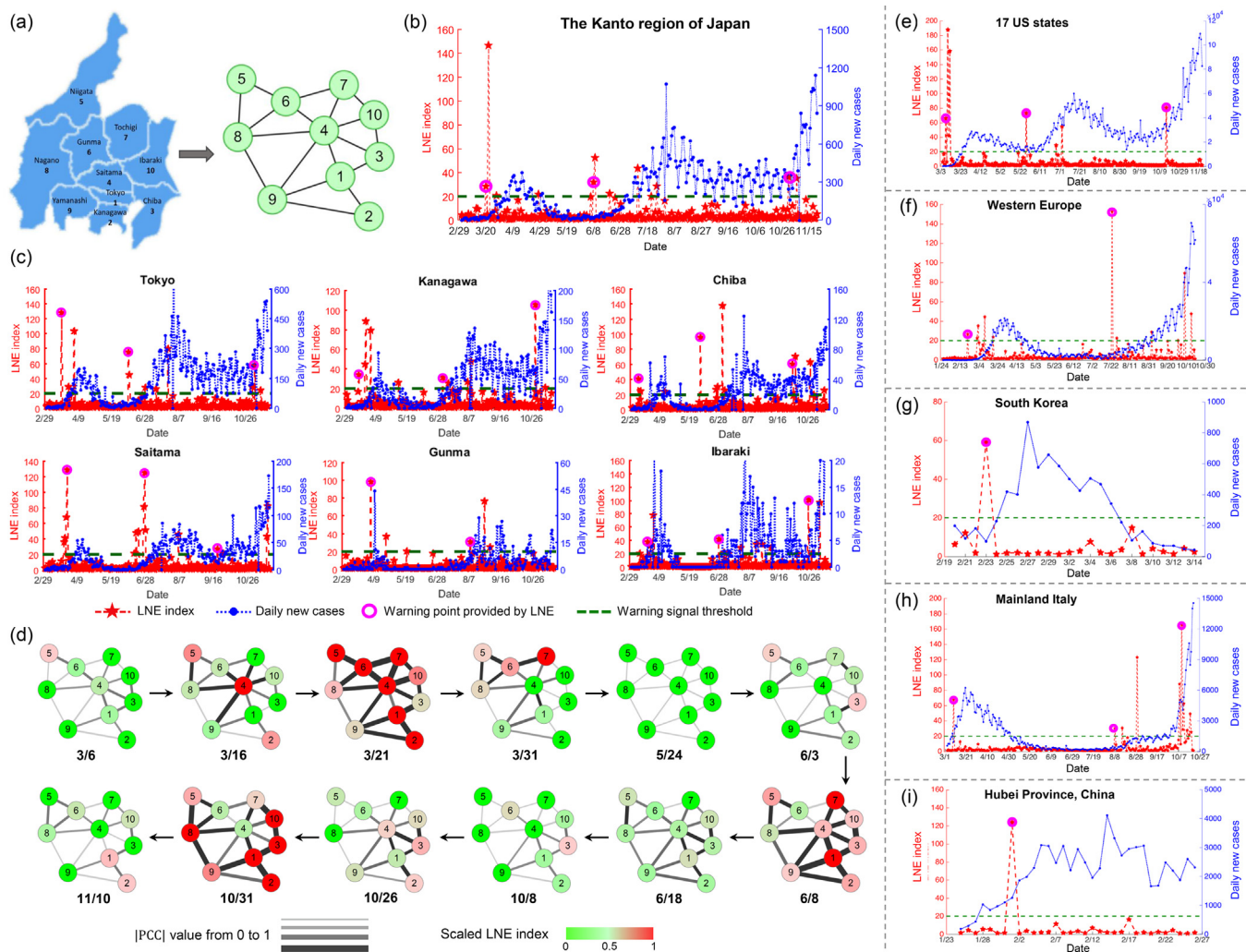


Fig. 1. Real-time monitoring of COVID-19 spreads and outbreaks in the Kanto region of Japan and other regions. (a) A ten-node network of the Kanto region. (b) The LNE index (the red curve) and the number of daily new cases (the blue curve) in the Kanto region. (c) The district-specific (local) LNE indices (the red curves) and the number of daily new cases (blue curves) for six districts. (d) The dynamic evolution of the district-network during the epidemic. (e–i) The results of other five regions. For all the LNE indices, the fuchsia circles represent the early warning points, which are ahead of the drastic increases in daily new cases. In the case of multiple warning points appearing within a short period of time, the first warning signal is marked.

and state-specific LNE indices for all 17 US states are shown in Fig. S5 (online). For Western Europe, the first warning signal appeared on February 21 (Fig. 1f). Indeed, the warning signal was prior to the sharp increase in daily new cases from March 3. Furthermore, the signal was also before the emergency event on March 2, when the European Commission announced the establishment of a “Coronavirus Response Team”, to coordinate the response to the spread of COVID-19. The second warning signal was detected on July 23 (Fig. 1f), predicting the second wave of the epidemic 12 days later. On August 4, Paris made mask-wearing compulsory for pedestrians outdoors in the city and extended the area several times over the following weeks. The local LNE indices for all 10 countries in Western Europe are shown in Fig. S6 (online). Furthermore, for the COVID-10 outbreaks in South Korea (Fig. 1g and Fig. S7 online), mainland Italy (Fig. 1h and Fig. S8 online), and Hubei Province, China (Fig. 1i and Fig. S9 online), the warning points provided by LNE are all ahead of the abrupt increase in the daily new cases. The warning signals were also backed up by the emergency events officially released by the governments. In addition, LNE predicted the surge of COVID-19 cases in the local areas of Heilongjiang Province (Fig. S10 online) and Shijiazhuang City (Fig. S11 online), both of which had typical sporadic bursts in China in January 2021. More details on predict-

ing the COVID-19 outbreaks are provided in [Supplementary materials Note S3](#) (online).

The LNE index can not only predict the COVID-19 outbreaks but also characterize the criticality of the dynamic spread of seasonal influenza. We collected historical longitudinal outpatient records of influenza from clinics distributed in 23 wards in Tokyo, Japan, between January 1, 2016 and March 31, 2020. To profile influenza transmission in the city, a 23-node network is constructed according to the actual locations of the 23 wards and their adjacency relationships (Fig. 2a). As shown in Fig. 2b, early-warning signals were detected by LNE for seasonal outbreaks of influenza from 2016 to 2020. For each influenza epidemic later developing into a massive outbreak, the LNE index is sensitive and increases significantly at least 3 weeks (i.e., the lead time) before the increase in outpatient counts, indicating the emergence of the critical transition into an epidemic. The dynamic evolution of the city network during an annual influenza outbreak is presented in Fig. 2c. More details of the prediction of influenza outbreaks are presented in [Supplementary materials Note S4](#) (online).

Based on the LNE method, a web tool is developed to facilitate the identification of the tipping point or the pre-outbreak stage during the dynamic process of disease spread (Fig. 2d, www.rp-computationalbiology.cn/monitor). Based on this web tool, we

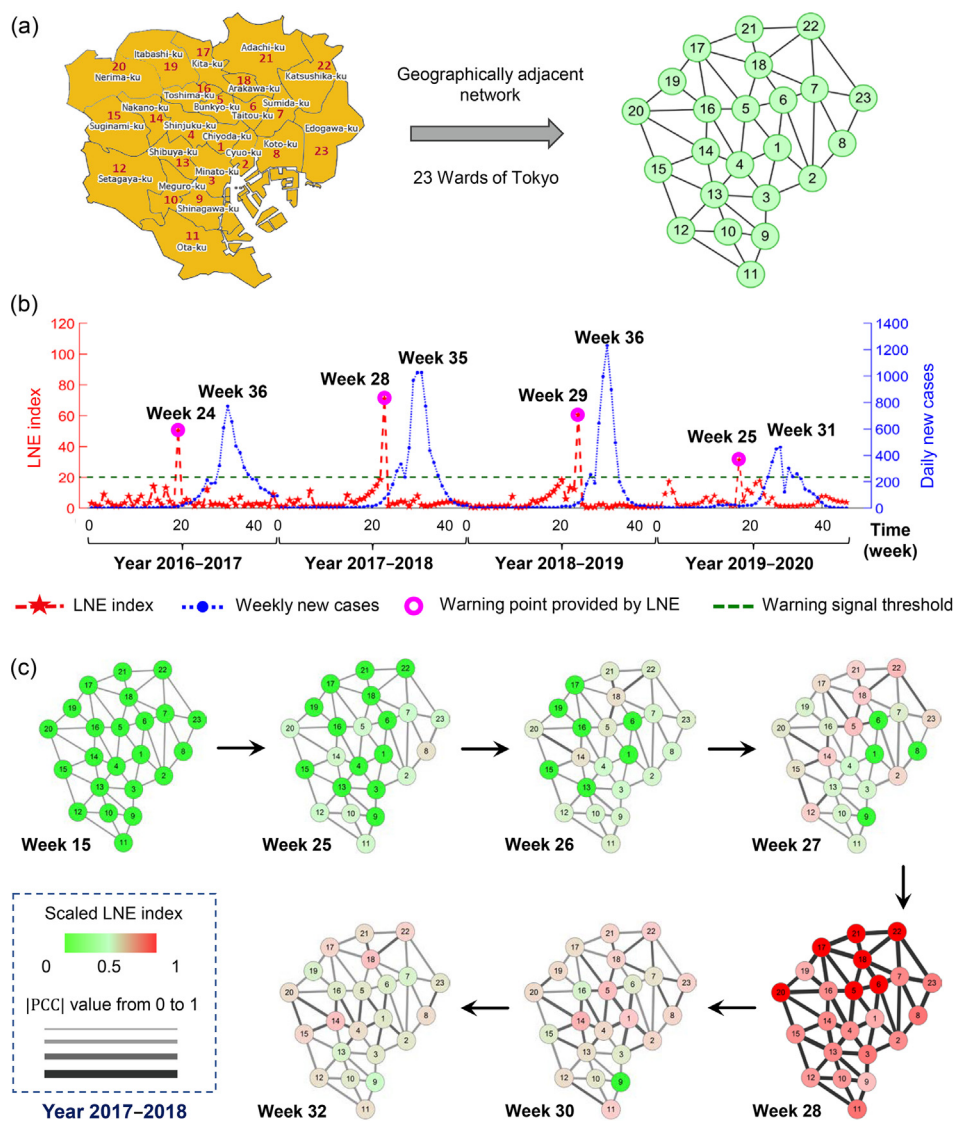


Fig. 2. Real-time monitoring of seasonal influenza outbreaks for 23 wards in Tokyo, Japan, between 2016 and 2020. (a) A 23-node city network of Tokyo. (b) The annual monitoring results of Tokyo for 2016–2017, 2017–2018, 2018–2019, and 2019–2020. The red curve represents the LNE index, and the blue curve shows the weekly number of clinic visits caused by influenza. The fuchsia circle represents the first signaling point, which occurs before the sharp increase in the number of clinic visits. The x-axis represents a period spanning from the 17th week (the first week in May) to the 60th or 61st week (the last week in March). The y-axis is the average number of weekly clinic visits. (c) The dynamic evolution of the city network during the influenza outbreak of 2017–2018. (d) An illustration of the web tool for the real-time monitoring of infectious disease outbreaks.

enabled the real-time monitoring function to detect early-warning signals to the outbreaks of infectious diseases, including COVID-19 and influenza, in seven regions. In each region, both the global and local LNE indices were provided. More details of the web tool were provided in [Supplementary materials Note S5](#) (online).

Epidemic models have provided rich prospective information on disease transmission. Through literature and information mining, we collected the basic reproduction number R_0 , which is an important and critical parameter in well-studied epidemic models, reflecting the transmission strength of a virus [14,15]. As shown in [Fig. S12](#) (online), the dynamic change in LNE and that of the basic reproduction number R_0 were compared. The signal point of LNE (i.e., the significant change point of LNE with significance, where $P < 0.05$) almost corresponds to the change in R_0 ; that is, there are signals from LNE when R_0 is approximately 1. In the early period in the United States, there was no obvious relation between the LNE signal and the R_0 values, which may be caused by the fact that

the number of reported cases was inaccurate because of the lack of test kits. Theoretically, the basic reproduction number R_0 is the dominant eigenvalue of the next-generation matrix [14] and thus $R_0 = 1$ corresponds to the bifurcation point of the nonlinear disease-transmission dynamical system [15], which can be quantified by the LNE. Therefore, the proposed LNE method can also provide early-warning signals for $R_0 = 1$ before $R_0 > 1$ or before the critical state transition from a stable fixed point ([Fig. S12](#) online). More details were shown in [Supplementary materials Note S6](#) (online).

To illustrate how the specific topological structures of the district networks affect the warning performance, we applied the LNE method to a different district-network structure for each region and show the results in [Fig. S13](#) (online). For example, no signal was detected for Hubei Province with a fully-connected network, while false positive signals were observed for the 17 US states under a geographically adjacent network structure (see

Supplementary materials Note S7 online for more details). Time-window entropy is also presented for the single-point outbreak scenarios without any geographical structure in Supplementary materials Note S8 and Fig. S14 (online).

In summary, the LNE method requires information only from geographic district networks and high-dimensional data of daily new cases. As a model-free and data-driven method, LNE requires neither feature selection nor a model/parameter training procedure. It performed well in predicting the COVID-19 and seasonal influenza outbreaks (Figs. 1, 2, and Table S2 online). The first signaling point predicted by the LNE index is ahead of an abrupt increase in the daily new cases, providing appropriate timing for implementing a proactive strategy to control the spread of infectious diseases. Some signaling points indicate the tipping point with the opposite direction at which the epidemic will soon be under control. The applications and analyses in this study have shown that by incorporating the high-dimensional dynamic information of district networks, LNE can accurately detect the pre-outbreak stages. Therefore, LNE is potentially useful in practical real-time monitoring of public health management strategies. The source codes are publicly available at https://github.com/zhongjiayuna/LNE_Project. Data and materials used in this study are presented in Supplementary materials Note S9 (online).

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgments

This work was supported by the National Key R&D Program of China (2017YFA0505500), National Natural Science Foundation of China (11771152, 12026608, 11901203, 31930022, and 31771476), Guangdong Basic and Applied Basic Research Foundation (2019B151502062, and 2021A1515012317), Strategic Priority Research Program of the Chinese Academy of Sciences (XDB38040400), Shanghai Municipal Science and Technology Major Project (2017SHZDZX01), Japan Society for the Promotion of Science KAKENHI (15H05707), Japan Science and Technology Agency Moonshot R&D (JPMJMS2021), Japan Agency for Medical Research and Development (JP20dm0307009), UTokyo Center for Integrative Science of Human Behavior (CiSHuB), and the International Research Center for Neurointelligence (WPI-IRCN) at The University of Tokyo Institutes for Advanced Study (UTIAS).

Author contributions

Luonan Chen, Rui Liu, Pei Chen, and Kazuyuki Aihara conceived and supervised the research. Rui Liu and Jiayuan Zhong designed the method. Jiayuan Zhong performed the formal analysis. Renhao Hong and Ely Chen performed data curation. Luonan Chen, Rui Liu, Pei Chen, and Kazuyuki Aihara wrote the original draft. Luonan Chen, Rui Liu, Pei Chen, Kazuyuki Aihara, Jiayuan Zhong, and Renhao Hong revised the manuscript.

Appendix A. Supplementary materials

Supplementary materials to this article can be found online at <https://doi.org/10.1016/j.scib.2021.03.022>.

References

- [1] Tang B, Wang X, Li Q, et al. Estimation of the transmission risk of the 2019-nCoV and its implication for public health interventions. *J Clin Med* 2020;9:462.
- [2] Hellewell J, Abbott S, Gimma A, et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *Lancet Glob Health* 2020;8:e488–96.
- [3] Wilson K, Brownstein JS. Early detection of disease outbreaks using the Internet. *Can Med Assoc J* 2009;180:829–31.
- [4] Huang J, Zhang L, Liu X, et al. Global prediction system for COVID-19 pandemic. *Sci Bull* 2020;65:1884–7.
- [5] Chen P, Liu R, Aihara K, et al. Autoreervoir computing for multistep ahead prediction based on the spatiotemporal information transformation. *Nat Commun* 2020;11:1–15.
- [6] Benvenuto D, Giovanetti M, Vassallo L, et al. Application of the arima model on the COVID-2019 epidemic dataset. *Data Brief* 2020:105340.
- [7] Drake JM, Brett TS, Chen S, et al. The statistics of epidemic transitions. *PLoS Comput Biol* 2019;15:e1006917.
- [8] Chen P, Li Y, Liu X, et al. Detecting the tipping points in a three-state model of complex diseases by temporal differential networks. *J Transl Med* 2017;15:217.
- [9] Chen P, Chen E, Chen L, et al. Detecting early-warning signals of influenza outbreak based on dynamic network marker. *J Cell Mol Med* 2019;23:395–404.
- [10] Chen L, Liu R, Liu ZP, et al. Detecting early-warning signals for sudden deterioration of complex diseases by dynamical network biomarkers. *Sci Rep* 2012;2:342–9.
- [11] Liu R, Chen P, Chen L. Single-sample landscape entropy reveals the imminent phase transition during disease progression. *Bioinformatics* 2020;36:1522–32.
- [12] van Nes E, Scheffer M. Slow recovery from perturbations as a generic indicator of a nearby catastrophic shift. *Am Nat* 2007;169:738–47.
- [13] Liu R, Wang J, Ukai M, et al. Hunt for the tipping point during endocrine resistance process in breast cancer by dynamic network biomarkers. *J Mol Cell Biol* 2019;11:649–64.
- [14] Hethcote HW. The mathematics of infectious diseases. *SIAM Rev* 2000;42:599–653.
- [15] van den Driessche P, Watmough J. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Math Biosci* 2002;180:29–48.



Rui Liu is a full professor at the School of Mathematics, South China University of Technology. He received his B.S. and Ph.D. degrees in Applied Mathematics from Peking University. His research interest includes nonlinear dynamics, modelling, and computational methods.



Jiayuan Zhong is a Ph.D. candidate at South China University of Technology. His research interest mainly focuses on developing computational approaches to detect the tipping points of nonlinear dynamical systems.



Pei Chen received her B.S. and M.S. degrees from Peking University, and Ph.D. degree from South University of Technology. Currently she is an assistant researcher at South China University of Technology. Her research interest includes deep learning, data mining, and computational biology.



Luonan Chen is a professor at Shanghai Institute of Biochemistry and Cell Biology, Center for Excellence in Molecular Cell Science, Chinese Academy of Sciences. He received his B.S. degree from Huazhong University of Science and Technology (China), and M.S. and Ph.D. degrees from Tohoku University (Japan). His research interest includes nonlinear dynamics, optimization, and computational systems biology.