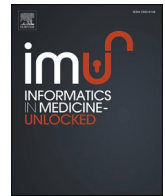




Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Similarity maps and pairwise predictions for transmission dynamics of COVID-19 with neural networks

Pitoyo Hartono

Chukyo University, Yagotohonmachi, 101-2, Nagoya, Japan

ARTICLE INFO

Keywords:

COVID-19
Transmission dynamics
Neural network
Long short-term memory
Time series prediction
Topological representations

ABSTRACT

On March 11, 2020, the World Health Organization declared COVID-19 as a pandemic. Since then, many countries have experienced the rapid transmission of this respiratory disease among their populations and have exercised many strategies to mitigate the spread of this disease. The prediction of the transmission dynamics serves important roles in designing mitigation strategies. However, due to the unknown characteristics of this disease, as well as the geographical and political factors, building efficient models of the dynamics for many countries is difficult. The objective of this study is to develop a transmission dynamics predictor that takes advantage of the time differences among many countries with respect to transmission of this disease, in that some countries experienced earlier outbreaks than others. The primary novelty of the proposed method is that, unlike many existing transmission predictors that require parameters based on prior knowledge of the epidemiology of past viruses, the proposed method only requires the transmission similarities between countries in the publicly available data for this current disease. In this paper, the viability and limitations of the proposed method are reported and discussed.

1. Introduction

At the end of 2019, China reported some cases of pneumonia of an unknown cause that rapidly became an outbreak [1]. On January 30, 2020, the World Health Organization (WHO) declared the outbreak as a public health emergency of international concern, and on February 11, 2020, the name for the coronavirus disease was announced as COVID-19. On March 11, COVID-19, the human respiratory disease caused by the new coronavirus, SARS-CoV-2, was declared as a pandemic by the WHO, with more than 100,000 cases in 114 countries as of that date. As of April 19, the completion date of this paper, there were more than 2.4 million cumulative cases in total.

As the possibility of a prolonged pandemic is imminent [2], it is urgent to understand the transmission dynamics of this disease for the prediction of future cases in various areas and, accordingly, the design of mitigation strategies.

Data from this outbreak have already been compiled and made available for research communities [3–5], which are rapidly studying various aspects of the transmission of this disease.

There are rich data collections for the study of the transmission of infectious diseases through populations. For example [6,7], and [8] are the seminal studies that have become the basis for the compartmental

SIR model which is often used at the present time. The SIR model divides a population into three compartments, namely, susceptible (S), infected (I), recovered or removed (R), and expresses the dynamics of the compartments with a series of ordinary differential equations, in which the analytical solutions are explained in Ref. [9]. Modifications to the basic SIR model have also been presented, for example, the SEIR Model in which one additional compartment, namely, exposed (E), is added to the SIR model [10–13].

In the wake of the current outbreak, some new models for transmission dynamics of this disease have been proposed [14–17]. The analysis of the correlation between the effectiveness of government interventions and the transmission dynamics has also been provided [18].

Artificial intelligence communities have been starting to approach many aspects of the COVID-19 crisis, as effectively reviewed in Ref. [19]. The proposed work utilizes a new neural network for transmission dynamics mapping, and further uses it for predicting the future cases in many countries. The proposed method takes advantage of the time variance with respect to the occurrences of the outbreaks in different countries, in that some countries or territories observed earlier outbreaks than others. However, while the transmission dynamics of many countries may start at different times, they often share similarities. In this preliminary study, a recently proposed topological neural

E-mail address: hartono@sist.chukyo-u.ac.jp.

<https://doi.org/10.1016/j.imu.2020.100386>

Received 25 April 2020; Received in revised form 25 June 2020; Accepted 25 June 2020

Available online 15 July 2020

2352-9148/© 2020 The Author.

Published by Elsevier Ltd.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

network, named the topological autoencoder (TA) [20], is utilized to map the transmission dynamics for more than 250 countries and territories. Given the time series of the confirmed cases in many countries as input, the TA produces a two-dimensional, and thus visible, map that projects the topological structures of those dynamics, in that similar dynamics are located adjacently, while dissimilar dynamics are separated by large spaces on the map. By setting a target country to be predicted, this map can then be used to locate a reference country with similar dynamics which started earlier, and can use the reference dynamics to train a neural network predictor, namely, long short-term memory (LSTM) [21,22], which is a deep recurrent neural network known to perform strongly with respect to modeling time series. Due to the lack of prior knowledge in the transmission dynamics for this disease and the relatively short period for the available time series data, predicting target dynamics using similar but longer dynamics can be efficient. The LSTM has also been applied in predicting the transmission dynamics of this disease in Ref. [23,24], although they differ significantly from the proposed work in that the LSTMs in the previous study were not utilized in a pairwise manner. In this paper, the efficiency and limitations of this pairwise prediction method are empirically evaluated.

The paper is structured as follows. Section 2 explains the mathematical concept of the topological autoencoder in the context of COVID-19 transmission dynamics data. Section 3 explains the experiments, in which the initial characteristics of the topological transmission

dynamics maps are explained in detail, while in the latter part, the pairwise predictions are explained. The conclusions are presented in the final section.

2. Topological autoencoder

The outline of the topological autoencoder (TA) with the framework for short-term prediction of transmission dynamics is illustrated in Fig. 1. The TA is a hierarchical neural network having a low-dimensional topological hidden layer that is a simplified version of the soft-supervised topological autoencoder (STA) proposed in Refs. [20], in which the basic mathematical properties were proposed in Ref. [25]. Originally, the STA can be trained in a supervised manner, an unsupervised manner or the mix of both, and produces a two-dimensional topological map that allows human users to visualize not only the structure of high-dimensional data but also their given concepts. The TA is a simplified version of the STA in that it is only trained as an autoencoder, which is a neural network that is trained to reproduce the given input in its output layer and thus encodes inherent structures of the inputs in its hidden layer. The TA here is different from the conventional autoencoder in that it has a two-dimensional topological layer that can be visibly observed by a human user.

Here, the inputs to the TA are the transmission dynamics of COVID-19 for more than 250 countries or territories that are characterized by

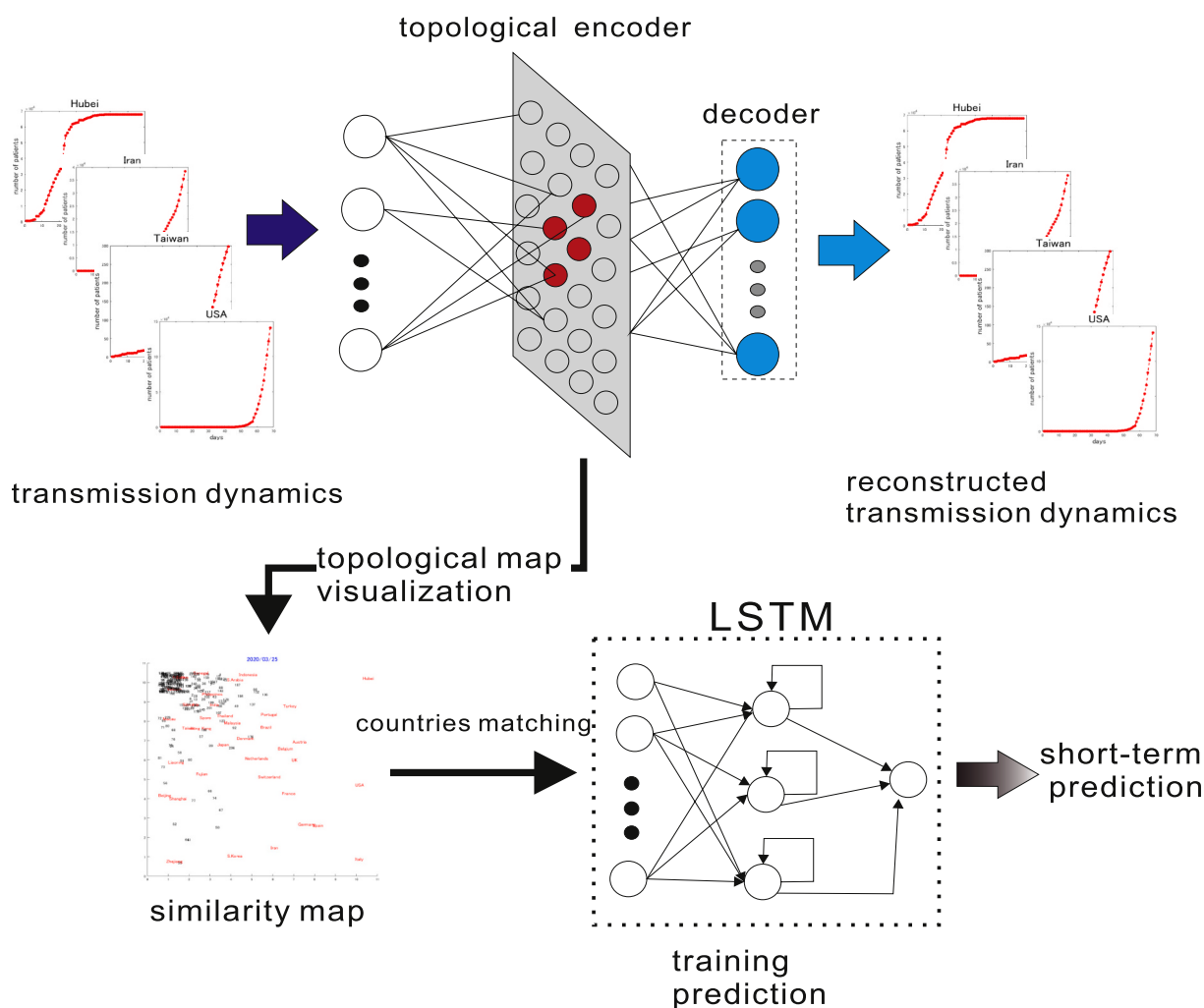


Fig. 1. Topological autoencoder and the framework for short-term prediction The TA is trained using transmission dynamics of more than 250 countries. During the learning process, a low-dimensional topological representation is organized in the hidden layer of the STA, resulting in a displayable similarity map. Given a target country, a country with similar but longer dynamics is chosen to be used to train an LSTM. The LSTM is then utilized for short-term prediction for the target country.

their daily numbers of confirmed cases as compiled by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University [5] and are made available in a GitHub repository [4].

The hidden layer of the TA is a two-dimensional topological map similar to Kohonen's self-organizing maps (SOM) [26,27], where hidden neurons are arranged in a two-dimensional grid, and each neuron is associated with a reference vector that has the same dimensionality as the input. The TA is then trained to reproduce the given input in its output layer, and hence is expected to extract the inherent structure of the inputs in its hidden layer. Since the hidden layer is a two-dimensional grid, the extracted structure of the inputs can be visualized. Through this visualization, human users can gain intuitive insights regarding the similarities and dissimilarities among many countries' transmission dynamics.

The mathematical explanations for the TA's learning mechanism are as follows.

Here, the transmission dynamics for country k , characterized by the daily confirmed cases in that country, recorded from January 22, 2020 for d days, are denoted by $\mathbf{X}^k \in \mathbb{R}^d$. Observing \mathbf{X}^k as an input at time t , the TA selects the best matching unit, $win^k(t)$, among all of the reference vectors associated with the hidden units of the TA as in Eq. (1), where $\mathbf{W}_j(t) \in \mathbb{R}^d$ is the reference vector associated with the j -th hidden unit at time t . Here, $\mathbf{W}_{win^k(t)}$ becomes the internal representation for the input \mathbf{X}^k at time t .

The output of the j -th hidden neuron, $H_j(t)$ at time t , is shown in Eq. (2), where $N(j, win(t))$ is the neighborhood function defined in Eq. (3). $N_0 > N_\infty > 0$ are the initial and the final values of the annealing term, t_∞ is the termination time, while σ is an empirically defined constant.

$$win^k(t) = \arg \min_j \|\mathbf{X}^k - \mathbf{W}_j(t)\| \quad (1)$$

$$H_j(t) = N(j, win^k(t)) e^{-\frac{\|\mathbf{x} - \mathbf{w}_j(t)\|^2}{2\sigma^2}} \quad (2)$$

$$N(j, win^k(t)) = e^{-\frac{dist(win^k(t), j)}{S(j)}} \quad (3)$$

$$S(t) = N_\infty + \frac{1}{2}(N_0 - N_\infty) \left(1 + \cos \frac{\pi t}{t_\infty}\right) \quad (4)$$

$$dist(win^k(t), j) = \|\mathbf{pos}(win^k(t)) - \mathbf{pos}(j)\|^2 \quad (4)$$

In Eq. (4), $\mathbf{pos}(j) \in \mathbb{R}^2$ is the coordinate of the j -th neuron in the two-dimensional hidden layer.

This neighborhood function ensures locality in the activation of hidden neurons in that the winning neuron generates the maximum output, while the outputs of other neurons are exponentially decreased along with their distance to the winning neuron on the two-dimensional grid of the hidden layer.

The values of the l -th output neuron, O_l , are defined in Eq. (5), where $f(x) = \frac{1}{1+e^{-x}}$.

$$O_l(t) = f((\mathbf{V}_l(t))^T \mathbf{H}(t))$$

Here, \mathbf{V}_l denotes the weight vector leading from the hidden layer to the l -th output neuron, and $\mathbf{H}(t) = (H_1(t), H_2(t), \dots, H_{N_{hid}}(t))^T$ is the hidden layer output vector, in which N_{hid} is the number of hidden neurons, and T indicates transposition.

Because the objective of the TA is to reconstruct the input in the output layer, the loss function for this neural network can be defined as in Eq. (5).

$$L(t) = \frac{1}{2} \sum_l (O_l(t) - X_l^k)^2 \quad (5)$$

Here, X_l^k is the l -th component of the input \mathbf{X}^k .

The TA is then trained to minimize the loss function with the standard stochastic gradient descent by modifying the weight vectors and the reference vectors.

The connection weights leading to the l -th output neuron are modified as follows.

$$\mathbf{V}_l(t+1) = \mathbf{V}_l(t) - \eta \frac{\partial L(t)}{\partial \mathbf{V}_l(t)} \quad (6)$$

$$\frac{\partial L(t)}{\partial \mathbf{V}_l(t)} = \delta_l^{out}(t) \mathbf{H}(t) \quad (7)$$

$$\delta_l^{out}(t) = (O_l(t) - X_l^k) O_l(t) (1 - O_l(t)) \quad (8)$$

Similarly, the modification of the j -th reference vector is modified as follows.

$$\mathbf{W}_j(t+1) = \mathbf{W}_j(t) - \eta \frac{\partial L(t)}{\partial \mathbf{W}_j(t)} \quad (9)$$

$$\frac{\partial L(t)}{\partial \mathbf{W}_j(t)} = \delta_j^{ref}(t) H_j(t) (\mathbf{X}^k - \mathbf{W}_k(t)) \quad (10)$$

$$\delta_j^{ref}(t) = \frac{1}{\sigma^2} \sum_l \delta_l^{out} v_{lj} \quad (11)$$

In Eq. (6) and Eq. (9), $\eta > 0$ is an empirically set learning rate, while V_{lj} in Eq. (11) is the j -th component of weight vector \mathbf{V}_l connecting the l -th neuron to the hidden layer.

The elaborations for the loss function derivations and the difference in reference vector modifications with respect to SOM are explained in detail in Ref. [20,25,28].

2.1. Experiments

In the first half of this section, the characteristics of the topological maps produced by the TA are explained, while in the later part, the utilization of these maps to run pairwise prediction, the evaluation and the limitations of the proposed pairwise prediction are explained.

2.2. Topological maps for COVID-19 transmission

Fig. 2 shows the topological maps for transmission dynamics for more than 250 countries or regions on February 20, 2020, March 6, 2020, March 30, 2020, and on April 13, 2020. It can be determined that the appearances of the maps change over time, in that on February 20 only a few countries or regions observed the transmission of this disease in their populations, while almost two months later, on April 13, most countries or regions had developed their unique transmission dynamics. The TA produces topological maps that project similar dynamics into adjacent regions on the map, while separating dissimilar dynamics. Hence, given a target country to be predicted, some countries that have similar dynamics can be selected for pairwise-prediction, which will be explained in the next subsection of this paper.

It must be noted that the topological maps here are produced during the learning process of the TA, in which the transmission dynamics of each country or region are treated as high-dimensional vectors, and thus the TA reduces their dimensions in its two-dimensional topological hidden layer. Here, the TA executes a nonlinear dimensionality-reduction process that, unlike the conventional principal component analysis (PCA), where the reduced dimensions are the linear combinations of the original ones, the dimensions cannot be easily interpreted. Hence, no unit is assigned to the axes in the map, which is common to nonlinear dimensionality reduction methods such as kernel PCA [29], Isomap [30], stochastic neighborhood embedding (SNE) [31] and its variant t-SNE [32], and the recently proposed UMAP [33].

The topological maps in Fig. 2a show the transmission dynamics up

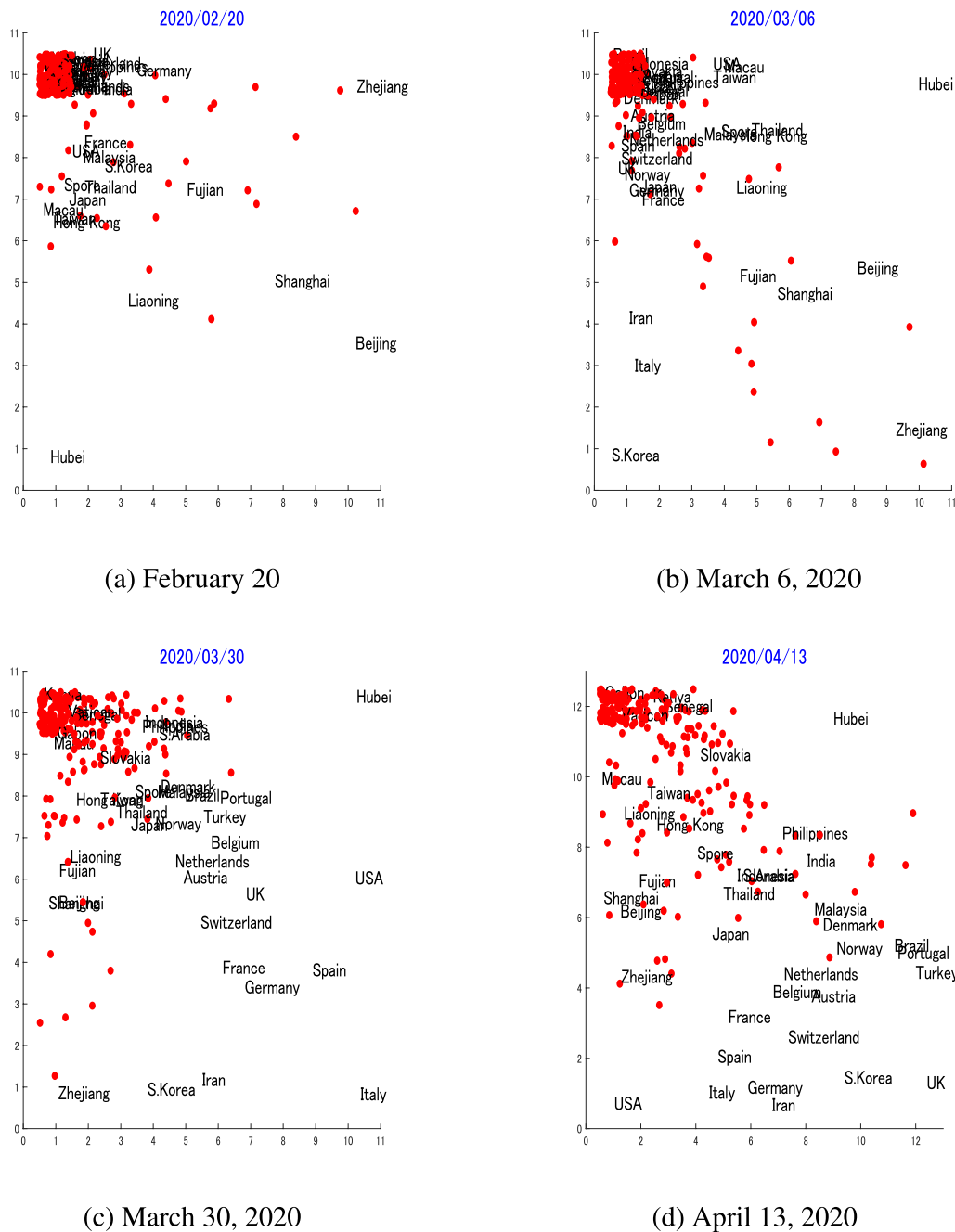


Fig. 2. Topological transmission dynamics maps. These figures show the topological transmission dynamics maps using the time series of the number of cases in over 250 countries or regions collected from Jan 22, 2020, until the specified date on each map. These maps show the development of the similarities of the transmission dynamics between countries or regions over the period from February 20 to April 13, 2020.

until February 20, 2020, the early stage when very few COVID-19 cases were detected outside China. From this map, it can be observed that many countries or regions occupied the upper left area on the map. The overlapping was due to their dynamics, in that most of them had identified zero or very few cases at the time. At this time, Hubei, the epicenter for the outbreak, was already an outlier on the map, as the number of cases in Hubei was of a different magnitude compared to other areas. It can also be observed here that the USA, France, South Korea, and many Asian countries or regions such as Malaysia, South Korea, Singapore, Thailand, Taiwan and Hong Kong were adjacently positioned due to their similar dynamics, while many of China's provinces such as Zhejiang, Shanghai, Liaoning and, naturally, Hubei, were outliers due to their early outbreaks. From Fig. 2b, it can be observed

that the upper left area of the map was still crowded, but many countries can be seen to escape from this overlapping area. This is because those countries or regions were beginning to develop different transmission dynamics. It is notable that Italy and Iran moved considerably from the crowded cluster due to the beginning of their outbreaks, while the USA still shared the space, and thus exhibits similar dynamics with Macau and Taiwan. The topological map in 2c shows the map until March 30, when there were already massive outbreaks in the USA and Italy that were directly reflected by their positions on the map. As of April 13, the world has observed more than 2 million confirmed cases, with the USA becoming the center of these outbreaks. The map shown in Fig. 2d has developed a new structure, on which the USA and many European countries with the largest numbers of cases are clustered within the left

bottom area on the map, while Hubei, the initial epicenter which has since flattened the dynamics, occupies the top right corner by itself, due to the lack of other similar dynamics.

Detailed explanations of the topological characteristics of the map are given as follows, with the objective of elaborating the connections between countries' dynamics with their representations on the map. The explanations are given for the March 30 map.

Fig. 3 shows the actual transmission dynamics until March 30 for some countries or regions that were grouped based on the regions they occupied on the map in Fig. 2c. From the left area, as shown in Fig. 3a, Beijing, Zhejiang and Vatican City exhibited dissimilar dynamics that were reflected by their large distances on the map. Beijing and Zhejiang have both contained the outbreaks but had different dynamics and cumulative patient numbers, while Vatican City had very few patients. In the middle of the map, Japan, the Netherlands and Switzerland can each be observed with considerably different dynamics, as shown in Fig. 3b. From the right side of the map, Hubei, the USA and Italy are shown in Fig. 3c. It is obvious that Hubei had stabilized, while patient numbers were growing steeply in the USA and Italy.

Fig. 4 shows the relation between the transmission dynamics and their representation on the map. Here, the representation of the particular dynamics of a particular country or region is the reference vector for the winning neurons for that country. Fig. 4a shows the representations for the dynamics of Beijing, Zhejiang and Vatican City, Fig. 4b shows the representations for Japan, the Netherlands and Switzerland, while Fig. 4c shows those of Hubei, the USA and Italy. It is important to notice here that different dynamics are represented with different representations, and the similarity and dissimilarity of the dynamics are reflected in the positions of the representations on the map.

To further explain the internal representation characteristics, the dynamics of South Korea and the UK are shown in Fig. 5b. These two countries are deliberately chosen because of their contrasting transmission dynamics, which are reflected by their largest distance on the map. South Korea is known for success in containing the spread of this disease after an early outbreak, while the UK was still struggling on March 30. Their internal representations are shown in Fig. 5b, which clearly shows the difference between these two countries.

In contrast, Fig. 6a shows the similar dynamics for Belgium and the Netherlands that were naturally placed adjacently on the map in Fig. 2c. The similarity in the transmissions dynamics was reflected in the similar representations in Fig. 6b, as well as their positions on the map.

Through these preliminary examples, it can be empirically shown that the hidden representation of the TA is able to capture the topological relations in different dynamics, and thus can be utilized for selecting a pair of countries with similar dynamics to be used as in the

pairwise prediction mechanism.

2.3. Short-term pairwise prediction

In this part of the paper, the pairwise prediction based on the dynamics similarity on the map is explained. Here, because of the relatively short length in time series, only short time prediction for three days is explained. However, since the data are updated daily, the extension of the prediction period can be performed without changing the proposed method.

For the short time prediction, daily numbers of confirmed cases in approximately 250 countries or regions from January 22 until April 15 are utilized. Here, the time series of confirmed cases until April 12 were utilized to build a predictor, while the data for the remaining three days were used for testing the predictor.

In this method, given a target country to be predicted, a reference country, which is another country with similar dynamics but longer time series, must be chosen by visually observing the transmission dynamics map generated by TA, as follows.

Here, let X^{target} be the transmission dynamics of a target country, where t^{target} is the time when the target country exceeded a specified number of cases, $threshold(target)$. Y^{target} is then a truncated X^{target} , starting at t^{target} and terminating on April 12, the end of the time series. Setting the value for $threshold(target)$ and finding t^{target} , $Y^j (j \neq target, t^j < t^{target})$ is the truncated time series of country j that reached the $threshold(target)$ earlier than the target country. Y^{target} starts from t^j and is arranged to have the same length as Y^{target} . Using the collections of the truncated transmission dynamics of some countries that are ahead of the target countries with respect to reaching a certain number of cases, a new topological map can be generated. Similarly to the previous map, this map also reflects the topological structure of the truncated time series, only this time, the maps are limited to the target country and the countries that were ahead of it.

By visually observing the generated map, a reference country can be manually chosen. The dynamics of the reference country, starting from t^j until April 12, are then used to train a deep recurrent neural network, long short-term memory (LSTM) [21,22], to build a short time predictor for the target country. If a number of candidates for a reference country were found, the one that started the earliest, and thus had the longest time series, was chosen. The reference country is ensured to have a longer time series than the target country, since the dynamics began earlier, but ended on the same day, April 12, as those of the target country. The idea here is to run a pairwise prediction, using similar but longer dynamics, to predict a shorter time series. In the COVID-19 crisis, this idea can be useful since past dynamics to train a predictor may not

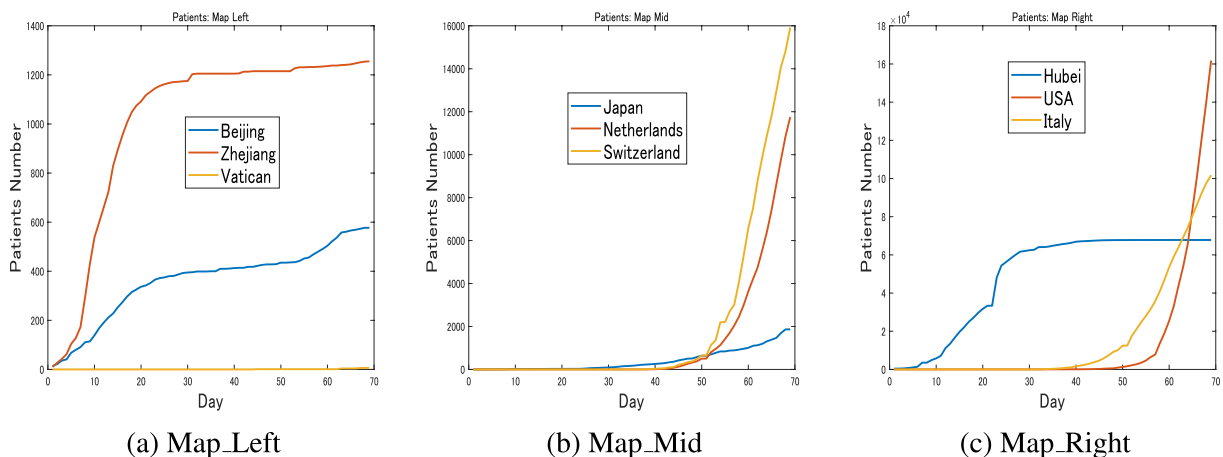


Fig. 3. Transmission dynamics: March 30. These figures show the time series of the number of cases for some countries or regions until March 30. It can be observed that countries or regions that are positioned adjacently on the map share similar dynamics, while dissimilar dynamics are associated with large distances on the map.

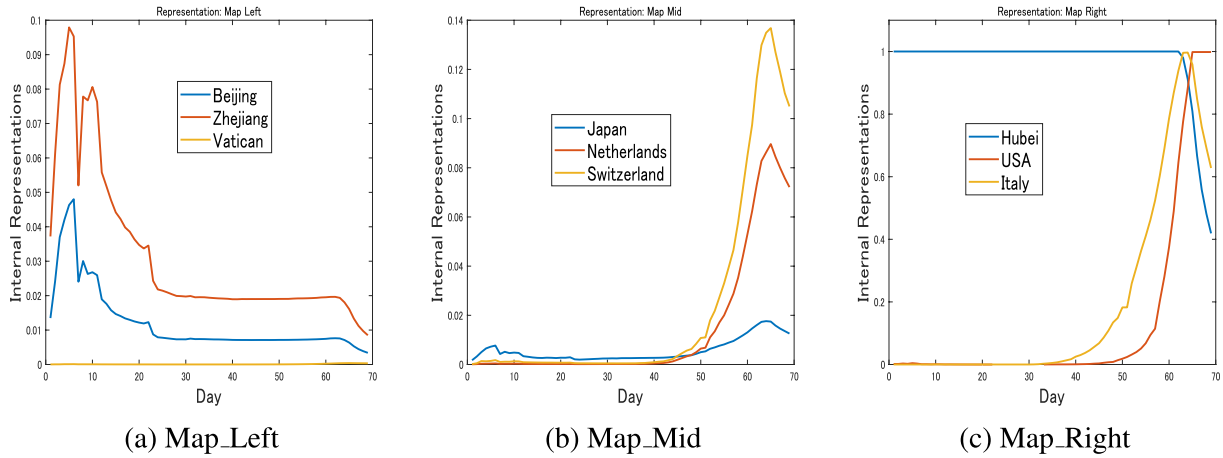


Fig. 4. Internal Representations: March 30 These figures show the representations of the transmission dynamics in Fig. 3. It can be observed here that similar dynamics are represented similarly, while dissimilar dynamics are represented differently.

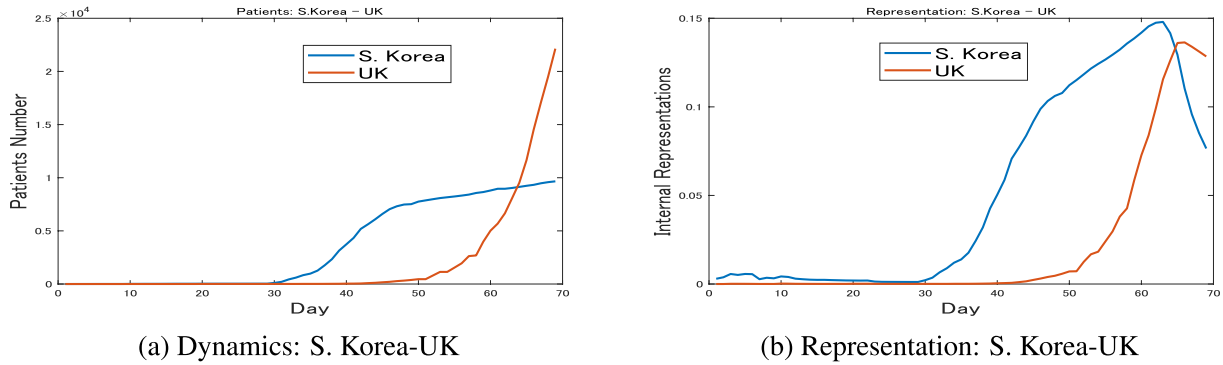


Fig. 5. S. Korea vs. UK: March 30 These figures show the contrasting transmission dynamics and their representations for South Korea and the UK up until March 30. It is obvious that the differences between the dynamics are directly represented in the representations.

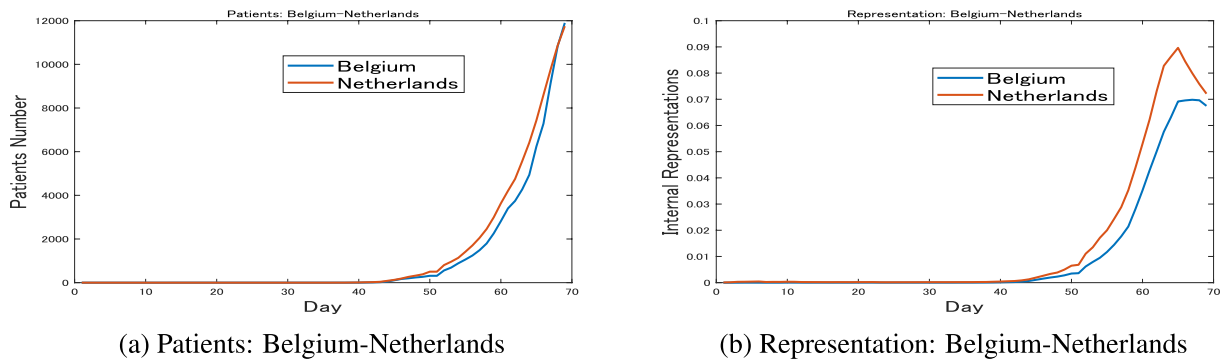


Fig. 6. Belgium vs. the Netherlands: March 30 These figures further emphasize the strong relation between the shapes of the dynamics and their representations. Here, the similarities between Belgium and the Netherlands dynamics are directly reflected by the similarity of their representations.

be available, and many countries still have a relatively short time series containing nonzero components.

Some of the experiments for 3 days of predictions are as follows.

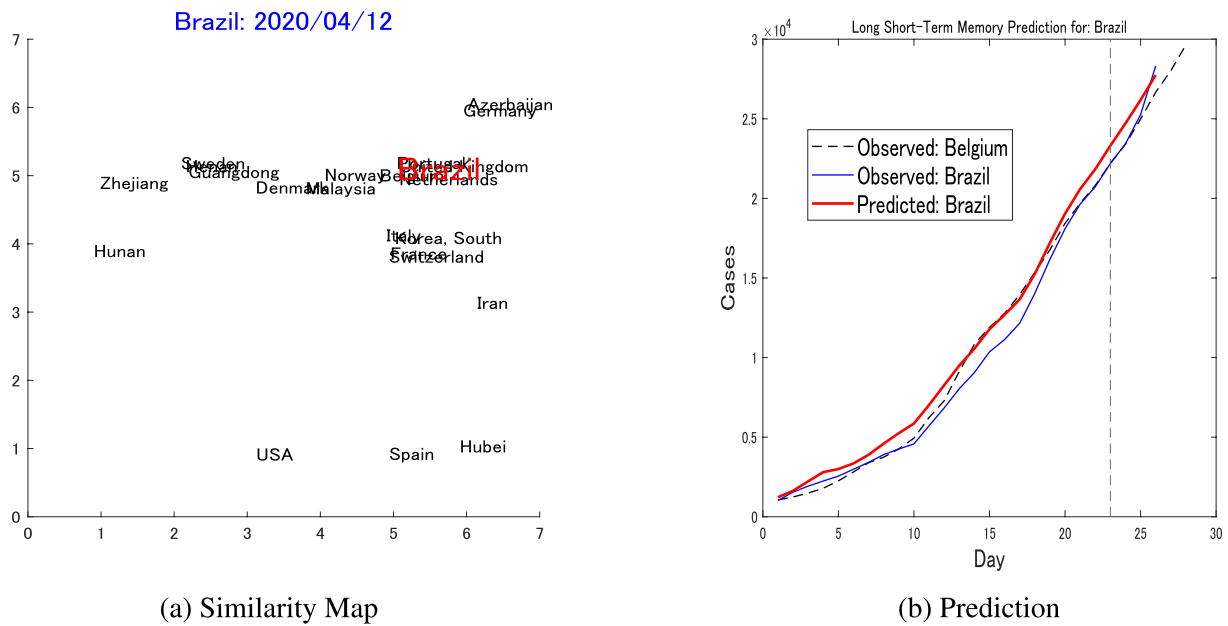
In the first prediction, Brazil is the target country. Brazil recorded its first few cases on February 26, and as of April 15 had recorded 28,320 cases. As the dynamics of this country are relatively long, and the patient number is relatively large, a $threshold(brazil) = 1000$ was set to search for similar dynamics. The TA-generated map of dynamics is shown in Fig. 7a, from which Belgium is hand-chosen as a reference country. While Brazil exceeded 1000 cases on March 21, Belgium exceeded that threshold on March 16 and was thus 5 days ahead of Brazil. Fig. 7b

shows the prediction graph, in which the horizontal axis shows the number of days since the two countries exceeded the threshold. The prediction error, defined in Eq. (12), is calculated for the three days from April 13–15, as shown on the right-hand side of the vertical dotted line.

$$err = \frac{1}{3} \sum_t \frac{|X^{target}(t) - X^{pred}(t)|}{X^{target}(t)} \times 100\% \quad (12)$$

In Eq. (12), $X^{pred}(t)$ denotes the prediction for the target country at time t .

The prediction error for Brazil was 3.66%. This prediction was then compared against a different LSTM that was trained only on the target



(a) Similarity Map

(b) Prediction

Fig. 7. Pairwise prediction: Brazil-Belgium(a) Shows the similarity map considering the dynamics of Brazil (threshold = 1000, up to April 12) as the target, while Belgium is one of the nearest neighbors that is chosen to be the reference for the LSTM's pairwise prediction. (b) Shows the short-term pairwise prediction between Brazil and Belgium.

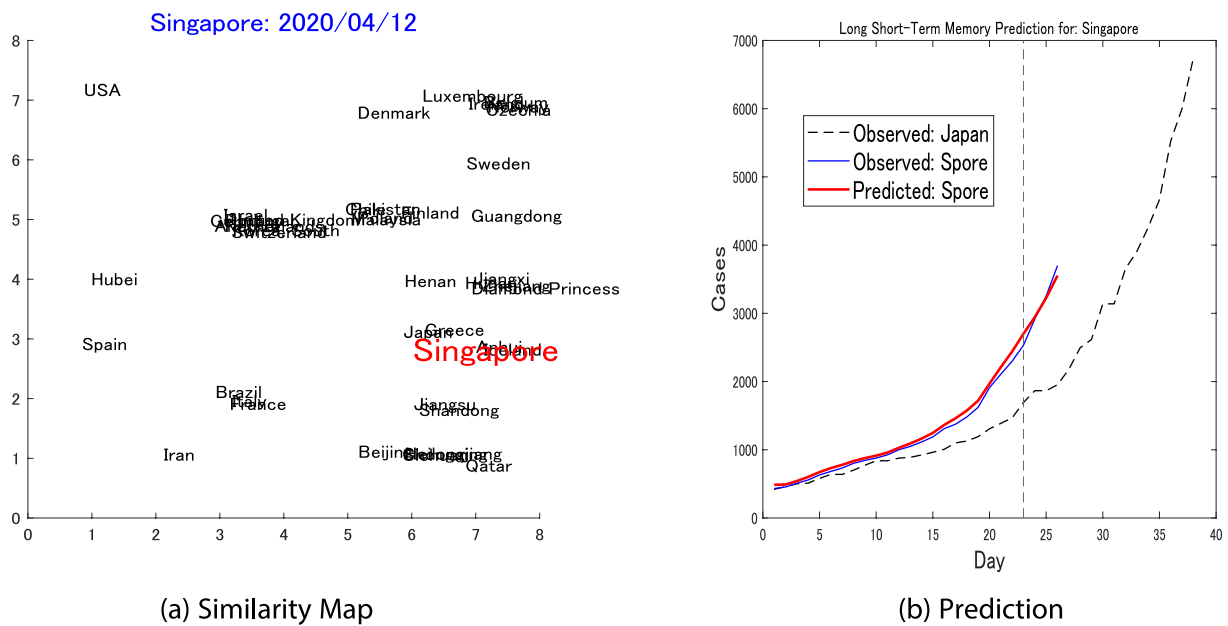
country dynamics up to April 12 and was then used to predict the next three days, for which the prediction error was 7.91%.

The second target country was Singapore. For this country, a *threshold(spore) = 400* was set and resulted in a similarity map as shown in Fig. 8a. The reference country chosen for Singapore was Japan. While Singapore reached the threshold on March 21, Japan reached this point on March 6. The prediction was depicted in Fig. 8b and resulted in a prediction error of 2.39%, which is significantly superior compared to the LSTM trained only on Singapore dynamics that generated a 15.72% prediction error.

The next prediction was for New Zealand, in which a *threshold(nz) = 300* was set and resulted in a similarity map as shown in Fig. 9a, from

which Guangdong was chosen as a reference. New Zealand reached the threshold on March 27, while Guangdong reached it on January 30, almost two months ahead. The pairwise prediction for New Zealand resulted in 0.74% prediction error, while the LSTM trained only on New Zealand dynamics generated a 3.72% prediction error.

The final prediction is for Taiwan, which is known to successfully mitigate the spread of the disease before the outbreak. Setting a *threshold(taiwan) = 300* resulted in a similarity map in Fig. 10a, which paired Taiwan with Beijing. Taiwan reached the threshold on March 30, while Beijing reached the same threshold on February 8. The pairwise prediction is depicted in Fig. 10b, which generates a result of 0.40% compared against the LSTM trained only on Taiwan's dynamics, which



(a) Similarity Map

(b) Prediction

Fig. 8. Pairwise prediction: Singapore-Japan(a) Shows the similarity map considering the dynamics of Singapore (threshold = 400, up to April 12) as the target, while Japan is one of the nearest neighbors and is chosen to be the reference for the LSTM's pairwise prediction. (b) Shows the short-term pairwise prediction between Singapore and Japan.

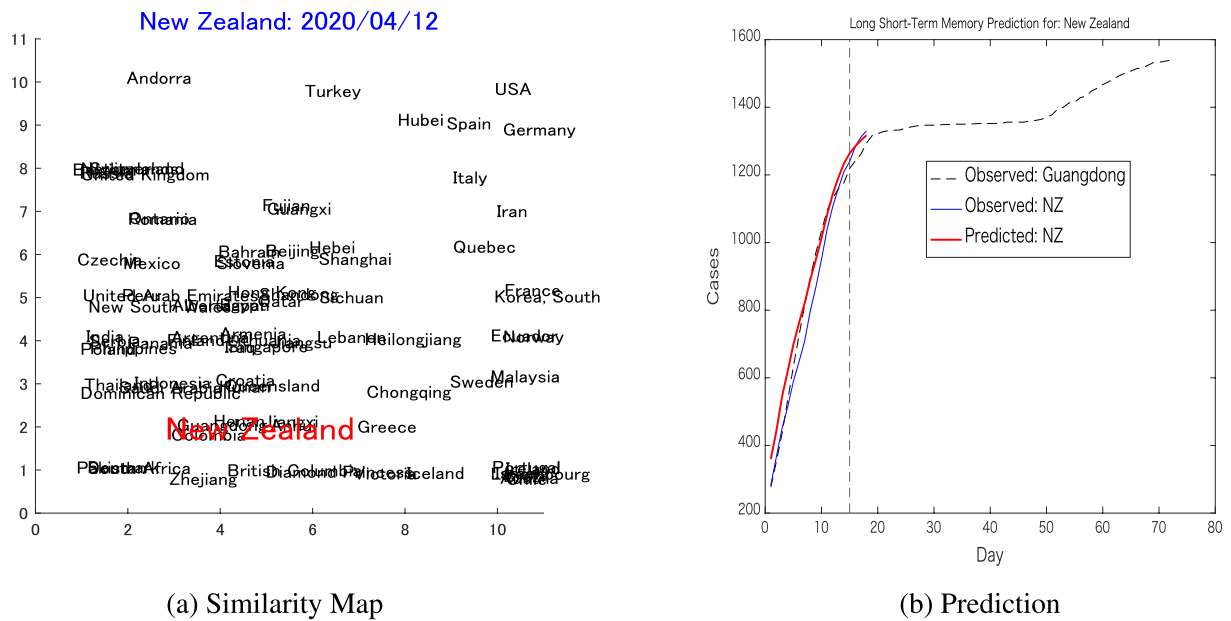


Fig. 9. Pairwise prediction: New Zealand-Guangdong (a) Shows the similarity map considering the dynamics of New Zealand (threshold = 300, up to April 12) as the target, while Guangdong is one of the nearest neighbors that is chosen to be part of the pair for the LSTM’s pairwise prediction. (b) Shows the short-term pairwise prediction between New Zealand and Guangdong.

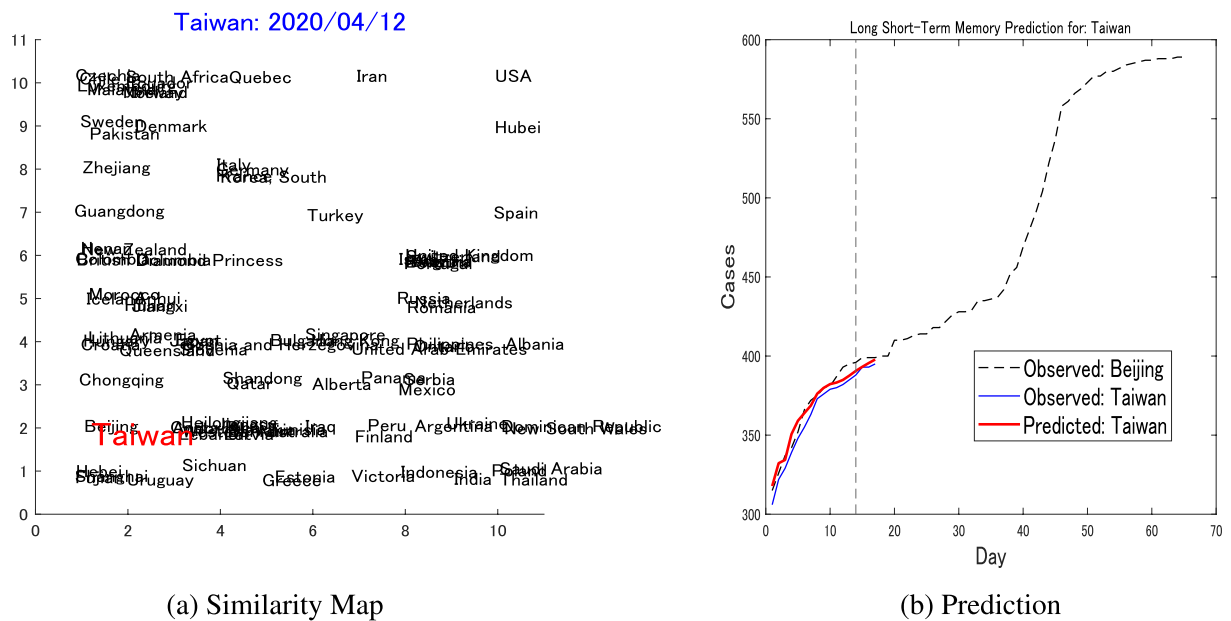


Fig. 10. Pairwise prediction: Taiwan-Beijing (a) Shows the similarity map considering the dynamics of Taiwan (threshold = 300, up to April 12) as the target, where Beijing is one of the nearest neighbors and is chosen to be the reference for the LSTM’s pairwise prediction. (b) Shows the short-term pairwise prediction between Taiwan and Beijing.

generated a prediction error of 1.32%.

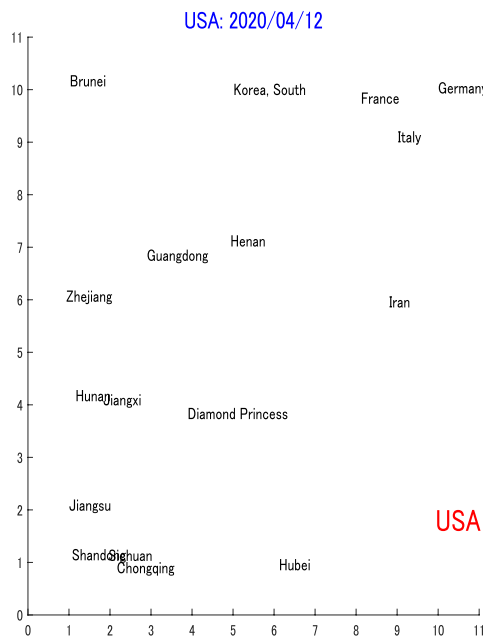
Some experiments in many areas of the world with various geographical, political and environmental factors empirically indicate the validity of the pairwise prediction mechanism. In the absence of any prior knowledge of the transmission dynamics that can be influenced by many factors, utilizing similar dynamics of other countries is one reasonable approach.

2.4. Limitations

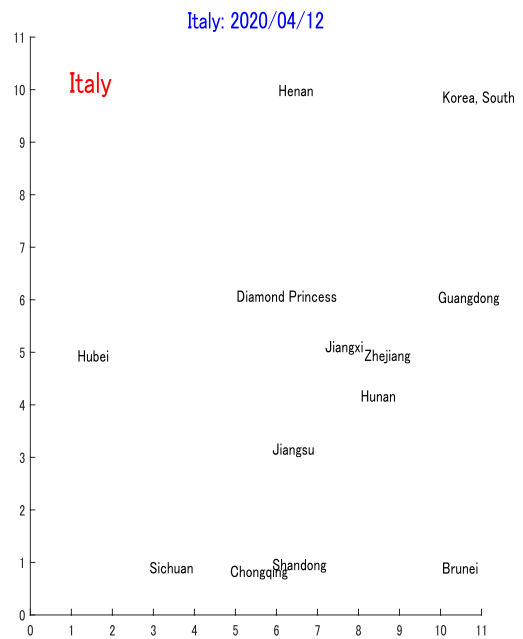
As the proposed pairwise prediction depends on the similarity be-

tween a target country’s dynamics and those of its reference countries, naturally it is possible that a suitable reference for a target area cannot be found. For example, Hubei, which experienced the first outbreak, did not have any counterpart, as there was no geographical location that preceded it with a certain number of cases. Some other countries did not have any counterparts due to their unique dynamics, for example the USA and Italy. The similarity maps, both with *threshold* = 500, are shown in Fig. 11, in which it is obvious that the two countries do not have any pair that can be used to predict their future dynamics.

The next example shows a target country that has a reference located relatively far from it. The target country in this example was Kuwait, and



(a) USA



(b) Italy

Fig. 11. Limitations These two maps illustrate the limitation for the proposed pairwise prediction. It is obvious from (a) and (b) that as of April 12, the USA and Italy developed unique dynamics that have no neighbors on the map, and no reference was found for their pairwise predictions.

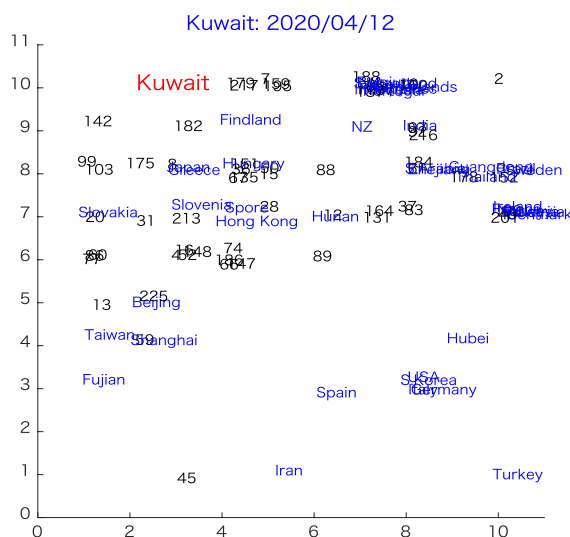
the reference country is Finland, whose relative positions can be observed in Fig. 12a, while the prediction was depicted in Fig. 12b. Here, the pairwise prediction error was 10.70%, while the LSTM trained on Kuwait's dynamics generates a prediction error of 15.48%. It can be observed here that the pairwise prediction error of these two relatively distant countries is significantly worse compared to the previous cases when close pairs of the target countries were found.

It should also be mentioned that the value of $threshold(x)$ for a country x that is required for searching a reference country must be

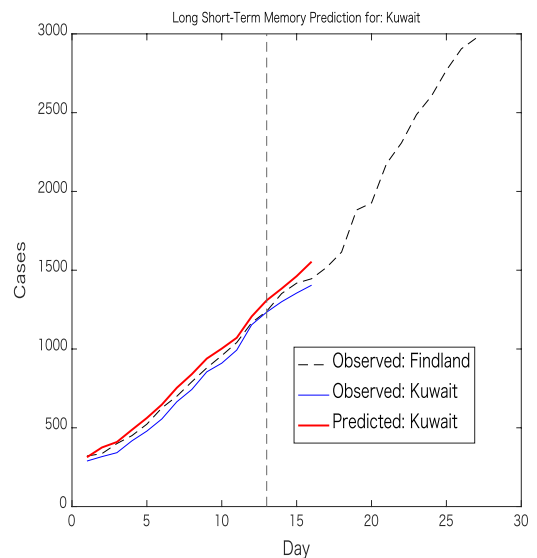
empirically decided, since it depends on the shape and length of its transmission dynamics. However, as a rule of thumb, the threshold is set to be 5%-25% from the number of cases on the last day of the time series to be used for training the LSTM.

3. Conclusion

In this paper, a recently proposed topological neural network is utilized for generating a transmission dynamics map on which



(a) Similarity Map



(b) Prediction

Fig. 12. Pairwise prediction: Kuwait-Finland These figures further illustrate the limitation of the proposed pairwise prediction by showing that choosing a distant country on the map for reference results in large prediction error, as is the case for Kuwait-Finland pairwise prediction.

similarities and dissimilarities of the dynamics among many countries can be visually observed. For this preliminary study, the reference country is manually chosen, although automatic selection of the reference country can easily be executed. After finding the pair, the longer dynamics of the reference country can then be used to train a neural network predictor, namely, LSTM. The primary novelty of the proposed pairwise prediction method is in its simplicity, in that if a target country follows the transmission dynamics of a reference country that generated its dynamics earlier and thus included a longer time series, then it is reasonable to expect the target country to follow the dynamics further, at least during the short-term. In the absence of other information, such as epidemiological prior knowledge, geographical and demographical characteristics and governmental intervention information, the proposed method for pairwise prediction can be one of the reasonable choices to predict further transmission dynamics in many countries. This approach will be potentially better in the future, especially as of the time this paper is completed; many countries and territories have started to stabilize their dynamics, and thus can potentially provide strong references for many countries that have yet to contain the spread of this disease.

The efficiency and the limitations of the proposed method are discussed in this paper. In this preliminary study, only three days of predictions were evaluated, but with the increase of the time series over time, longer predictions can be accomplished without any change in the proposed method. Continuous prediction is also made possible by repeating the map generation and LSTM training every few days.

Declaration of competing interest

None declared under financial, general, and institutional competing interests.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.imu.2020.100386>.

References

- [1] Tian J-H, Pei Y-Y, Yuan M-L, et al. A new coronavirus associated with human respiratory disease in China. *Nature* 2020;579:265–9.
- [2] Kissler SM, Tedijanto C, Goldstein E, Grad YH, Lipsitch M. Projecting the transmission dynamics of sars-cov-2 through the postpandemic period. *Science* 2020;368(6493):860–8.
- [3] Xu B, Gutierrez B, Mekaru S, et al. Epidemiological data from the covid-19 outbreak, real-time case information. *Scientific Data* 2020;7(106).
- [4] “2019 novel coronavirus covid-19 (2019-ncov) data repository by johns hopkins csse. <https://github.com/CSSEGISandData/COVID-19>; 2020.
- [5] Dong E, Du H, Gardner L. An interactive web-based dashboard to track covid-19 in real time. *Lancet Infect Dis* 2020;20(5):533–4.
- [6] Kermack WO, McKendrick AG, Walker GT. A contribution to the mathematical theory of epidemics. *Proc R Soc Lond - Ser A Contain Pap a Math Phys Character* 1927;115(772):700–21.
- [7] Kermack WO, McKendrick AG. Contributions to the mathematical theory of epidemics ii. the problem of endemicity. *Bull Math Biol Mar* 1991;53:57–87.
- [8] Kermack WO, McKendrick AG. Contributions to the mathematical theory of epidemics iii. further studies of the problem of endemicity. *Bull Math Biol Mar* 1991;53:89–118.
- [9] Harko T, Lobo FS, Mak M. Exact analytical solutions of the susceptible-infected-recovered (sir) epidemic model and of the sir model with equal death and birth rates. *Appl Math Comput* 2014;236:184–94.
- [10] Fang Haiping, Chen Jixiu, Hu Jun. Modelling the sars epidemic by a lattice-based monte-carlo simulation. In: 2005 IEEE engineering in medicine and biology 27th annual conference; 2005. p. 7470–3.
- [11] Saito MM, Imoto S, Yamaguchi R, Sato H, Nakada H, Kami M, Miyano S, Higuchi T. Extension and verification of the seir model on the 2009 influenza a (h1n1) pandemic in Japan. *Math Biosci* 2013;246(1):47–54.
- [12] Smirnova A, deCamp L, Chowell G. Forecasting epidemics through nonparametric estimation of time-dependent transmission rates using the seir model. *Bull Math Biol Nov* 2019;81:4343–65.
- [13] Yang Z, Zeng Z, Wang K, Wong S-S, et al. Modified seir and ai prediction of the epidemics trend of covid-19 in China under public health interventions. *J Thorac Dis* 2020;12(3).
- [14] Giordano G, Blanchini F, Bruno R, Colaneri P, Di Filippo A, Di Matteo A, et al. Modelling the covid-19 epidemic and implementation of population-wide interventions in Italy. *Nat Med Apr* 2020;26:855–60.
- [15] Kucharski AJ, Russell TW, Diamond C, et al. Early dynamics of transmission and control of covid-19: a mathematical modelling study. *Lancet Infect Dis* 2020/04/18;20(5):553–8.
- [16] Perc M, Gorišek Miksić N, Slavinec M, Stožer A. Forecasting covid-19. *Frontiers in Physics* 2020;8:127.
- [17] Zhang J, Litvinova M, Wang W, Wang Y, Deng X, et al. Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside hubei province, China: a descriptive and modelling study. *Lancet Infect Dis* 2020;20(7):793–802.
- [18] Y. Fang, Y. Nie, and M. Penny, “Transmission dynamics of the covid-19 outbreak and effectiveness of government interventions: a data-driven analysis,” *J Med Virol*, vol. 92, no. 6, pp. 645–659.
- [19] Bullock J, Alexandra Luccioni, Pham KH, Lam CSN, Luengo-Oroz M. Mapping the landscape of artificial intelligence applications against covid-19. 2020.
- [20] Hartono P. Mixing autoencoder with classifier: conceptual data visualization. *IEEE Access* 2020;8:105301–10.
- [21] Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput* 1997;9(8):1735–80.
- [22] Sherstinsky A. Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network. *Phys Nonlinear Phenom* 2020;404:132306.
- [23] Tomar A, Gupta N. Prediction for the spread of covid-19 in India and effectiveness of preventive measures. *Sci Total Environ* 2020;728:138762.
- [24] Chimmula VKR, Zhang L. Time series forecasting of covid-19 transmission in Canada using lstm networks. 109864–109864 *Chaos, Solit Fractals May* 2020;135:32390691[pmid].
- [25] Hartono P, Hollensen P, Trappenberg T. Learning-regulated context relevant topographical map. *IEEE Trans. on Neural Networks and Learning Systems* 2015; 26(10):2323–35.
- [26] Kohonen T. Self-organized formation of topologically correct feature maps. *Biol Cybern* 1982;43:59–69.
- [27] Kohonen T. Essential of self-organizing map. *Neural Network* 2013;37:52–65.
- [28] Hartono P. Classification and dimensional reduction using restricted radial basis function networks. *Neural Comput Appl* 2018;30(3):905–15.
- [29] Schölkopf B, Smola A, Müller K-R. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comput* 1998;10(5):1299–319.
- [30] Tenenbaum JB, Silva Vd, Langford JC. A global geometric framework for nonlinear dimensionality reduction. *Science* 2000;290(5500):2319–23.
- [31] Hinton GE, Roweis ST. Stochastic neighbor embedding. In: Becker S, Thrun S, Obermayer K, editors. *Advances in neural information processing Systems*, vol. 15. MIT Press; 2003. p. 857–64.
- [32] van der Maaten L. Visualizing high-dimensional data using t-sne. *J Mach Learn Res* 2008;9:2579–605.
- [33] McInnes L, Healy J, Melville J. Umap: uniform manifold approximation and projection for dimension reduction. 2018.