Real-time forecast of influenza outbreak using dynamic network marker based on minimum spanning tree

Kun Yang
South China University of Technology
Jialiu Xie
University of North Carolina at Chapel Hill
Rong Xie
Guangdong University of Economics and Finance
Yucong Pan
Guangdong Science and Technology Infrastructure Center
Rui Liu
South China University of Technology
Pei Chen (chenpei@scut.edu.cn)

Technical advance

Keywords: dynamical network marker (DNM), minimum spanning tree, early-warning signal, influenza outbreak, logistical regression

Posted Date: January 31st, 2020

DOI: https://doi.org/10.21203/rs.2.22426/v1

Version of Record: A version of this preprint was published at BioMed Research International on October 1st, 2020. See the published version at https://doi.org/10.1155/2020/7351398.

Real-time forecast of influenza outbreak using dynamic network marker based on minimum spanning tree

Kun Yang[1], Jialiu Xie[2], Rong Xie[3], Yucong Pan[4], Rui Liu[5*], Pei Chen[5*]

[1] School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China

[2] Department of Biostatistics, South China University of Technology, 27514, USA

[3] School of Information, Guangdong University of Finance and Economics, Guangzhou 510320, China

[4] Guangdong Science and Technology Infrastructure Center, Guangzhou, 510033, China

[5] School of Mathematics, South China University of Technology, Guangzhou 510640, China

* Corresponding authors

Email addresses: K.Y. csyk@mail.scut.edu.cn; J.X. jialiux2@live.unc.edu;

R.X. rongxie2005@qq.com;

Y.P. 807376553@qq.com; R.L. scliurui@scut.edu.cn and P.C. chenpei@scut.edu.cn

Abstract

Background: The influenza pandemic is a wide-ranging threat to people's health and property all over the world. Developing effective strategies for predicting the influenza outbreak which may prevent or at least get ready for a new influenza pandemic is now a top global public health priority.

Methods: Owing to the complexity of influenza outbreaks that are usually involved with spatial and temporal characteristics of both biological and social systems, however, it is a challenging task to achieve the real-time monitoring of influenza outbreaks. In this study, by exploring the rich dynamical information of the city network during

influenza outbreaks, we developed a computational method, the minimum-spanning-tree-based dynamical network marker (MST-DNM), to identify the tipping point or critical stage prior to the influenza outbreak.

Results: With historical records of influenza outpatients between 2009 and 2018, the MST-DNM strategy has been validated by accurate predictions of the influenza outbreaks in three Japanese cities/regions respectively, i.e., Tokyo, Osaka, Hokkaido. These successful applications show that the early-warning signal was detected 4 weeks on average ahead of each influenza outbreak.

Conclusion: The results show that our method is of considerable potential in the practice of public health surveillance.

Keywords: dynamical network marker (DNM), minimum spanning tree, early-warning signal, influenza outbreak, logistical regression

Background

Influenza, a seasonal, contagious and widespread respiratory illness, has always been a huge threat to people's health. According to the World Health Organization, up to 650,000 deaths annually are associated with respiratory diseases caused by seasonal influenza [1]. In the United States, the influenza pandemic leads to an average of 610,660 deaths per year and 3.1 million hospitalized days [2]. It is estimated that the total economic burden caused by influenza reaches 81.7 billion US dollars each year [3]. Therefore, from both public health and economic respective, it is crucial to detect the early-warning signal of imminent influenza outbreak so that timely preventive measures can be carried out to prevent a new influenza pandemic, or at least reduce the magnitude of influenza

46 outbreaks [4,5]. However, it is usually a challenging task to predict the influenza outbreak due to
47 the complexity of its temporal and spatial characteristics. First, the records of world-wide influenza
48 pandemics showed that each outbreak differed from the others with respect to etiologic agents,
49 epidemiology, and disease severity [6]. Second, there is a major obstacle for most developing
50 countries to deploy influenza forecasts, that is, the national surveillance system for infectious
51 disease could be either too costly or inaccurate [7]. Therefore, it is of great concern to develop a
52 cost-effective computational method for predicting the outbreak of influenza only based on the
53 available data.

54 In this study, by exploring the rich dynamical information provided by high-dimensional
55 records of clinic hospitalization data, we developed a practical computational method, i.e., the
56 minimum-spanning-tree-based dynamical network marker (MST-DNM), to quantitatively measure
57 the dynamical change of a city network and thus detect the early warning signal of an influenza
58 outbreak. The theoretical basis of MST-DNM is our recently proposed concept, the so-called
59 dynamical network marker (DNM) [8], which is a dominant group of variables satisfying three
60 generic properties for the impending critical transitions, that is, (1) the correlation between any pair
61 of members in the DNM group rapidly increases; (2) the correlation between one member of the
62 DNM group and any other non-DNM member rapidly decreases; (3) the standard deviation or
63 coefficient of variation for any member in the DNM group drastically increases. Different from
64 traditional biomarkers, DNM method aims at detecting the early-warning signal of the critical state
65 before the occurrence of a catastrophic event, by mining the critical information from high
66 dimensional time series data [8,9].The DNM method has been applied to real-world datasets and
67 successfully identified the critical states for a number of biological processes, such as the critical

state of cell differentiation [10], the tipping point during the cell fate decision process [11], the critical transition in the immune checkpoint blockade-responsive tumor [12], the multi-stage deteriorations of T2D [13], acute lung injury [14], HCV induced liver cancer [15], cancer metastasis [16], and other complex diseases [16-20]. However, to accurately predict the influenza outbreak, new computational method is required to explore and measure the criticality from a network perspective by considering the geographic information of a city.

The MST-DNM is a novel network-based computational method combined with minimum spanning tree for accurate detection of early-warning signal to the influenza outbreak. The spread of infectious diseases in a region is described as the dynamical evolution of a nonlinear system, while the influenza outbreak is regarded as a qualitative state transition of the dynamical system. Without loss of generality, there are three states for the influenza outbreak (Figure 1), that is, a normal state with high stability and robustness to disturbances, standing for the period with few clinic visits; a pre-outbreak state (critical state) with low resilience and high convertibility, representing the critical stage just before the emergence of massive clinic visits; and an outbreak state with high stability and robustness, which is an irreversible state or severe flu pandemic with massive clinic visits. Clearly, identifying the pre-outbreak state is crucial in influenza control since timely management may greatly reduce the magnitude and duration of influenza outbreak. Specifically, by combining the geographically adjacent information, transportation, population, and the number of clinics of each city district, we constructed a city network with edge-weights which were assigned as the correlation between the clinic visit numbers of two adjacent districts. By analyzing the dynamical transmission of influenza in the city network, the proposed MST-DNM can accurately identify the pre-outbreak state and thus early signal influenza outbreaks or potential

pandemics. Specifically, the MST-DNM method was employed to probe useful dynamical information in a city network, which is modeled based on geographic location and traffic conditions, from the high-dimensional clinic-visiting data of influenza, which are from 175 clinics distributed in 23 wards of Tokyo, Japan, 139 clinics distributed in 30 cities of Hokkaido, Japan, and 197 clinics distributed in 11 wards of Osaka, Japan. Clearly, such real-time data could be much more readily available for a large-scale surveillance system. The results indicate that the MST-DNM method is capable of monitoring the infection process of the flu in real time and timely identifying the warning signal before the outbreak of influenza. Moreover, by analyzing the dynamic changes of the minimum-spanning tree in a city network, it provides a new approach to study the epidemic spread in a city. Therefore, this method is of great applicable potential in setting up a real-time surveillance system, which could be great favorable for preventive care or the implementation of interventions to a health epidemic.

Methods

Theoretical Background

The influenza spread and outbreak is a complex dynamic process of a nonlinear system. According to the DNM theory, when a complex system approaches to a tipping point or critical transition point, there is a dominant group, i.e., the DNM, which satisfies the following three essential properties [8]:

- The correlation (PCC_{in}) between each pair of members in the DNM group dramatic increases;
- The correlation (PCC_{out}) between a member of the DNM group and a non-DNM member rapidly decreases;
- The standard deviation (SD_{in}) for each member in the DNM group drastically increases.

In general, the above properties can be roughly understood as that the emergence of the DNM group with violent fluctuation and high correlation signifies the upcoming critical transition. Thus, these properties can be utilized as three criteria to identify the critical state of a complex biological system.

Based on the DNM theory, we developed the MST-DNM method in order to accurately predict the early-warning signal to the influenza outbreak, by combining with the minimum spanning tree in a city network. According to our method, the evolution process of flu outbreak could be modeled as three diverse stages or states (Figure 1): (i) the normal stage, which is a stable state with high resilience. (ii) the pre-outbreak stage, which is an unstable critical state with low resilience. This critical state is the limit of the normal state and at the edge of transition into an epidemic outbreak of influenza. (iii) the outbreak stage, which is a steady and irreversible stage with a large number of clinic visits caused by influenza. It would bring heavy economic burdens to people and society and strongly impact the existing social health security system once in this status. Consequently, it's crucial to identify the warning signal of the pre-outbreak state to prevent people and social from the catastrophic flu outbreak in some effective measures.

Algorithm

The sketch of MST-DNM method was presented in Figure 2. First, it is noted that MST-DNM method is applied to a city network for monitoring the influenza spread and outbreak in such a city. Therefore, the first step of our method is to model a city network by combining the information of geographically adjacent relationship, transportation, population, the number of clinics of each city district. Then, a weight was assigned to each edge of the city network, which was the correlation between the numbers of clinic visits of two adjacent districts. Based on such weighted city network,

133 our method is implemented. Specifically, in order to detect the critical state of influenza outbreak,
134 the procedure of MST-DNM method can be described as the following detailed steps. Its pseudocode
135 is illustrated in Table 1.

136 **(i) Modeling a city network structure**

137 A city network is modeled based on its administrative divisions' geographic location and their
138 adjacent information. As demonstrated in the Figure 2, for example, there are 23 districts in
139 Tokyo, so that 23 nodes are added into the Tokyo city network. Furthermore, the edges between
140 nodes in the network are established based on the adjacency relations of those corresponding
141 districts.

142 **(ii) Data pre-processing**

143 For each district of a city, it is necessary that the raw data should be averaged in terms of the
144 total number of clinics within the district, owing to the enormous discrepancy of the number
145 of visits between different clinics. Afterwards, the processed data is mapped to the city network.

146 **(iii) Implement**

147 The city network can be represented as a graph $G = (V, E)$, where $V = \{v_i\}_{i=1}^{M}$ is a set of M
148 vertexes in this network, and $E = \{e_k\}_{k=1}^{N}$ is set of N edges in this network. There are the
149 following procedures.

150 First, we consider the number of clinic visits per week of a district as a sample *s*, forming a
151 series of time series data. In other words, when the city network is at the week *t*, there is a
152 sequence of clinic-visiting data $\{s_1, s_2, \ldots, s_t\}$ for each vertex v_i.

153 Second, for each edge e^k of the city network at the week *t*, calculate the correlations between
154 the two vertexes v_i, v_j of this edge to give it a weight W_t^k:

$$W_t^k = \delta||PCC_t(v_i, v_j)| - |PCC_{t-1}(v_i, v_j)||$$

where $PCC_t(v_i, v_j)$ represents the Pearson Correlation Coefficient (PCC) between the two vertexes v_i, v_j at week *t*, and $PCC_{t-1}(v_i, v_j)$ represents the Pearson Correlation Coefficient between the two vertexes v_i, v_j at week *t*-1, parameter δ is of the following form:

$$\delta = ||SD_t(k)| - |SD_{t-1}(k)||,$$

where $SD_t(k)$ represents the standard deviation (SD) of all simple data of the two vertexes of this edge e^k at week *t* and $SD_{t-1}(k)$ represents the standard deviation of all simple data of the two vertexes of this edge e^k at week *t*-1. After this step, we have obtained a set of weighted differential-network $\{N_1, N_2, \ldots, N_t, \ldots\}$.

Third, when the city network is at the week *t*, in order to better describe its evolution as the number of visits changes, it's required to obtain its minimum spanning tree. In this study, the Kruskal's algorithm is applied to the time-specific weighted differential-network N_t (such network is generated specifically for a timepoint) to obtain its minimum spanning tree MST_t. The detailed flow of Kruskal's algorithm is presented in the Table 2. Then we can calculate the weight sum L_t of this minimum spanning tree as the MST-DNM score:

$$L_t = \sum_{i=1}^{K} Weight_i$$

where $Weight_i$ represents the weight of edge e_i in MST_t, and K represents the total number of edges of MST_t.

In the ideal case, when the network system approaches a tipping point, there are the following two properties for the relationship between nodes in the network:

- The nodes in the city network are all DNB members. The standard deviation of these members and the Pearson's correlation coefficient between these members both dramatic

177 increases.

178 ● There are DNB and non-DNB members in the city network. The standard deviation of the
179 DNB members dramatic increases but the Pearson's correlation coefficient between DNB
180 members and non-DNM members decreases significantly, i.e., its absolute value increases
181 significantly.

182 Therefore, the city network's MST-DNM score L_t, which is based on the standard deviations
183 of these members and their Pearson's correlation coefficients, could be employed as an index
184 for quantitatively analyzing the significant change of the city network, and thus detecting the
185 warning signal of the critical point.

186 **(iv) Identifying the critical state**

187 After the above procedure, it is possible to quantitatively analyze and monitor the dynamical
188 process of influenza spreading based on the indicator L_t. Nevertheless, it's still a tough task to
189 confirm the tipping point. In some previous studies, the fold-change thresholds were used to
190 detect the warning signal [21]. However, such empirical or tunable threshold is not a universal
191 method for different data or network structure. In this study, the logistic regression is applied
192 to determine the appearance of the tipping point, which is widely employed in the biological
193 field [22] due to its intrinsic advantage that the threshold is determined by the data itself. In
194 view of the sufficient training data (several years of clinic-visiting records), the learning-based
195 approach would be an optimal option.

196 Logistic regression, which essentially is a linear regression model based on the sigmoid
197 function, is used to analysis the dataset with duality to explore relationship between its internal
198 independent variables, i.e., solving a two-classes (0 or 1) problems. Assume a dataset with m

199 samples and n feature, and each sample with a binary label. Then, we will get a sample matrix
200 $X = (x_1, x_2, \dots, x_m)^T \in R^{mn}$, where x_i is a column matrix with n features, and corresponding
201 label $Y = (y_1, y_2, \dots, y_m)$, where y_i represents a binary label (0 or 1) . Usually, we will add
202 an extra item to X as a bias, therefore each x_i is represented by $x_i = (x_i^0, x_i^1, \dots, x_i^n)$. Then
203 the sigmoid function is applied to calculate the probability for x_i belonging to 1:

204 $$P(y_i = 1|x_i; \omega) = 1/(1 + \exp(-x_i^T \omega))$$

205 According to the above form, the key to the logistic regression model is to train a suitable
206 parameter ω based on the given sample X and label Y. Therefore, the following loss
207 function based on the negative log-likelihood is applied to optimize our logistic regression
208 model to obtain a suitable ω :

209 $$L(\omega) = -\sum_{i=1}^{n} (y_i \log x_i^T \omega + (1 - y_i) \log(1 - x_i^T \omega)) + ||\omega||_1$$

210 In order to prevent our model from overfitting, the l_1 norm was added into the loss function.
211 Since there is no direct solution to this loss function at present, we used coordinate descent to
212 minimize this loss function with respect to ω.
213 In this study, we used the MST-score of each week as the X and the relevant state as label Y,
214 where 1 represents the critical state and 0 represents others. For a certain year, the logistical
215 regression model is trained by other years' datasets, we tested whether the week $T = t$ is the
216 tipping point. As long as the probability of x_t belonging to 1, i.e., $P(y_t = 1|x_t; \omega) = 1/(1 +$
217 $\exp(-x_t^T \omega))$, is greater than 0.5, this week is considered to be the critical state. Otherwise,
218 this week is classified as the normal state. Then, the week $T = t + 1$ is selected as the new
219 test point to carry on.

Results

Predict the outbreak of seasonal influenza in Tokyo

It's usually too complicated to mathematically express the influenza transmission kinetics before a sudden outbreak, because the influenza spread involves massive parameters from both biological and social systems. Based on the dynamical systems theory, there exists a so-called bifurcation point when the dramatic fluctuations or a qualitative transformation in a network from its normal status [20,23]. It means that the state transition of a dynamical system would gradually be restricted in a one- or two-dimensional space so that the system can be simply expressed and understood while approaching to the bifurcation point [8]. According to this theory, it's achievable that developing a general method to detect the tipping point of influenza outbreak only based on the observed data.

As shown in the Figure1, we collected the historical clinic-visiting data caused by influenza from clinics in 23 districts of Tokyo, Japan from January 1, 2009 to May 31, 2019. It can be regarded as the outbreak point of flu when the number of total clinic visits reaches the peak in each year. According to the proposed method, MST-DNM, the following procedures will be carried out to identify the critical state of flu outbreak in Tokyo. First, we modeled a 23-node network according to the geographic location of 23 wards and their adjacency. Second, we mapped the clinic-visiting numbers into corresponding node, assign weights (i.e., the correlations between two adjacent nodes, the detailed calculation is in Section Methods) to edges and calculate the weight sum of minimum spanning tree of this network for each week. Finally, an analyzed data matrix constituted by MST-DNM scores were obtained, which were employed to train a logistic regression though leave-one-out cross validation, and further detect the tipping point of influenza for each year.

As presented in the Figure3, the early-warning signals of the seasonal influenza outbreak were

242 detected by our MST-DNM method. It can be seen that the flu outbreak of each year is quite regular
243 except year 2009. The worldwide large-scale outbreak of influenza A (H1N1) in 2009, which was
244 reported first in Mexico, led to a massive long-term outbreak of influenza in Tokyo. It is explicit
245 that the peak of L_t appears earlier than the peak of the clinic-visiting counts for 4 weeks on average.
246 Therefore, before the outbreak of influenza, our MST-DNM score is quite sensitive and the index
247 L_t increases drastically, which implies the appearance of critical state of the influenza outbreak.
248 In order to better demonstrate the dynamical process of the influenza spread in the network
249 level, the evolutions of minimum spanning tree of the city network can also be presented. As shown
250 in Figure 4, it is seen that there are almost no influenza cases at each node/ward and the correlations
251 between these adjacent nodes/wards are relatively low at the beginning. In the city network, when
252 the correlations between the adjacent nodes/wards drastically increase, which are the necessary
253 conditions of the DNM features, it indicates that the influenza spread in this city is closed to its
254 outbreak point. Furthermore, the edges of the minimum spanning tree become thicker before the
255 nodes turn red in week 54, which means that the early warning signals of our method appears before
256 the flu outbreak point. The dynamical evolution of minimum spanning tree of the city network
257 illustrates that the system base on MST-DNM method is able to monitor the whole process of
258 influenza outbreak in real time and issue an early-warning signal in time.

259 **Application of MST-DNM in Osaka and Hokkaido**

260 In order to illustrate the universality of our MST-DNB method, we also applied it to detect the early-
261 warning signals of flu outbreak in Hokkaido and Osaka. Similar to the processing flow in Tokyo
262 city, a 30-node city network modeled for Hokkaido region and a 11-node for Osaka city. Then we
263 mapped the clinic-visiting data to the corresponding network and calculate the minimum spanning

tree. Finally, a logistic regression model trained by data consisting of MST-DNM scores was applied to detect the tipping point of influenza for each year.

As shown in Figures S1-S2 of Supplementary Information [see Additional file 1], the critical state of the influenza outbreak was smoothly detected by our method MST-DNB in Hokkaido between 2011 to 2015 and in Osaka between 2012 to 2017 respectively. In other words, the MST-DNG method is quite general and robust irrelevant to the scale of city network. The dynamic evolutions of the minimum spanning tree of Hokkaido city network and Osaka city network was shown in Figures S3 and Figures S4 respectively.

Discussion

The key role of the minimum spanning tree

In order to demonstrate the key role of the minimum spanning tree in our approach, we compared the effect of the MST-DNM method on the presence or absence of the minimum spanning tree in 2010, which was presented in the Figure 5A. It can be seen that the early-warning signal detected by a DNM method without minimum spanning tree is far away from the influenza outbreak point but another signal appears in an appropriate time point.

An undirected and edge-weighted minimum spanning tree is the smallest tree model that minimizes the sum of the weights of all connected edges in the original network. It's able to reflect the overall changes of the network structure and could avoid the impact caused by local abnormal correlations around the node 7 in week 45, which indicates that the minimum spanning tree plays a key role in the prediction process of outbreak points.

Performance comparison with other methods

In the previous work, we developed a groundbreaking network-based approach for predicting influenza outbreaks, so-called landscape dynamic network marker, which used empirical fold-change threshold to recognize the significant changes in DNM score to get the early-warning signal. We compared the performance of the proposed method MST-DNM with different tipping point determination strategies, that is, threshold determined from logistic regression and empirical threshold, which was presented in Figure 6. It is clear that the performance of MST-DNM method based on logistic regression is better than which on the fold-change threshold. Actually, the logistical regression has natural advantages relative to the traditional early-warning signal determination methods. The logistical regression model is a more general and more robust method only with some appropriate training measures.

Conclusions

Japan suffered a serious influenza outbreak at the beginning of year 2019. According to the reports of about 5000 designated medical institutions across Japan, there was an average of 57.09 influenza patients per institutions in the week from January 21st to 27th, which hit a new historical high since the first statistics in 1999. The influenza epidemic causes school suspension and the absence of a large number of workers, which would further result in a decline in social productivity and affect the economic development. It is estimated that the direct economic losses caused by the 2009 influenza pandemic to countries are about 0.5% to 1.5% of gross domestic product (GDP) [24]. However, the actual losses may be higher, due to the underestimate for the indirect economic losses caused by other infection prevention and control measures, such as the decline of tourism. Therefore, in order to better prevent the outbreak of influenza, it's quite essential to establish a real-time

monitoring system only based on available and robust data, such as the number of clinic visits issued by the relevant health department.

Based on the DNB theory, which was applied to detect the tipping point or analysis critical transition of complex diseases on related genomic data in our previous works, combined with minimum spanning tree and logistic regression, a novel computable method called MST-DNM was developed to identify the early-warning signal of influenza outbreak in Tokyo, Osaka and Hokkaido of Japan. In our MST-DNM method, we first extract the crucial characteristics of the pre-outbreak state of influenza using DNB and minimum spanning tree from high-dimensional and longitudinal clinic-visiting counts. Then, the logistic regression trained by leave-one-out cross validation is applied to identify the pre-outbreak state and issue an early-waring signal based on these crucial characteristics. As shown in Figure 3 and Figure 4, the MST-DNM method could timely detect the early-waring signal of influenza outbreak, which makes it quite possible to construct a real-time and effective influenza surveillance system. Nevertheless, there are still a few ways to improve the performance of our algorithm, such as using other robust but hardly obtainable data like population movement between wards and flu epidemic report to calculate the Pearson correlation coefficient and standard deviation, which is one of our future topics.

List of abbreviations

DNB: dynamic network biomarker; DNM: dynamic network marker; MST: minimum spanning tree; PCC: Pearson's correlation coefficient; SD: standard deviation; ROC: receiver operating characteristic curve; MST-DNM: the minimum-spanning-tree-based dynamical network marker.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

The historical raw data is available from Tokyo Metropolitan Infectious Disease Surveillance Center (Link: https://survey.tokyo-eiken.go.jp/epidinfo/weeklyhc.do), Hokkaido Infectious Disease (Link: http://www.iph.pref.hokkaido.jp/kansen/501/data.html) and Osaka Infectious Disease (Link: http://www.iph.pref.osaka.jp/infection/2-old.html) respectively.

Competing interests

The authors declare that they have no competing interests.

Funding

The work was supported by National Natural Science Foundation of China (Nos. 11771152, 11901203, 11971176); Guangdong Basic and Applied Basic Research Foundation (2019B151502062); China Postdoctoral Science Foundation funded project (No. 2019M662895); the Fundamental Research Funds for the Central Universities (2019MS111). The funders had no role in the design of the study, data collection and analysis, and interpretation of data or preparation of the manuscript.

Author's Contributions

KY, RL and PC designed the research; KY performed the experiments and created the figures; KY, RL and PC wrote the manuscript. RX, ZL and YCP helped with the experiments. All authors read and approved the final manuscript.

Acknowledgement

The authors are grateful to Professor Yongjun Li for the valuable discussion.

References

1. World Health Organization: Influenza (seasonal). https://www.who.int/en/news-room/fact-sheets/detail/influenza-(seasonal) (2018). Accessed 20 October 2019.
2. Zhang J, Nawata K. Multi-step prediction for influenza outbreak by an adjusted long short-term memory. Epidemiology and Infection. 2018;146(7):809-816.
3. Molinari NM, Ortega-Sanchez IR, Messonnier ML, Thompson WW, Wortley PM, Weintraub E, et al. The annual impact of seasonal influenza in the US: Measuring disease burden and costs. Vaccine. 2007;25(27):5086-96.
4. Kelso JK, Milne GJ, Kelly H. Simulation suggests that rapid activation of social distancing can arrest epidemic development due to a novel strain of influenza. BMC Public Health. 2009;9(1):117.
5. Zhong NS, Zeng GQ. Our strategies for fighting severe acute respiratory syndrome (SARS). American Journal of Respiratory and Critical Care Medicine. 2003;168(1):7-9.
6. Kilbourne ED. Influenza pandemics of the 20th century. Emerging Infectious Diseases. 2006;12(1):9-14.
7. Milinovich GJ, PhD, Williams GM, Prof, Clements, Archie C A, Prof, Hu W, PhD. Internet-based surveillance systems for monitoring emerging infectious diseases. Lancet Infectious Diseases, The. 2014;14(2):160-168.
8. Chen L, Liu R, Liu Z, Li M, Aihara K. Detecting early-warning signals for sudden deterioration of complex diseases by dynamical network biomarkers. Scientific Reports. 2012;2(1):342.
9. Liu R, Wang X, Aihara K, Chen L. Early Diagnosis of Complex Diseases by Molecular Biomarkers, Network Biomarkers, and Dynamical Network Biomarkers. Medicinal Research Reviews. 2014;34(3):455-478.
10. Richard A, Boullu L, Herbach U, Bonnafoux A, Morin V, Vallin E, et al. Single-Cell-Based Analysis Highlights a Surge in Cell-to-Cell Molecular Variability Preceding Irreversible Commitment in a Differentiation Process. PLoS Biology. 2016;14(12): e1002585.
11. Mojtahedi M, Skupin A, Zhou J, Castaño IG, Leong-Quong RYY, Chang H, et al. Cell Fate Decision as High-Dimensional Critical State Transition. PLoS Biology. 2016;14(12): e2000640.
12. Joost Lesterhuis W, Bosco A, Millward MJ, Small M, Nowak AK, Lake RA. Dynamic versus static biomarkers in cancer immune checkpoint blockade: Unravelling complexity. Nature Reviews Drug Discovery. 2017;16(4):264-272.
13. Li M, Zeng T, Liu R, Chen L. Detecting tissue-specific early warning signals for complex diseases based on dynamical network biomarkers: Study of type 2

diabetes by cross-tissue analysis. Briefings in Bioinformatics. 2014;15(2):229-243.
14. Liu R, Yu X, Liu X, Xu D, Aihara K, Chen L. Identifying critical transitions of complex diseases based on a single sample. Bioinformatics. 2014;30(11):1579-1586.
15. Liu R, Li M, Liu Z, Wu J, Chen L, Aihara K. Identifying critical transitions and their leading biomolecular networks in complex diseases. Scientific Reports. 2012;2(1):813.
16. Liu R, Chen P, Chen L. Single-sample landscape entropy reveals the imminent phase transition during disease progression. Bioinformatics, 2019. DOI: 10.1093/bioinformatics/btz758.
17. Liu R, Wang J, Ukai M, Sewon K, Chen P, Suzuki Y, Wang H, Aihara K, Okada-Hatakeyama M, Chen L. Hunt for the tipping point during endocrine resistance process in breast cancer by dynamic network biomarkers, Journal of Molecular Cell Biology, 2019, 11(8): 649–664.
18. Chen P, Liu R, Li Y, Chen L. Detecting critical state before phase transition of complex biological systems by hidden Markov model. Bioinformatics. 2016;32(14):2143-50.
19. Chen P, Li Y, Liu X, Liu R, Chen L. Detecting the tipping points in a three-state model of complex diseases by temporal differential networks. Journal of Translational Medicine. 2017;15(1):217-215.
20. Chen P, Liu R, Chen L, Aihara K. Identifying critical differentiation state of MCF-7 cells for breast cancer by dynamical network biomarkers. Frontiers in Genetics. 2015;6:252.
21. Chen P, Chen E, Chen L, Zhou XJ, Liu R. Detecting early-warning signals of influenza outbreak based on dynamic network marker. Journal of Cellular and Molecular Medicine. 2019;23(1):395-404.
22. Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review. Journal of Biomedical Informatics. 2002;35(5):352-9.
23. Mather JN. Catastrophe Theory for Scientists and Engineers. American Scientist. 1982;70(2):210-211.
24. Smith RD, Keogh-Brown MR, Barnett T, Tait J. The economy-wide impact of pandemic influenza on the UK: a computable general equilibrium modelling experiment. BMJ. 2009;339(7733):1298.

Figure title and legend

Figure 1 Schematic illustration of detecting the early-warning signal of influenza outbreak based on MST-DNM. (A), the historical records of clinic visits caused by influenza between 1 January 2009 and 1 May 2019 were collected from three regions of Japan, including Tokyo, Osaka

and Hokkaido. (B), though building a city network, weighting, and the changes of the minimum spanning tree of this network, the MST-DNM method can real time monitor the progress of the influenza and issue early-warning signals in a timely manner. (C), based on the MST-DNM method, the outbreak process of influenza could be divided into three states, i.e., the normal state, the pre-outbreak state and the flu outbreak state. The abrupt increase of MST-DNM score means the arrival of the pre-outbreak state.

Figure 2 The overall algorithm structure of MST-DNM method. First, model a city network based on its administrative divisions and the geographical relationship and map the corresponding clinic-visiting record matrix into the city network. Then, regard a week t as a candidate tipping point, weight the city network and calculate its minimum spanning tree's length as the MST-DNM score L_t. Finally, according to a logistical regression model trained by other years' dataset, Calculate the probability of L_t belonging to 1, i.e.,$P(y_t = 1 \mid L_t; \omega) = 1/(1 + exp(-L_t^T \omega))$. If this probability is greater than or equal to 0.5, the week t is considered as the tipping point. Otherwise, the week *t* is classified to the normal state, and the algorithm carries on with the week $t + 1$.

Figure 3 The predictions of annual influenza outbreak in Tokyo city between 2009 and 2019. For each year, our MST-DNM method timely issues the early-warning signal of influenza outbreak only based on the clinic-visiting information. For each figure, the x-axis represents the time evolution from the 20th week to 72nd week (roughly a seasonal-outbreak period), the y-axis represents the MST-DNM score and average number of clinic visits, respectively. The red hollow triangle represents the early-

449 warning signal detected by the MST-DNM method, and the explosion symbol is the actual outbreak point
450 of influenza, i.e., the peak of the clinic-visiting number.
451
452 **Figure 4 The dynamic evolution of the minimum spanning tree of the city network in Tokyo during**
453 **years 2013-2014.** The nodes are colored by the average number of clinic visits of the corresponding
454 district, and the thickness of the edges represents the correlations between corresponding nodes (the
455 detailed calculation is in Section Methods). It's clear that the edges become thicker before the nodes turn
456 red in week 54, which indicates that the early warning signals from our method appears before the flu
457 outbreak.
458
459 **Figure 5 The comparison result of the MST-DNM method on the presence or absence of the**
460 **minimum spanning tree in 2010.** A, the early-warning signal of a DNM method without the minimum
461 spanning tree is far away from the real influenza outbreak point, however the MST-method's is
462 measurable. B, the minimum spanning tree avoids abnormal correlations around the node 7 in week 45,
463 though which the MST-DNM method is more accurate.
464
465 **Figure 6 The performance of MST-DNM method in different critical status determination**
466 **strategies, that is logistic regression and 2-fold change threshold.** It can be seen that the MST-DNM
467 method based on logistic regression is better than that based on 2-fold change threshold. The AUC of
468 MST-DNM with logistic regression is 0.8986 while that of MST-DNM with 2-fold change threshold is
469 0.7391.
470

Table

Table 1. Algorithm for MST-DNM

Algorithm 1 MST-DNM

Input: One-year hospitalization data caused by influenza;
Output: The tipping point of the flu outbreak of this year.

Model a city network N for a specific city
Map the hospitalization data into the corresponding nodes in the network
for week t in a certian year **do**
    for each edge $e^k \in N$ **do**
        Weight the edge e^k with $W_t^k = \delta||PCC_t(v_i, v_j)| - |PCC_{t-1}(v_i, v_j||$
    end for/*obtained a weighted undirected graph N_t*/
    MST_t=Kruskal(N_t) /*obtained the minimum spanning tree using the Algorithm 2*/
    Calculate the minimum spanning tree's weight sum L_t as the MST-DNB score
    if $\frac{1}{1+e^{-L_t^T \omega}} \geq 0.5$ **then** /*the parameter ω was trained by other year's dataset*/
        the week t is deemed to the tipping point
        Break
    end if
end for

Table 2. Algorithm for Kruskal

Algorithm 2 Kruskal

Input: A weighted undirected graph G=(V,E);
Output: A minimum spanning tree of this graph.

$MST \leftarrow \Phi$
for each vertex $v \in$ V **do**
    MAKE-SET(v)
end for
sort the edges of E into nondecreasing order by weight w
for each edge$(u, v) \in$ E,taken in nondecreasing order by weight **do**
    if FIND-SET(u) $\neq$ FIND-SET(v) **then**
        $A \leftarrow A \cup (u, v)$
        UNION(u, v)
    end if
end for
return MST

Additional Files

Additional file 1 — Additional file 1.pdf

Figure S1: The predictions of annual influenza outbreak in Hokkaido city between 2011 and 2015. Figure S2: The predictions of annual influenza outbreak in Osaka city between 2012 and 2017. Figure S3: The dynamic evolution of the minimum spanning tree of the city network in Hokkaido during years 2014-2015. Figure S4: The dynamic evolution of the minimum spanning tree of the city network in Osaka during years 2017-2018.

Figures

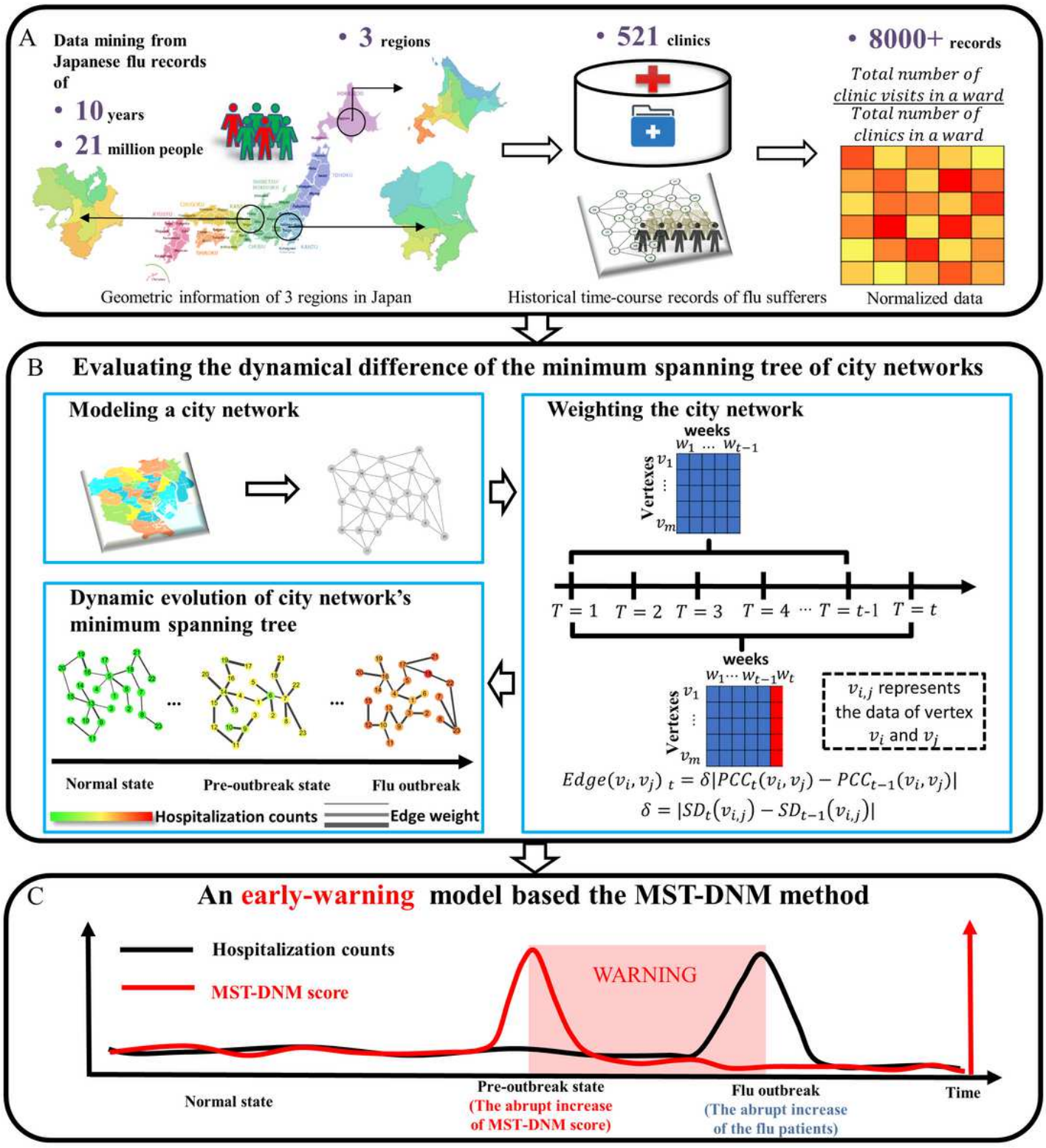

Figure 1

Schematic illustration of detecting the early-warning signal of influenza outbreak based on MST-DNM. (A), the historical records of clinic visits caused by influenza between 1 January 2009 and 1 May 2019 were collected from three regions of Japan, including Tokyo, Osaka and Hokkaido. (B), though building a

city network, weighting, and the changes of the minimum spanning tree of this network, the MST-DNM method can real time monitor the progress of the influenza and issue early-warning signals in a timely manner. (C), based on the MST-DNM method, the outbreak process of influenza could be divided into three states, i.e., the normal state, the pre-outbreak state and the flu outbreak state. The abrupt increase of MST-DNM score means the arrival of the pre-outbreak state.

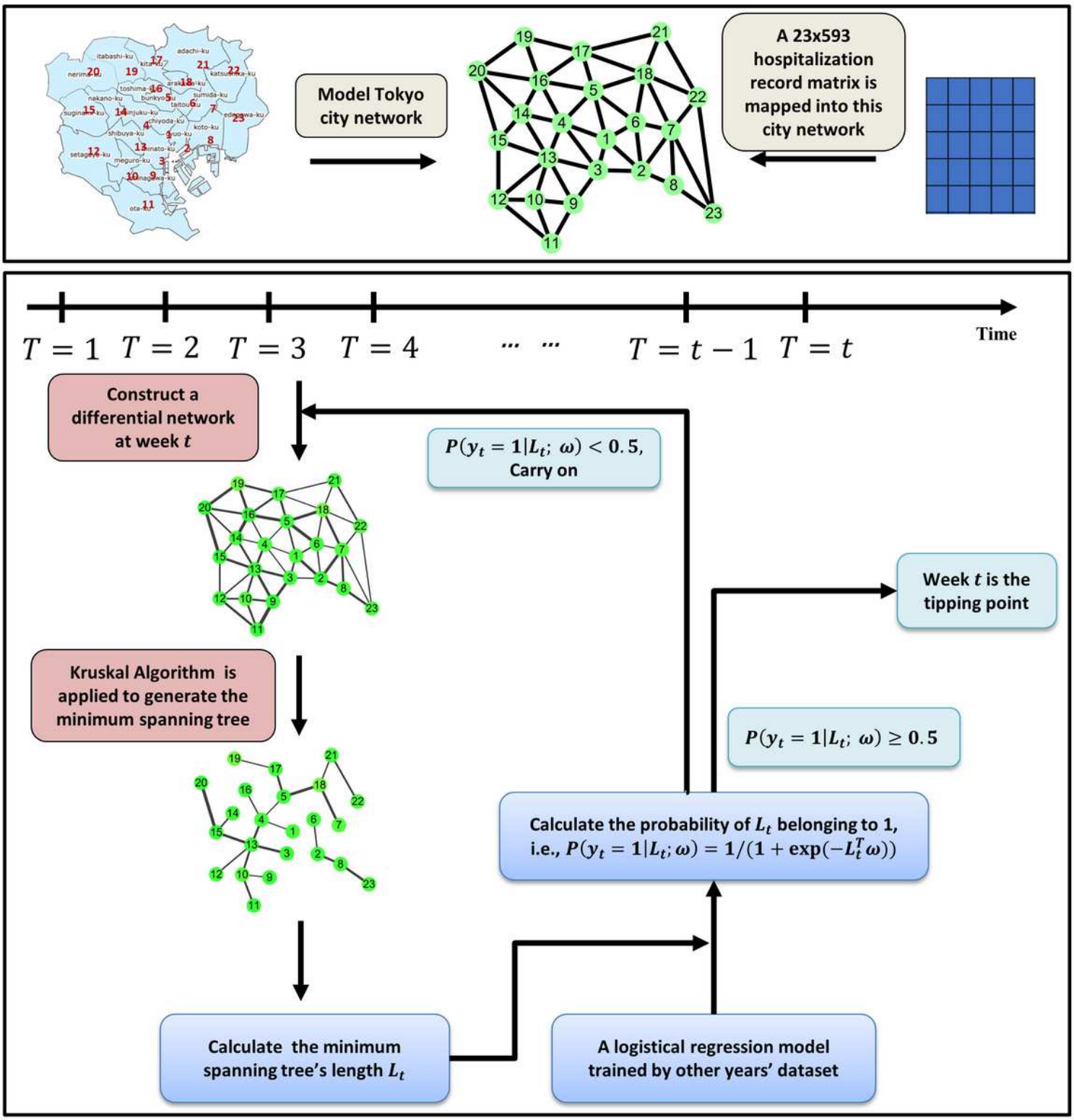

Figure 2

The overall algorithm structure of MST-DNM method. First, model a city network based on its administrative divisions and the geographical relationship and map the corresponding clinic-visiting record matrix into the city network. Then, regard a week t as a candidate tipping point, weight the city network and calculate its minimum spanning tree's length as the MST-DNM score L_t. Finally, according to a logistical regression model trained by other years' dataset, Calculate the probability of L_t belonging to 1, i.e.,P(y_t=1|L_t;ω)=1/(1+exp(-L_t^T ω)). If this probability is greater than or equal to 0.5, the week t is considered as the tipping point. Otherwise, the week t is classified to the normal state, and the algorithm carries on with the week t+1.

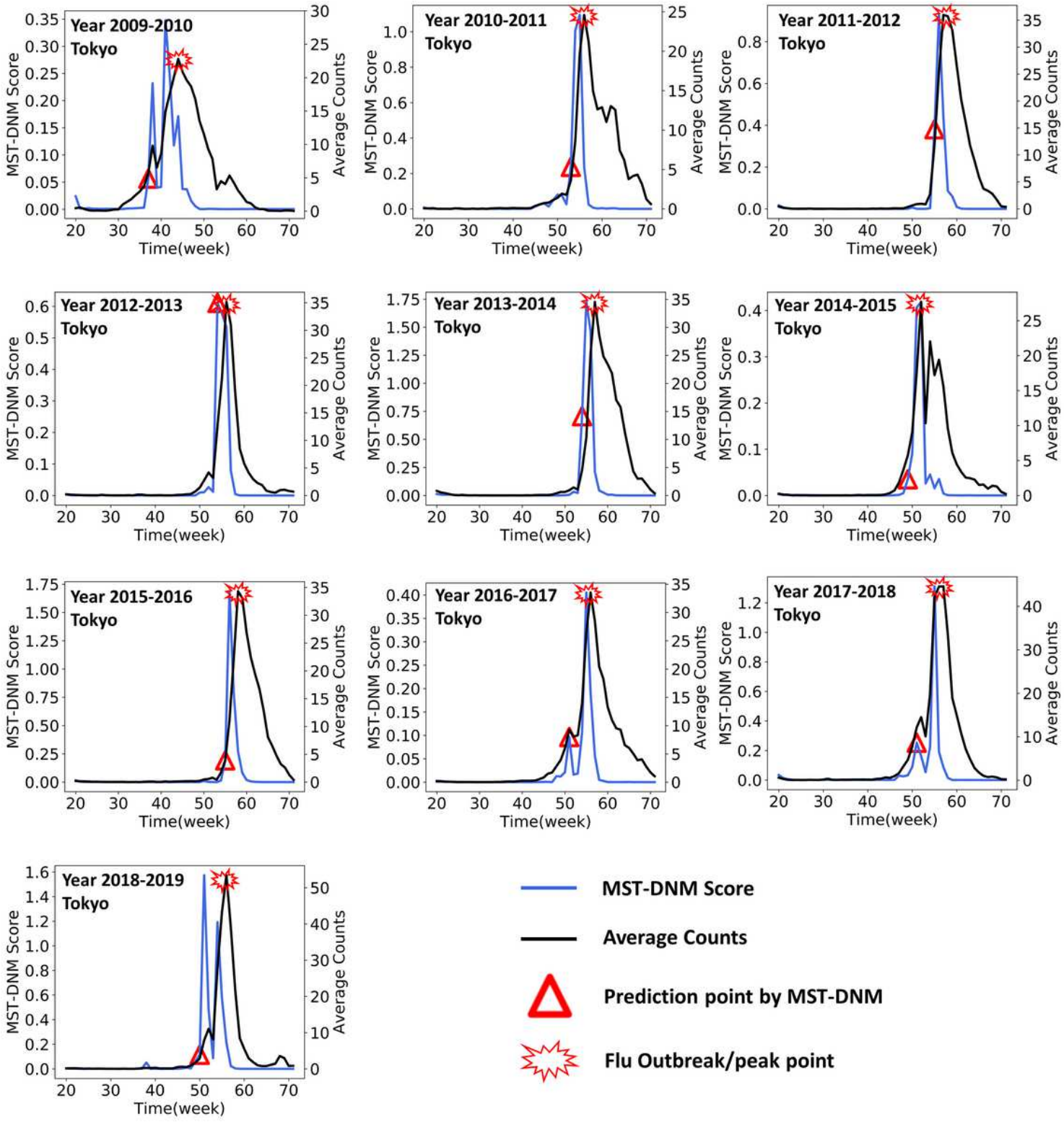

Figure 3

The predictions of annual influenza outbreak in Tokyo city between 2009 and 2019. For each year, our MST-DNM method timely issues the early-warning signal of influenza outbreak only based on the clinic-visiting information. For each figure, the x-axis represents the time evolution from the 20th week to 72nd week (roughly a seasonal-outbreak period), the y-axis represents the MST-DNM score and average number of clinic visits, respectively. The red hollow triangle represents the early-warning signal detected

by the MST-DNM method, and the explosion symbol is the actual outbreak point of influenza, i.e., the peak of the clinic-visiting number.

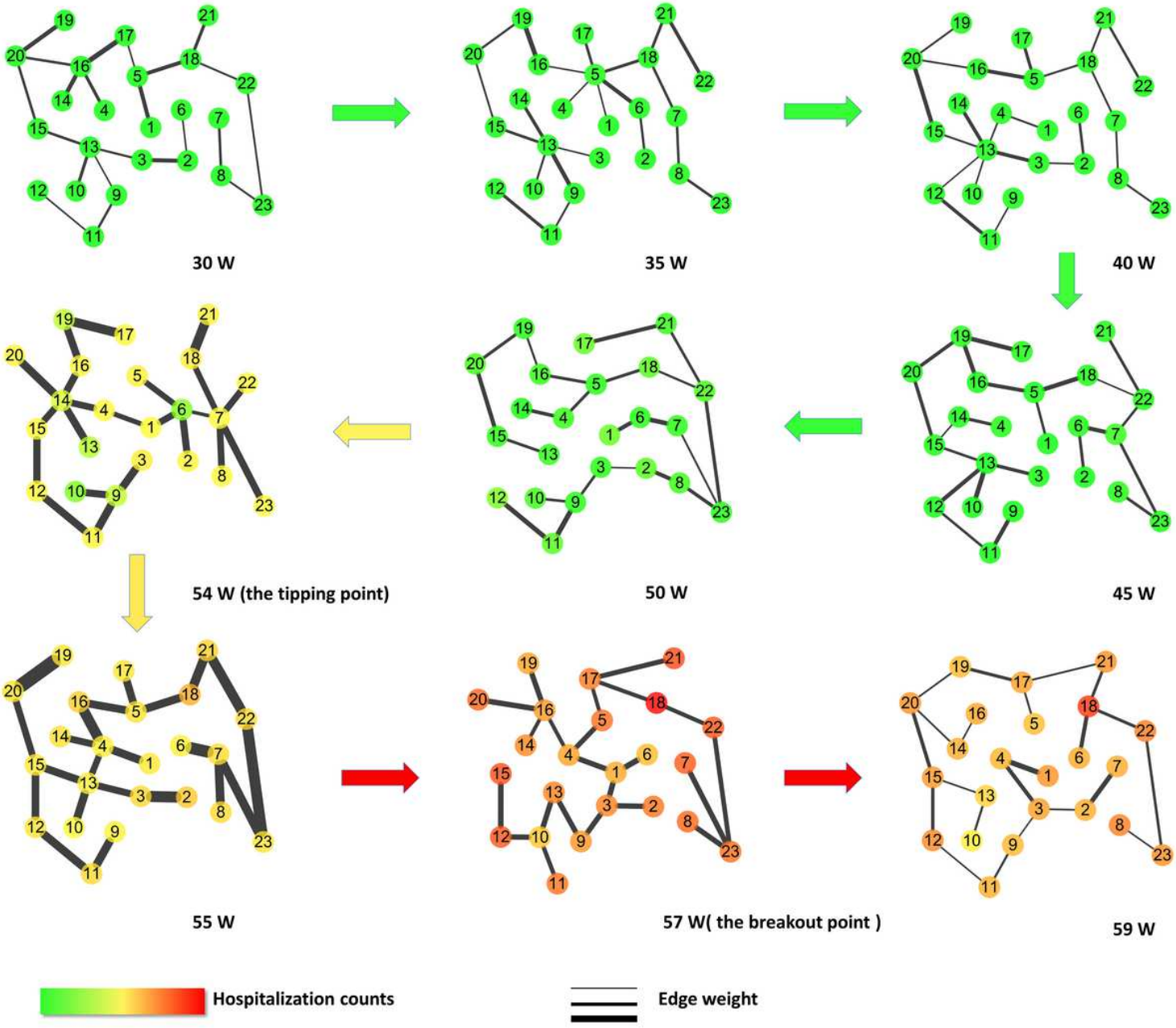

Figure 4

The dynamic evolution of the minimum spanning tree of the city network in Tokyo during years 2013-2014. The nodes are colored by the average number of clinic visits of the corresponding district, and the thickness of the edges represents the correlations between corresponding nodes (the detailed calculation is in Section Methods). It's clear that the edges become thicker before the nodes turn red in week 54, which indicates that the early warning signals from our method appears before the flu outbreak.

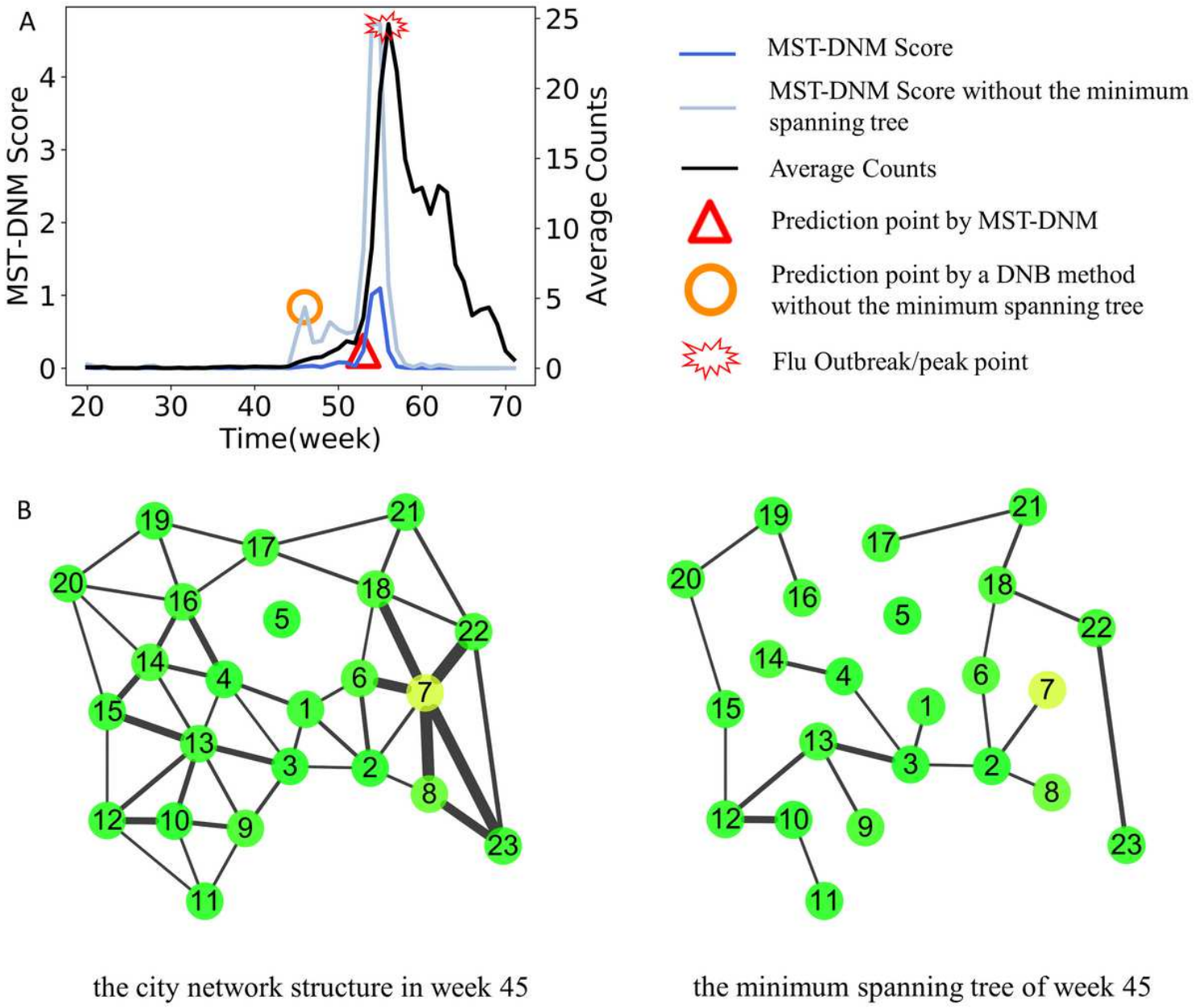

Figure 5

The comparison result of the MST-DNM method on the presence or absence of the minimum spanning tree in 2010. A, the early-warning signal of a DNM method without the minimum spanning tree is far away from the real influenza outbreak point, however the MST-method's is measurable. B, the minimum spanning tree avoids abnormal correlations around the node 7 in week 45, though which the MST-DNM method is more accurate.

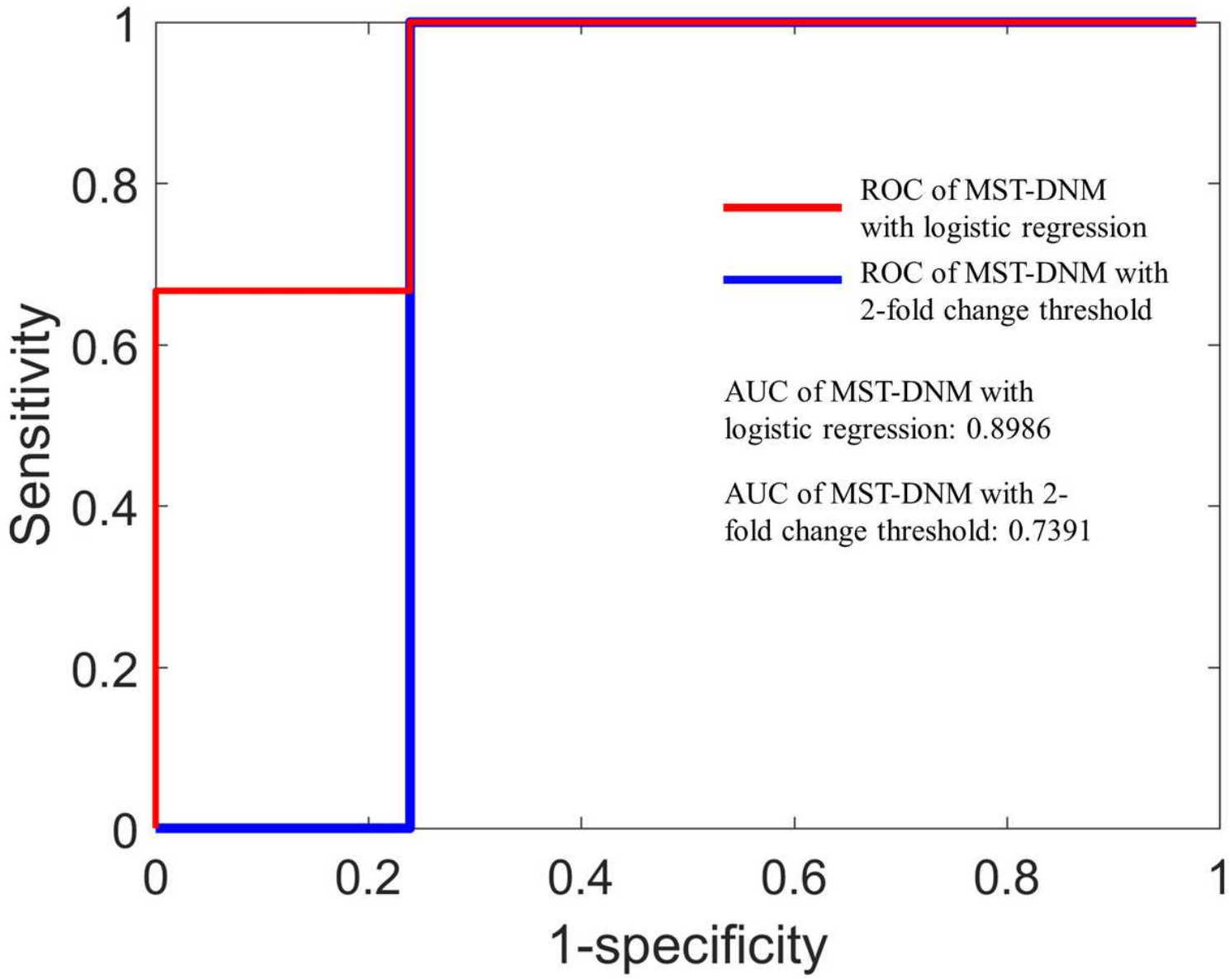

Figure 6

The performance of MST-DNM method in different critical status determination strategies, that is logistic regression and 2-fold change threshold. It can be seen that the MST-DNM method based on logistic regression is better than that based on 2-fold change threshold. The AUC of MST-DNM with logistic regression is 0.8986 while that of MST-DNM with 2-fold change threshold is 0.7391.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1.pdf