

# SCIENTIFIC REPORTS



OPEN

## A geo-computational algorithm for exploring the structure of diffusion progression in time and space

Wei-Chien-Benny Chin<sup>1</sup> , Tzai-Hung Wen<sup>1</sup>, Clive E. Sabel<sup>2</sup> & I-Hsiang Wang<sup>1</sup>

A diffusion process can be considered as the movement of linked events through space and time. Therefore, space-time locations of events are key to identify any diffusion process. However, previous clustering analysis methods have focused only on space-time proximity characteristics, neglecting the temporal lag of the movement of events. We argue that the temporal lag between events is a key to understand the process of diffusion movement. Using the temporal lag could help to clarify the types of close relationships. This study aims to develop a data exploration algorithm, namely the TrACKing Progression In Time And Space (TaPiTaS) algorithm, for understanding diffusion processes. Based on the spatial distance and temporal interval between cases, TaPiTaS detects sub-clusters, a group of events that have high probability of having common sources, identifies progression links, the relationships between sub-clusters, and tracks progression chains, the connected components of sub-clusters. Dengue Fever cases data was used as an illustrative case study. The location and temporal range of sub-clusters are presented, along with the progression links. TaPiTaS algorithm contributes a more detailed and in-depth understanding of the development of progression chains, namely the geographic diffusion process.

A geographic diffusion process is the evolution of space-time clusters of entities. Geographic diffusion processes are a scientific field of research that focuses on the movement of events, goods, information, ideas, or people through space and time<sup>1,2</sup>, that is, how do things spread from one place to another through time. In the literature, there are three types of diffusion: contagious, relocation, and hierarchical<sup>3-5</sup>. Contagious diffusion is concerned with proximate contact and is highly influenced by the friction of distance. Relocation processes involve larger leaps in spatial distance. Hierarchical diffusion is influenced by inherent hierarchies of geographical space, such as demographic, socio-economic, or the mobility structure of a region. While modeling geographic diffusion processes from the original event point locations, there are two critical points to consider: the occurrence of events, and the transmission of events.

The concept of event occurrence focuses solely on the spatial-temporal locations of events. To model the process of diffusion, first, we need to know where and when events occurred. For example, the onset date and the residential or working locations of the patients of a disease outbreak have to be recorded and analyzed in the model. Point pattern analysis methods are designed to describe the pattern of the locations of the events, such as disease cases<sup>6-8</sup>, accidents<sup>9</sup>, crime locations<sup>10,11</sup>, and disaster locations<sup>12,13</sup>. Point pattern analysis can be classified into distance-based or density-based techniques<sup>1</sup>. Distance-based techniques, such as nearest neighbor analysis, use information on the spacing of points to define a pattern<sup>14,15</sup>. Density-based techniques, such as quadrat analysis and kernel density estimation, rely on various characteristics of the frequency distribution of the observed numbers of points in regularly defined sub-regions in the study area<sup>16,17</sup>. In spatial epidemiology, kernel density estimation has been used to estimate the spatial distribution of potential risk factors<sup>7</sup>. For example, Sabel *et al.*<sup>7</sup> mapped the spatial distribution of the relative risk based on patients' residential locations, and spatial temporal trends of the groups of patients based on their age groups. In summary, point pattern analysis detects spatial clustering<sup>18-22</sup> and describes the spatial pattern of the occurrences of events.

On the other hand, the concept of the transmission of events focuses on movement through space and time. Diffusion of diseases has been studied for decades. Using Iceland as the study area, Cliff, Haggett and their team intensively worked in the 1960s on the spread of infectious disease within a closed island community in time

<sup>1</sup>Department of Geography, National Taiwan University, Taipei City, 10617, Taiwan. <sup>2</sup>Department of Environmental Science, Aarhus University, 4000, Roskilde, Denmark. Correspondence and requests for materials should be addressed to T.-H.W. (email: [wenthung@ntu.edu.tw](mailto:wenthung@ntu.edu.tw))

and space<sup>3,23–25</sup>. They attempted to link epidemic models with spatial theory and had some success in revealing underlying mechanisms of movement of disease through time and space. Aside from modeling diffusion from space-time characteristics, recent studies have used graph theory and complex network analysis to explicitly model relationships between the events<sup>26–28</sup>. Transmission relationships were modeled at various scales of networks, which included individual social networks<sup>29–31</sup>, meta-population and sub-population networks<sup>32,33</sup>, buildings network<sup>34</sup>, and cities or countries networks<sup>35,36</sup>, by converting the objects of study into nodes and the contacts or interactions between them into links. Using complex network theories to analyze the transmission relationships provides clear topological structure of contacts in terms of nodes and links for revealing the process of complex interactions<sup>37,38</sup>. These studies attempted to understand the process of the exposure to disease through an agent- or equation-based simulation, or an integrated modeling approach. For example, Meloni *et al.*<sup>32</sup> investigated infectious disease spread using a meta-population system, a network composed of subpopulations. They modeled changes in human movement behavior in response to the status of disease at the location, and simulated the transmission of disease under these scenarios. They presented the concept of an invasion tree that shows disease progression by defining a directional link from the origin to the destination subpopulations of an infection process. Disease diffusion in space and time has also been modelled by spatial dynamic models to understand the spatial pattern formation<sup>39–41</sup>. Spatial dynamic modeling is a mathematical approach that captures dynamic behaviors with patch-based spatial interaction models (i.e. cellular automata), for revealing population dynamic processes, such as the disease spreading, predator-prey interaction, and the interaction between population density and the fitness of individuals<sup>42–44</sup>. Therefore, we see that the major function of modeling the transmission between events is to understand and detect the movement process.

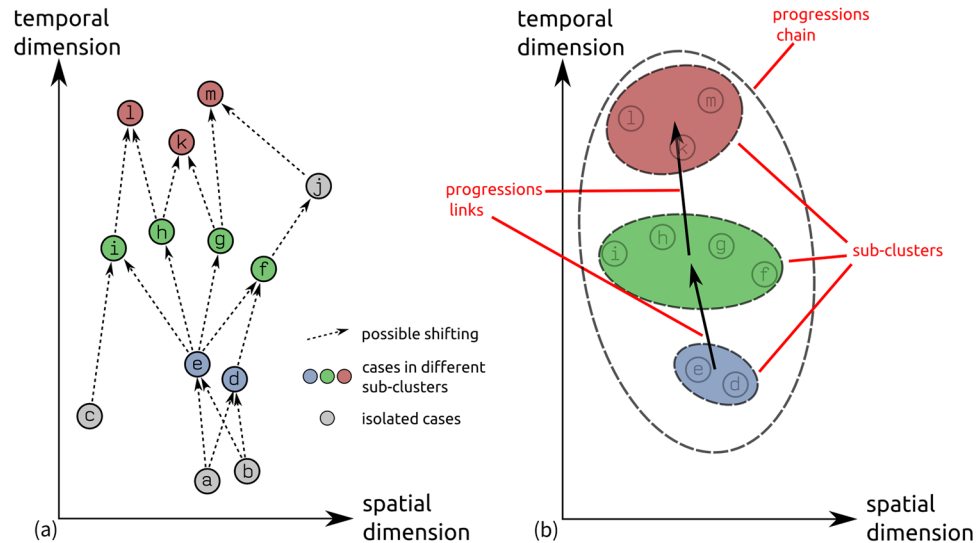
The temporal dimension is fundamental to understanding human activity<sup>45</sup>. Taking the temporal dimension into consideration is crucial for investigating space-time clustering and diffusion processes<sup>46</sup>. Previous studies of space-time point data analyses aim to investigate two spatial temporal phenomena: space-time interactions and space-time clustering. Space-time interactions determine whether a significant association between short distances in time and space exist. For example, the Knox test, a method which uses a critical space and time to determine whether a pair of events is spatially and temporally close. If the distances in space and time are correlated, a space-time interaction exists<sup>47</sup>. In the spatial epidemiological field, these tests can determine whether epidemics have contagious characteristics<sup>48–50</sup>. On the other hand, space-time clustering focuses on detecting clusters of events that are close with each other in both spatial and temporal dimensions. Space-time clustering methods can be used to detect a statistically significant excess of events occurring within a limited space-time continuum, which indicate where and when a situation becomes more serious. SaTScan, a space-time scan statistic method, which differs from space-time interaction tests, can identify when and where clusters are, and has been used to detect space-time clusters<sup>51</sup>. By considering the temporal dimension, not only the spatial location of the clusters but also the temporal periods of the clusters can be revealed.

Diffusion processes emphasize the movement of events through space and time<sup>29</sup>. But, neither space-time interactions nor space-time clustering phenomena are designed to capture the temporal differences of movements. Two events are considered related in space-time dimensions if they happen at the same place in the same time, i.e., two events happened in a small spatial range and temporal differences. However, while diffusion processes describe the spread of events through space and time, it means a temporal lag must be in between the source and target events, i.e., the second event should have occurred some time after the first event, and also not too far from it. This is the case especially in disease diffusion processes, where transmissions may experience a temporal lag for an incubation period, that is the time between infection and disease emergence. Thus a temporal lag between the transmission pairs should be considered in the understanding of disease diffusion<sup>52</sup>. To study disease diffusion processes, previous studies that used simulation approaches, including equation- and agent-based modeling, considered the shifting in temporal dimension as a key aspect in simulation models<sup>53–56</sup>. In disease diffusion simulation models, such as the susceptible-exposed-infectious-recovered (SEIR) model, a patient is exposed after physical contact with another infectious patient, and then waits for several time steps (depending on the particular disease etiology) before becoming an infectious patient<sup>53</sup>. The temporal lag effect has been considered in previous simulated diffusion studies, but the purpose of these studies was to understand the outcomes of different policy scenarios. In other words, a simulation approach cannot be used for empirical data exploration purpose.

From a data exploration perspective, the purpose of which is to identify patterns within space-time data, considering temporal lag can help clarify the relationships of space-time proximate events, and capture the progression of diffusion events. This study aims to develop a novel algorithm that utilizes temporal lag properties for understanding diffusion processes. Previous studies in diffusion have tended to focus on visualizing the spreading processes, whereas here we propose to extend this work by modeling the process, specifically to identify the space-time pattern and understand the structure and relationships between the clusters of events. First we describe the proposed algorithm, before we demonstrate its applicability using Dengue Fever data from Kaohsiung City, Taiwan.

## Method

In this section, we present a novel algorithm, namely the TrAcking Progression in Time and Space (TaPiTaS) algorithm. We decomposed the diffusion process into sub-clusters and progression links between sub-clusters. The TaPiTaS algorithm uses the spatial and temporal distance between each pair of points to identify the most probable common origin and to detect sub-clusters. A sub-cluster is formed by a group of spatially and temporally close points that are probably related to one or several common origins. By common origin, we mean the source or in an epidemiological context, the original infective agent or individual that is common to all subsequent sub-clusters. Then, sub-clusters are connected by progression links according to the spatial and temporal relationships of these sub-clusters. Finally, progression chains that are formed by several linked sub-clusters could



**Figure 1.** The illustration of sub-clusters in simplified space-time dimensions. In the figure, the cases all occur in a line on the spatial dimension, therefore the X-axis distance equals their spatial distance. (a) Illustrates the point distribution through space and time dimensions; (b) shows the sub-clusters, progression links, and the progression chain that are detected from the events distribution.

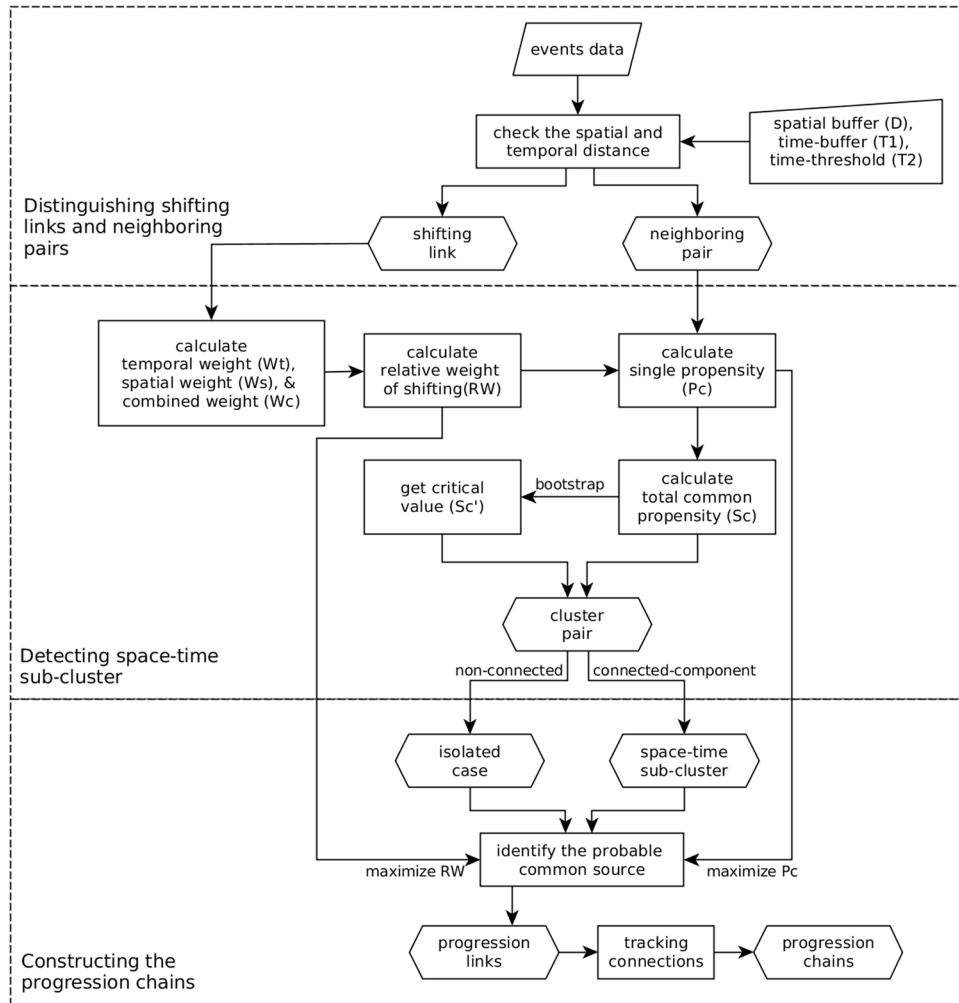
be revealed. The illustration of the sub-clusters is shown in Fig. 1, where the spatial dimensions are reduced to one dimension to show if the cases are close to each other.

The TaPiTaS algorithm is composed of three steps. The first step distinguishes the relationships of each pair of the spatially close events into two types: shifting link or neighboring pair. The second step focuses on identifying space-time sub-clusters. The third step aims to construct the progressions between sub-clusters. The algorithm framework is shown in Fig. 2. We applied the TaPiTaS algorithm to individual cases of Dengue Fever from 1998 to 2015 in Taiwan.

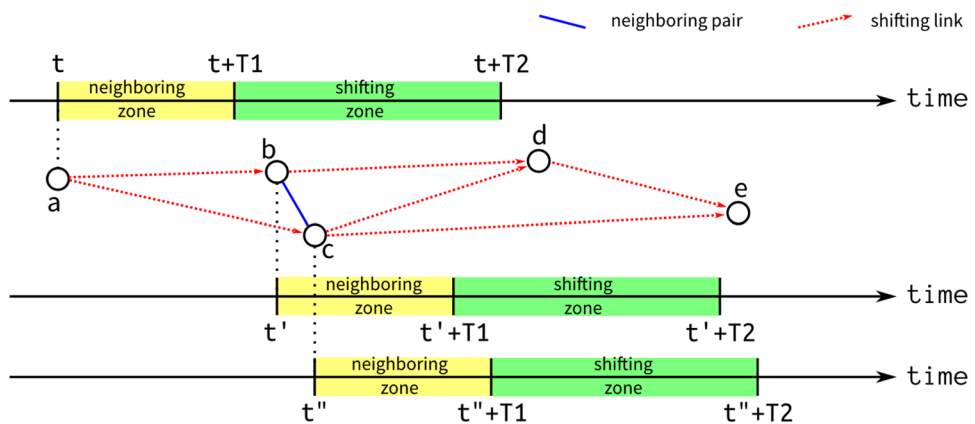
**Distinguishing shifting links and neighboring pairs.** While the spatial diffusion process is a concept describing the movements of events, it takes time to shift from one location to another. If two events happened in a same area at the same time, they can be considered as a space-time cluster of events, but not a spreading process. In point data analysis, the idea of the same area is captured by a spatial buffer zone, i.e., if one event happened within a distance buffer of another event, the two events are considered as spatially neighbors. For the temporal dimension, the idea of the same time can also be captured by a time buffer, i.e., if one event happened within a time buffer of another event, they are temporal neighbor. If two events happened in the same area, and the second event happened immediately after the time buffer from the first event, the second event can be considered as the outcome of the first event, that the first event has shifted to the location of the second event. This situation captures the concept of temporal lag, which is defined as the temporal interval that is needed for the outcome event to appear starting from the occurrence of the source event. But, if the second event happened a long time after the first event, they may be indirectly connected, but the relationship between the two events is not specified, thus, can be considered as not related.

The first step of the algorithm is to separate the spatially neighboring events into the three types of relationships (Fig. 3). To those pairs of events happened within a spatial buffer ( $D$ ), we denote the pair as a *neighboring pair* if the temporal-length (temporal lag) between the two events is shorter than or equal to a time-buffer ( $T1$ ). We denote the pair as a *shifting link* if the time-lag between the two events is longer than the time-buffer but shorter or equal to a time-threshold ( $T2$ ). And we denote the pairs with longer time-lag than the time-threshold as non-related, which are not included in the next procedure.

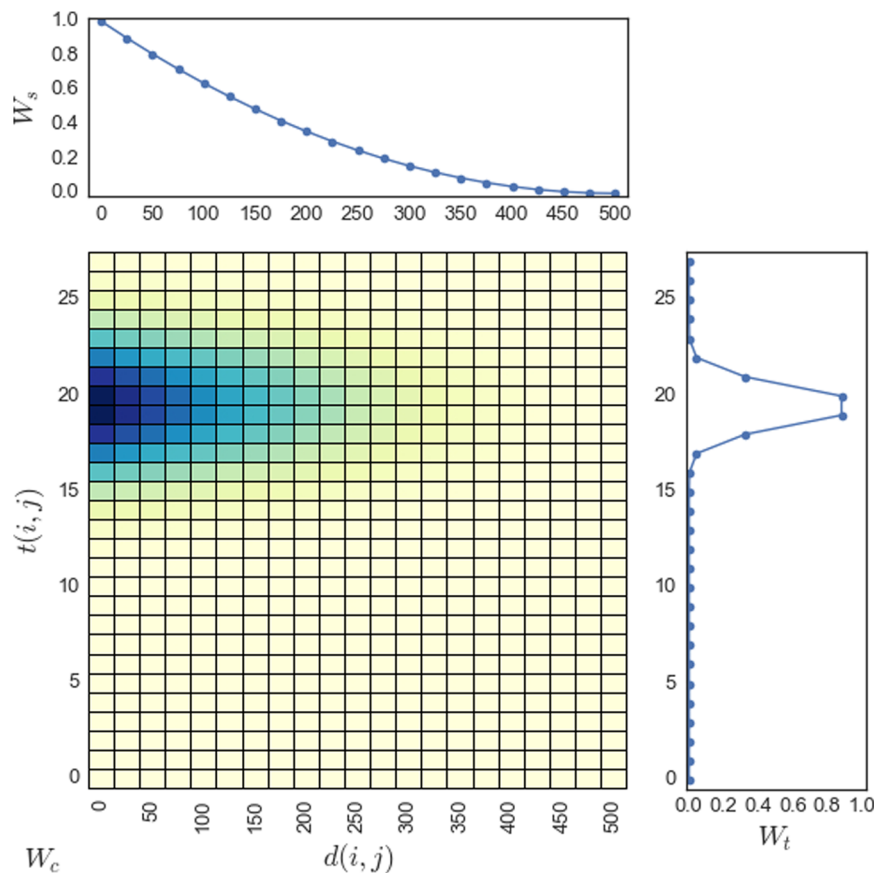
**Detecting space-time sub-clusters.** To detect space-time sub-clusters, the algorithm analyzed neighboring pairs and shifting links. Then, a group of nodes that are probably related to one or several common origins is determined by the shifting relationships. We define shifting links for capturing the opportunity of moving from one to a latter event, which is used to measure the chances if a pair of nodes have a strong common origin (or several common origins). The shifting relationship in time and space is defined as a spatial and temporal weighting function. Spatial weights decrease with the increasing distance between the two nodes and the strength temporal weights raises after  $T1$  until the middle of the range between  $T1$  and  $T2$  where the strength reaches a peak, and decreases after the middle point until  $T2$ . Therefore, the spatial weighting function is formulated as a distance-decay function with a threshold at  $D$  (Equation 2), and the temporal weighting equation is formulated as a bell shape function with the mean as the middle between  $T1$  and  $T2$  and the standard deviation as the half of the range between  $T1$  and  $T2$  (Equation 3). The results of the space-time weighting equations are illustrated in Fig. 4. The two weighting equations are normalized to range from zero (the lowest) to one (the highest). Thus, the shifting relationship between nodes is defined as a combined weight ( $W_c(i, j)$ ) of the occurrence of the link calculated based on the spatial and temporal weights with the Equations 2 and 3, respectively. We adopted the concept



**Figure 2.** The calculation framework of the algorithm.



**Figure 3.** The searching time-buffer ( $T1$ ) and time-threshold ( $T2$ ) from the cases. The onset time of illness of cases a, b, and c are  $t$ ,  $t'$ , and  $t''$ , respectively. Both onset time of case b and c fall in the shifting zone of the case a (the range between  $t + T1$  and  $t + T2$ ), thus two shifting links (a,b) and (a-c) are identified. The shifting links (b-d), (c-d), (c-e), and (d-e) are found using the same concept. The onset time of case c falls inside the neighboring zone of case b (the range between  $t'$  and  $t' + T1$ ), therefore the relationship between (b,c) is identified as a neighboring pair. There is no direct relationship between (a-d), (a-e), and (b-e), because the latter falls beyond the shifting zone of the former.



**Figure 4.** The illustration of the spatial ( $W_s$ ), temporal ( $W_t$ ), and combined ( $W_c$ ) weighting schemes. The top figure shows the spatial weight decrease with the increasing distance (calculated using Equation 2,  $D$  is set to 500); the figure on the right shows the temporal weight having a bell curve with the center around 19–20 (calculated using Equation 3, parameters are set to  $T1 = 12$ ,  $T2 = 27$ ); the bottom left figure shows the combined weight distribution using the colors, the darker the higher  $W_c$  (calculated using Equation 1).

of Mantel Index, a commonly-used space-time correlation statistic which is based on a product of normalized spatial distances and time intervals<sup>47</sup>. Thus, the combined weight (Equation 1) adopts a product of spatial and temporal weights. A higher combined weight value between nodes indicates a higher space-time association with each other.

$$W_c(i, j) = W_s(i, j) \times W_t(i, j) \tag{1}$$

$$W_s(i, j) = \begin{cases} (1 - \frac{d(i, j)}{D})^2, & \text{if } d(i, j) \leq D \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

$$W_t(i, j) = EXP \left( - \frac{\left( t(i, j) - \frac{T1 + T2}{2} \right)^2}{T2 - T1} \right) \tag{3}$$

where,  $W_s(i, j)$ ,  $W_t(i, j)$ , and  $W_c(i, j)$  are the spatial, temporal, and combined weighting functions;  $d(i, j)$  and  $t(i, j)$  are the spatial distance and temporal distance of the pair of nodes of a shifting link;  $D$  is the spatial buffer; and  $T1$  and  $T2$  are the time-buffer and time-threshold.

For each target case( $j$ ), we compare the combined weight of each of its incoming shifting link( $W_c(i, j)$ ) with the total combined weight of all of its incoming shifting links ( $\sum_{k \in I(j)} W_c(k, j)$ ). The higher the relative weight of a shifting link, the more likely the target case is shifted from the shifting link.

$$RW(i, j) = \frac{W_c(i, j)}{\sum_{k \in I(j)} W_c(k, j)} \tag{4}$$

where,  $RW(i, j)$  is the relative weight of shifting,  $I(j)$  is the set of incoming shifting links of  $j$ .

Using the relative weight of shifting, we calculate the single propensity ( $P_c(s, (a, b))$ ) of each neighboring pair ( $a, b$ ) from each common source ( $s \in I(a) \cap I(b)$ ) with Equation 5. The total common propensity ( $S_c(a, b)$ ) of each neighboring pair from all of their common sources ( $k \in I(a) \cap I(b)$ ) is calculated by Equation 6. Thus, the higher the total common propensity, the more likely the neighboring pair are shifted from one of their common sources.

$$P_c(s, (a, b)) = RW(s, a) \times RW(s, b) \quad (5)$$

$$S_c(a, b) = \sum_{k \in I(a) \cap I(b)} P_c(k, (a, b)) \quad (6)$$

where,  $P_c(s, (a, b))$  is the single propensity from one common source to a pair of nodes of a neighboring pair;  $S_c(a, b)$  is the total common propensity of a pair of neighboring nodes.

Based on the groups of total common propensity values, a non-parametric bootstrap procedure is used to identify pairs of neighboring relationships that are significantly stronger than the others. A critical value ( $S'_c$ ) is calculated as the threshold, from a bootstrapping process, to filter the neighboring pairs. Denotes  $N$  as the number of neighboring pairs, the bootstrapping process randomly sample  $N$ -pairs of neighboring pairs ( $a, b$ ), and calculates and records the mean ( $mean_s$ ) of the  $N$ -samples' total common propensity ( $S_c(a, b)$ ). The resampling process is repeated for  $M$  times, and the bootstrapped mean ( $mean_{boot}$ ) and standard deviation ( $sd_{boot}$ ) of the  $M$ -recorded means ( $mean_s$ ) is calculated for evaluating the critical value with Equation 7. Then, the neighboring pair with a common propensity ( $S_c(a, b)$ ) that is higher than or equal to the critical value ( $S'_c$ ) is defined as the *cluster pair*. The other neighboring pairs would then be neglected in the following procedures.

$$S'_c = mean_{boot} + 1.28 \times sd_{boot} \quad (7)$$

where,  $S'_c$  is the upper critical value from the bootstrapping process;  $mean_{boot}$  and  $sd_{boot}$  are the mean and standard deviation that represents the distribution of the  $M$ -times of resample means of the total common propensity values ( $mean_s$ ). To calculate the upper bound of the 80% interval of the distribution (two-tails), the standard deviation is multiplied by 1.28 and summed to the mean of the  $M$  samples.

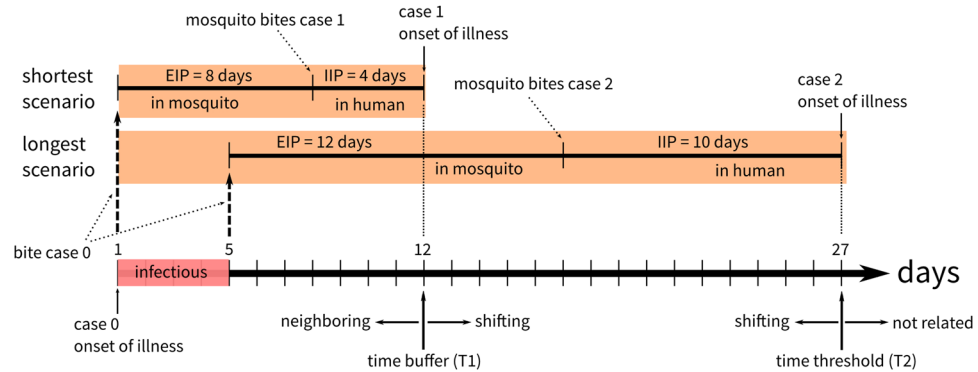
After all of the neighboring pairs are evaluated, we can construct a network where the nodes are events, and two nodes are connected if a cluster pair exists between them. We search for the connected components within the network, which represent the subgroup of nodes, in which each node has a connecting path to any node within the same subgroup. Each connected component is identified as a space-time sub-cluster. Therefore, a sub-cluster is composed of a group of events which are connected by a bunch of cluster pairs, indicating that they are more likely to have one or several common sources.

**Constructing the progression chains.** The progressions represent the connections between sub-clusters, that show how the sub-clusters influence one another, and the direction of the diffusion process. Two scales of progressions are included in this part: the progression links, and the progression chains. After all cluster pairs are found, the most probable common origin (the source(s) of the shifting links with the max  $P_c(s, (a, b))$ ) of each cluster pair are revealed and defined as common links. A *progression link* is constructed between two sub-clusters that has at least one common link exists between them, which represents the progressions from one sub-cluster to another. And, the sub-clusters that are connected with each other form a *progression-chain* (Fig. 1b), whereas the other sub-clusters that are not linked with any sub-clusters are called *isolated sub-clusters*.

### Application: the sub-clusters and progressions of Dengue Fever in Taiwan from 1998 to 2015

To test and demonstrate our TaPiTaS algorithm for understanding and visualizing the diffusion process, we used Dengue Fever data in Kaohsiung City, Taiwan from 1998 to 2015. Located in East Asia, Taiwan straddles tropical and subtropical zones. The tropical weather pattern of Kaohsiung City, hot temperature and high humidity in summer, provides suitable habitats for the vector of Dengue Fever (mainly the *Aedes aegypti* mosquitoes)<sup>57,58</sup>. Moreover, Kaohsiung City, which has the largest harbor and an international airport in Taiwan, is a major East-Asian transport hub that has a high volume of travelers from South-East Asia. This increases the opportunity of Dengue Fever importation to Taiwan<sup>59,60</sup>. The number of Dengue Fever cases in Kaohsiung City shows an annual cyclical pattern, which mainly starts in the early summer and ends in the late winter of the next year<sup>57</sup>. Therefore, we have separated the data in this study to start from April 1st of each year to March 31st of the following year. The locations of past outbreaks are mostly concentrated in southern Taiwan (including Kaohsiung City) due to the temperature suitability for the mosquitoes breeding<sup>61,62</sup>. Therefore, the aim of the case study is to detect sub-clusters from the annual Dengue Fever cases and identify processes between the sub-clusters.

Dengue Fever is a vector-borne disease, with a human-mosquito-human transmission cycle. There is a temporal lag between the time a case is infected and the time when the case become infectious. The infectious period after the first symptoms appear is about 4–5 days; the extrinsic incubation period (EIP) for mosquitoes is about 8–12 days; and the intrinsic incubation period (IIP) for humans is about 4–10 days<sup>63</sup> (Fig. 5). Therefore, the minimum temporal lag between a pair of related cases is 12 days (the first patient infects the mosquitoes on the first day, 8 days of extrinsic incubation period, following by 4 days of intrinsic incubation in the second patient), whereas the maximum time-lag is about 27 days (the first patient infects the mosquitoes on the 5th day, following by 12 days and 10 days of the extrinsic and intrinsic incubation period, respectively).



**Figure 5.** The time parameters settings were designed to capture the ranges of infectious period, extrinsic incubation period (EIP), and intrinsic incubation period (IIP) of Dengue Fever. The time buffer ranges from the first day when the case 0 is sick to the minimum days for EIP and IIP. The time threshold ranges beyond time buffer to the maximum days for EIP and IIP.

year	no. cases	no. SC	SC size	SC duration	no. iso-SC	no. PL	no. chains	chain size	chain duration
1998	113	5	3 (0)	5 (4)	2	2	1	3.0 (0)	66.0 (0)
1999	20	1	—	—	1	0	0	—	—
2000	—	—	—	—	—	—	—	—	—
2001	220	9	8 (9)	14 (15)	3	8	1	6 (0)	81 (0)
2002	4671	276	3 (1)	7 (6)	52	198	39	2 (0)	36 (22)
2003	34	2	2 (0)	2 (2)	2	0	0	—	—
2004	56	1	—	—	1	0	0	—	—
2005	96	6	4 (1)	6 (4)	4	1	1	2 (0)	19 (0)
2006	955	102	3 (1)	4 (4)	21	57	25	3 (1)	29 (18)
2007	168	7	4 (1)	12 (9)	3	2	2	2 (0)	31 (1)
2008	417	31	3 (1)	5 (6)	18	7	6	2 (0)	23 (7)
2009	751	75	3 (1)	6 (6)	11	52	17	3 (1)	36 (16)
2010	1044	129	3 (1)	5 (4)	36	68	28	3 (1)	35 (16)
2011	1158	119	3 (1)	5 (6)	34	61	25	2 (0)	34 (16)
2012	478	37	3 (1)	5 (4)	18	15	6	2 (0)	26 (9)
2013	64	3	3 (0)	3 (1)	1	1	1	2 (0)	17.0 (0)
2014	15011	433	3 (1)	5 (4)	99	292	56	2 (0)	39 (19)
2015	19520	484	3 (1)	5 (4)	115	356	44	3 (1)	50 (36)

**Table 1.** The descriptive statistics of the diffusion progression from 1998 to 2015, including cases, sub-clusters(SC), progression links(PL), and chains. The SC size represents the median number of cases within a sub-cluster; chain size represents the median number of sub-clusters within a chain; the SC duration represents the median temporal duration of the sub-clusters, which measures the number of days between the earliest and the latest cases within a sub-cluster; the chain duration represents the median temporal duration of the chains. The numbers in brackets are the median absolute deviation (MAD) of the corresponding columns.

Our data was provided by Taiwan Centers for Disease Control (Taiwan CDC), which records the daily number of Dengue Fever cases in each basic statistical unit (BSU) in Kaohsiung City, separated into imported (to Taiwan) and local indigenous cases, based on the epidemiological investigation records. The Kaohsiung City area includes the previously named Kaohsiung City and Kaohsiung County until 2010. Only the indigenous case data was used in this study to eliminate the external importation noise from the local diffusion process. On average for the whole of Taiwan, each BSU contains 400 people. There are three main parameters in the our TaPiTaS algorithm: a spatial distance parameter, and two temporal parameters. In this case study, the spatial neighboring parameter ( $D$ ) was set to 500 meters following Hsu and Tsai<sup>64</sup>. Based on the intrinsic and extrinsic incubation periods of Dengue Fever<sup>63</sup>, the time-buffer( $T1$ ) was set to 12 days; and the time-threshold parameter ( $T2$ ) was set to 27 days. Experimentally, the bootstrapping value converged after 99 iterations.

**Descriptive statistics of events, pairs of close events, sub-clusters, and chains.** Table 1 shows the descriptive statistics of the diffusion progression by year. The total number of cases varied between years: six out of the 18 years have less than 100 cases; in 2014 and 2015, the number of cases exceeded 10,000. The number of sub-clusters are related to the number of cases in the year, but are not always proportional to the number of cases. For example, 2010 had less cases than 2011, but more sub-clusters were detected. The sub-cluster size (SC size)

and duration (SC duration) shows the overall extent and the duration of continuity of sub-clusters. Sub-cluster size is measured by the number of cases in every sub-cluster, whereas the duration is measured by the days between the first case and the last case in each sub-cluster. Regardless of the significant differences in the total number of cases, the median number and the median absolute deviation of the sub-cluster sizes and durations over the 18 years are similar: sub-clusters consist of about 2 to 4 cases, and the duration median is 5 days (with 4 days MAD). MAD is the abbreviation for median absolute deviation<sup>65</sup>, which better describes our range than the normal standard deviation.

The number of isolated sub-clusters (iso-SC), progression links (PL), chains, and the characteristics of progression chains including the sizes and duration are also shown in Table 1. The chain size is measured by using the number of sub-clusters in each chain to show the extent of chains in the year; the chain duration is measured by using the temporal difference between the first case and the last case in each chain to show the temporal continuity in each year. Similar to the sub-cluster sizes and durations, the chain sizes and durations are also similar through the 18 years regardless of the differences in terms of the total size of cases. The chain sizes consist of 2 to 3 sub-clusters and the duration median is one month (with 2 weeks MAD).

**Visualizing the processes of diffusion.** From the descriptive statistics analysis (Table 1), in three years, 2002, 2014, and 2015, Kaohsiung City experienced the severest epidemics in the past 70 years<sup>65,66</sup>. The total number of confirmed Dengue Fever cases in these three years were 4671 cases, 15011 cases, and 19520 cases, respectively. The progression structures of large-scale Dengue transmission in these epidemic years can be distinctly illustrated and investigated quantitatively for demonstrating the functionality of our proposed algorithm. Therefore, in the following discussion, we focused on the three years. In 2002, our algorithm found 276 sub-clusters and 224 of these formed 39 progression chains. In 2014, 435 sub-clusters were detected, and 336 of these sub-clusters were found in 56 progression chains. In 2015, 484 sub-clusters were detected, and 368 formed 44 progression chains.

The spatial distribution of the cases, sub-clusters and progression links in 2002, 2014, and 2015 are shown in Fig. 6. The colors indicate the progression chains to which they belong. The sub-clusters are presented as standard ellipses using the XY coordinates of cases to determine the standard distances. The width of the progression links indicates the number of shifting links that connect the two sub-clusters. Cases were distributed throughout the city for the three years (Fig. 6a,d,g). In 2002, the algorithm found a significant spatial separation between the progression chains (Fig. 6a). Figure 6b shows that the progression chains were differentiated into ellipses by our algorithm based on the temporal differences between the cases. In order to show the evolution and strength between the sub-clusters, the progression links were mapped in Fig. 6c.

The spatial distribution of the progression chains in 2014 and 2015 were different from the year 2002, in that an extremely large sub-cluster appeared and covered the most highly populated area of the city (Fig. 6e,h). This is because the cases were not concentrated in a small area over the small temporal duration, but distributed over a larger region within the spatial search radius. Different from 2002, some smaller progression chains overlapped with the large sub-cluster in 2014. However, most of the smaller overlapping sub-clusters were linked with another large sub-cluster in 2015. Comparing the 2014 results to the 2015 results, the 2015 analysis revealed more sub-clusters but less progression chains. The chain sizes in 2015 were larger than those in 2014, indicating that the sub-clusters in 2015 were more connected to one another. For example, in the southern part of Kaohsiung City (lower part of the map), five progression chains were detected in 2014, but only one progression chain was found in 2015.

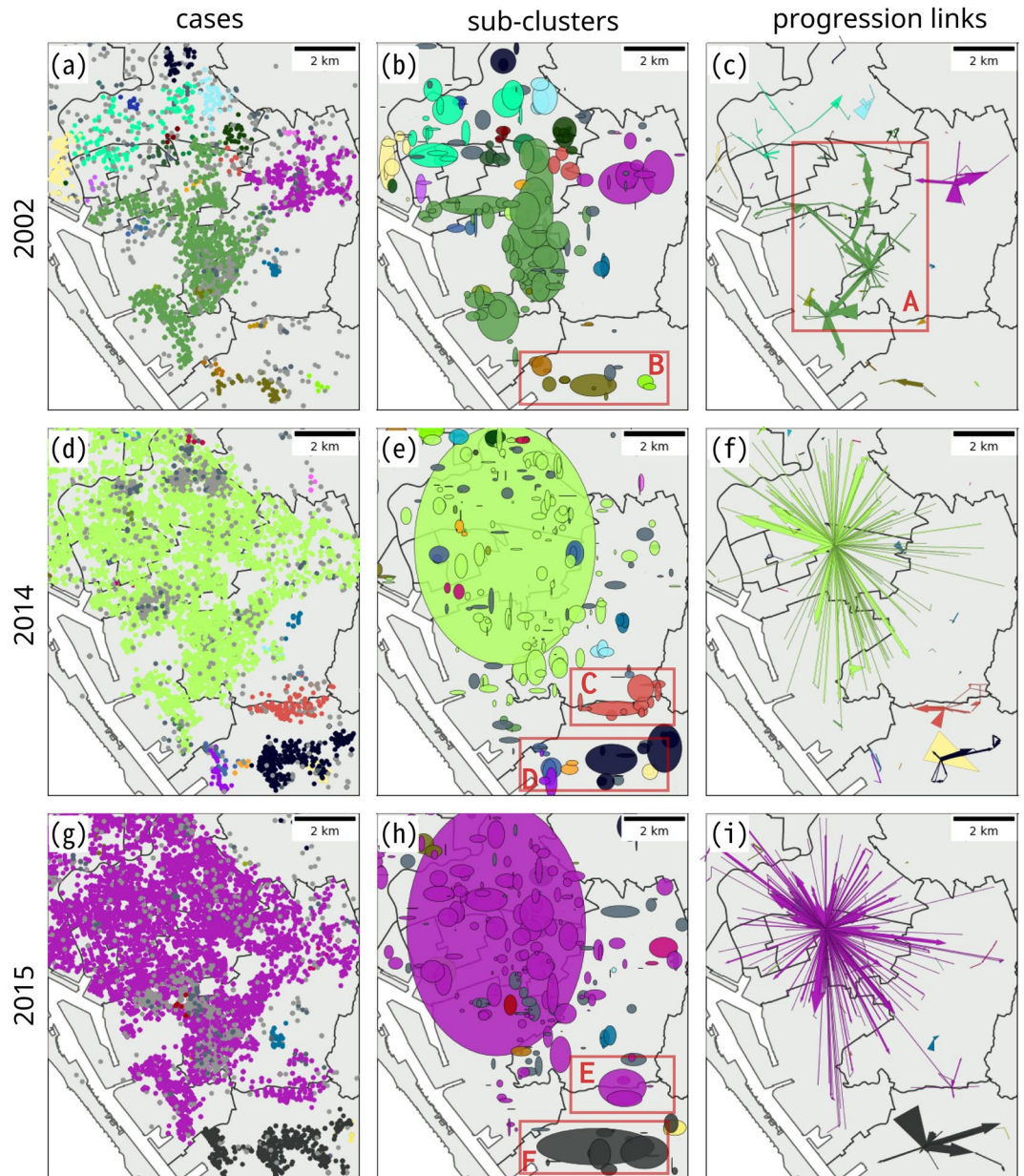
## Discussion

When exploring diffusion processes, especially for disease diffusion, the actual interactions between events (cases) are normally not known, since information about the transmission process (sub-cluster evolution over time) between large numbers of cases is difficult to measure retrospectively. The location and temporal case information is normally the only available data. To uncover any diffusion process from this type of data, we have developed the TaPiTaS algorithm for exploring and visualizing space-time point data to show the process of diffusion. TaPiTaS is a novel algorithm that utilizes the temporal lag within the diffusion process and the spatial distance between events to detect the spatial-temporal sub-clusters and to uncover the development of progression chains. Recall that a sub-cluster represents events that have high propensity to have common origins in the diffusion process. A progression chain represents the linked sub-clusters. Thus TaPiTaS is an important conceptual contribution to cluster structure and sub-cluster evolution understanding.

We anticipate that this algorithm can be used to explore diffusion phenomena, in which temporal lag is a key to the movement of the diffusion process. Unlike previous methods for detecting clusters that only included a time-window to capture the temporal proximity, our algorithm adds a lagged time-window (between T1 and T2) to capture the shifts between events, and weights the temporal lags to search for the sub-clusters and progressions. This enables us to distinguish the sub-clusters from the spatially and temporally clustered events.

To demonstrate the analysis process and algorithm outputs, we presented a case study, of the annual Dengue Fever diffusion process in Kaohsiung City, Taiwan, from 1998 to 2015. In the case study, despite varying sizes of epidemics in different years, we detected similar size sub-clusters with a median of 2 to 4 cases. The reason that most of the sub-clusters are composed of small numbers of cases is because the algorithm distinguishes the relationships between space-time proximate events using the lagged time-window, and the algorithm filtered neighboring pairs using a critical value to ensure that only the pairs of neighbors with compelling high total common propensity are used to detect sub-clusters. In other words, by connecting the events that happen in the same place at the same time (with buffer zones in space and time), our TaPiTaS algorithm aggregates cases into small groups, namely the sub-clusters, which could form part of a larger space-time cluster. Moving on to discuss the progression chains, by connecting the sub-clusters according to the evolution of shifting links, TaPiTaS uncovers the movement progression in space-time dimensions, namely the processes within a space-time cluster.





**Figure 6.** Spatial distribution of cases (left), sub-clusters (middle), and progression links (right) in 2002 (first row), 2014 (second row), and 2015 (third row). The colors differentiate the progression chains. The sub-clusters were determined by the standard ellipses' standard distances (long and short axes) and the orientation. The number of shifting links that form the progression links were used to determine the width of the progression links. The maps were created using Matplotlib (2.0.0) package in Anaconda (4.0.0, Python 2.7.11 version, url: <https://www.continuum.io/anaconda-overview>).

Our results differentiate diffusion structures in time and space among the severest epidemic years: 2002, 2014, and 2015. It implies Dengue Fever epidemics for these years in Kaohsiung could have been triggered by different sources of infection, driving forces of transmission, and the effectiveness of intervention measures. In 2002, its diffusion structure indicated the epidemic circulated around the different district-level administration areas. The district health authority is the basic operation unit for disease control and prevention. Thus, it implies that the peripheral areas may lack consistent intervention measures. The sub-clusters with green color originated from the box-A (Fig. 6c), near to the boundary of the Fengshan district, a frequent Dengue-epidemic region<sup>57,67</sup>, then spreads south-west, north-west, and north. Fengshan district is a satellite city of the Kaohsiung metropolitan area that has a high population density but its socio-economic level is relatively lower than the neighboring central business district (CBD) of Kaohsiung City. Recent studies have shown the areas with high urbanization levels, high population density, and low social-economic status would increase the risk of Dengue diffusion and tend to become the sources of diffusion<sup>52</sup>. Moreover, there is a group of small and less connected sub-clusters appearing

in the southern part (Fig. 6b, box-B) of the map, indicating that the areas may also be vulnerable to Dengue Fever and may require more attention in the future.

The epidemic progressions in 2014 and 2015 share some common diffusion structure characteristics. First, these two Dengue epidemics are composed of two major progression chains. One of the progression chains contains the largest size of sub-cluster, and these cases in the sub-cluster were mostly located in epidemic areas (Fig. 6e,h). Second, two smaller groups of progression chains appear in the southern part of Kaohsiung City, one of the progression chain groups contains the orange progression chain in 2014 (Fig. 6e, box-C), and part of the purple progression chain in 2015 (Fig. 6h, box-E). The other contains a group of progression chains in 2014 (Fig. 6e, box-D), and a black progression chain in 2015 (Fig. 6h, box-F). These locations are approximately the same vulnerable areas as identified in 2002 (Fig. 6b, box-B). Therefore, comparing these progression chains could reveal geographic epidemiological links among these three years. On the other hand, we also identified different progression directions in 2014 and 2015. The progression links in 2014 pointed northwest are stronger than the other directions; however, in 2015, some of the progression links pointed south are stronger. In 2014, the epidemic sub-clusters mainly diffused northward and north-westward, whereas in 2015, some significant southward spreading of sub-clusters was also observed.

Using standard ellipses to represent sub-clusters, and arrows between them to show the progression chains, the diffusion process is illustrated. Previous studies on space-time clustering issues using the space-time scan statistic (SaTScan) visualized the clustered area with a circle, and described the clustering periods of each cluster<sup>68,69</sup>. This visualization method shows a clear location and the magnitude of the detected clusters. But it ignores the process of cluster development. Other space-time clustering studies used kernel density techniques to visualize the clustered area in space-time cube plots superimposed on a 2D X,Y map<sup>21,22</sup>. Event temporal mobility was illustrated on the Z or 3rd dimension, and thus the event migration could be shown on a three-dimensional figure. However, without the ability to interact with a static 3D image, it is confusing to understand cluster progression. Furthermore, using density based techniques, simple temporal proximity relationships were assumed, not calculated from the data whereas in our TaPiTaS algorithm, the temporal lag between cases is explicitly included in the calculation procedure. In our study, by using standard ellipses to visualize sub-clusters, we suggest that readers can get a better understanding about the shape and directions of sub-clusters; using arrows to link the sub-clusters and to represent the progression links with the day of appearance of the sub-clusters, the process of diffusion can be revealed.

Profiling diffusion structures of Dengue Fever epidemics can provide important clues for health authorities to implement spatially-targeted intervention measures. Our proposed algorithm reveals three key geospatial characteristics of Dengue diffusion by identifying progression links: the source areas, the target areas, and the linkage between epidemic areas. The source areas of diffusion may represent the areas that are suitable for the breeding of disease vector. Eliminating the habitats of mosquitoes could be beneficial for the source area and also its adjacent areas. It would produce the trans-boundary externalities in disease control<sup>70,71</sup>. The target areas could be the regions with spatial risk factors of Dengue transmission. The areas with these risk factors are expected to be more vulnerable to the Dengue Fever transmission, including high urbanization levels, low social-economic status, favorable weather conditions and previous epidemic records<sup>52,72,73</sup>. The linkage between epidemic areas indicates the spatial interactions or communication through human movement. It also implies the possible route of the virus diffusing from one to another region. A longer progression linkage may cause a large-scale epidemic through long-distance human moving behaviors<sup>50,66,74</sup>.

Our TaPiTaS algorithm has several limitations. First, when a large amount of events happens within a short spatial distance of each other, over a short temporal period, such as occurred in the case study in 2014 and 2015 with more than 15,000 cases per year, our algorithm could not fully capture the diffusion process. Instead of identifying sub-clusters as separate events, our algorithm grouped them as a giant sub-cluster and which became a common source for the sub-clusters at different places. We think that this situation requires further investigation to understand the mechanism of the extreme sub-clusters or progression chains. Second, no significance evaluation procedure is included in our TaPiTaS algorithm. While clusters could happen by chance under a complete spatial randomness condition, the detected sub-clusters and the progression links could be a result of a space-time random distribution. To overcome this limitation, a Monte-Carlo significance test could be included to evaluate the level of significance of each sub-cluster and progression link. Third, the underlying population distribution is not considered. Other methods, such as SaTScan, consider the distribution of population at risk to capture the spatial inhomogeneity of population, and perform a risk normalization procedure in the calculation. Risk normalization is not included within our current algorithm, but it may be done in the data preparation stage before the algorithm, or in the results evaluation stage using additional calculation after the algorithm. Fourth, the definition of spatial proximity is measured only by the straight line distance between events, which simplifies the concept of distance. The measurement of distance is a key issue in most spatial analysis models and methods. Some alternatives, such as street network distance, time needed for movement, or cost of movement, could be used alternatively to measure the spatial proximity between events. Fifth, we used two equations to calculate the spatial and temporal weight separately (Equations 2 and 3), and multiplied them to calculate the integrated weight (Equation 1) in the algorithm. The equations may not be suitable for all events of interest in different studies. Therefore, the parameters for weighting spatial and temporal distance, and for the combined weight of shifting links could be modified according to the needs of the study subject.

## Conclusion

Temporal lag is essential in the understanding of diffusion processes. We have proposed a novel algorithm, TaPiTaS, that utilizes the spatial distance and temporal interval between events, to explore and to visualize the diffusion process. This algorithm can be used to explore point data with timestamps, and outputs sub-clusters, progression links, and progression chains to show the process of diffusion. As a method for space-time point

data exploration, TaPiTaS differentiates the relationships of events and detects the sub-clusters of events that are immediately proximate to each other. TaPiTaS then identifies the directional links between sub-clusters, which represents diffusion progressions. By additionally visualizing the detected sub-clusters and progression links, our TaPiTaS algorithm contributes a more detailed and in-depth understanding of the geographic diffusion process then currently exists. In summary, we propose our TaPiTaS algorithm as a tool for uncovering the evolution of space-time clusters of entities.

**Data availability.** The data that support the findings of this study are available from Center for Disease Control, Taiwan, under the Taiwan Open Government Data License, version 1.0 (<http://data.gov.tw/license#eng>). The datasets are available in the Daily reported Dengue Fever cases since 1998 repository (dataset url: <http://data.gov.tw/node/21025>, direct download url: [http://data.gov.tw/iisi/logaccess/61136?dataUrl=http://nidss.cdc.gov.tw/download/Dengue\\_Daily\\_EN.csv&ndctype=CSV&ndcnid=21025](http://data.gov.tw/iisi/logaccess/61136?dataUrl=http://nidss.cdc.gov.tw/download/Dengue_Daily_EN.csv&ndctype=CSV&ndcnid=21025)).

## References

- Haggett, P., Cliff, A. D. & Frey, A. E. *Locational analysis in human geography*. (Wiley, 1977).
- Sabel, C. E., Pringle, D. & Schaerstrom, A. Chapter 7: Infectious Disease Diffusion. In *A Companion to Health and Medical Geography* 111–132 (Wiley-Blackwell, 2010).
- Cliff, A. D., Haggett, P. & Smallman-Raynor, M. *Island Epidemics*. (Oxford University Press, 2000).
- Nekola, J. C. & White, P. S. The distance decay of similarity in biogeography and ecology. *J. Biogeogr.* **26**, 867–878 (1999).
- Meade, M. S. & Emch, M. *Medical geography*. (The Guilford Press., 2010).
- Bithell, J. F. An application of density estimation to geographical epidemiology. *Stat. Med.* **9**, 691–701 (1990).
- Sabel, C. E., Gatrell, A. C., Löytönen, M., Maasilta, P. & Jokelainen, M. Modelling exposure opportunities: estimating relative risk for motor neurone disease in Finland. *Soc. Sci. Med.* **50**, 1121–1137 (2000).
- Porphyre, T. *et al.* Vulnerability of the British swine industry to classical swine fever. *Sci. Rep.* **7**, 42992 (2017).
- Xie, Z. & Yan, J. Kernel Density Estimation of traffic accidents in a network space. *Comput. Environ. Urban Syst.* **32**, 396–406 (2008).
- Gerber, M. S. Predicting crime using Twitter and kernel density estimation. *Decis. Support Syst.* **61**, 115–125 (2014).
- Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P. & Tita, G. E. Self-Exciting Point Process Modeling of Crime. *J. Am. Stat. Assoc.* **106**, 100–108 (2011).
- Woo, G. Kernel estimation methods for seismic hazard area source modeling. *Bull. Seismol. Soc. Am.* **86**, 353–362 (1996).
- Galindo, I., Romero, M. C., Sánchez, N. & Morales, J. M. Quantitative volcanic susceptibility analysis of Lanzarote and Chinijo Islands based on kernel density estimation via a linear diffusion process. *Sci. Rep.* **6**, 27381 (2016).
- Clark, P. J. & Evans, F. C. Distance to Nearest Neighbor as a Measure of Spatial Relationships in Populations. *Ecology* **35**, 445–453 (1954).
- Lee, J., Lay, J.-G., Chin, W. C. B., Chi, Y.-L. & Hsueh, Y.-H. An Experiment to Model Spatial Diffusion Process with Nearest Neighbor Analysis and Regression Estimation. *Int. J. Appl. Geospatial Res.* **5**, 1–15 (2014).
- Hess, D., van Lieshout, M.-C., Payne, B. & Stein, A. A review of spatio-temporal modelling of quadrat count data with application to striga occurrence in a pearl millet field. *Int. J. Appl. Earth Obs. Geoinformation* **3**, 133–138 (2001).
- Silverman, B. W. *Density estimation for statistics and data analysis*. (Chapman and Hall, 1986).
- Openshaw, S., Charlton, M., Wymmer, C. & Craft, A. A Mark 1 Geographical Analysis Machine for the automated analysis of point data sets. *Int. J. Geogr. Inf. Syst.* **1**, 335–358 (1987).
- Gatrell, A. C., Bailey, T. C., Diggle, P. J. & Rowlingson, B. S. Spatial Point Pattern Analysis and Its Application in Geographical Epidemiology. *Trans. Inst. Br. Geogr.* **21**, 256–274 (1996).
- Cuzick, J. & Edwards, R. Spatial Clustering for Inhomogeneous Populations. *J. R. Stat. Soc. Ser. B Methodol.* **52**, 73–104 (1990).
- Demšar, U. & Vrrantaus, K. Space-time density of trajectories: exploring spatio-temporal patterns in movement data. *Int. J. Geogr. Inf. Sci.* **24**, 1527–1542 (2010).
- Lee, J., Gong, J. & Li, S. Exploring spatiotemporal clusters based on extended kernel estimation methods. *Int. J. Geogr. Inf. Sci.* **31**, 1154–1177 (2017).
- Cliff, A. D. & Haggett, P. Changes in the seasonal incidence of measles in Iceland, 1896–1974. *J. Hyg. (Lond.)* **85**, 451–457 (1980).
- Cliff, A. D., Haggett, P. & Graham, R. Reconstruction of diffusion processes at different geographical scales: the 1904 measles epidemic in northwest Iceland. *J. Hist. Geogr.* **9**, 29–46 (1983).
- Cliff, A. D., Haggett, P. & Ord, J. K. Forecasting epidemic pathways for measles in Iceland: the use of simultaneous equation and logit models. *Ecol. Dis.* **2**, 377–396 (1983).
- Keeling, M. J. & Eames, K. T. D. Networks and epidemic models. *J. R. Soc. Interface* **2**, 295–307 (2005).
- Pastor-Satorras, R., Castellano, C., Van Mieghem, P. & Vespignani, A. Epidemic processes in complex networks. *Rev. Mod. Phys.* **87**, 925–979 (2015).
- Sun, Y., Ma, L., Zeng, A. & Wang, W.-X. Spreading to localized targets in complex networks. *Sci. Rep.* **6**, 38865 (2016).
- Eubank, S. *et al.* Modelling disease outbreaks in realistic urban social networks. *Nature* **429**, 180–184 (2004).
- Ridenhour, B. J., Braun, A., Teyrasse, T. & Goldsman, D. Controlling the Spread of Disease in Schools. *PLoS ONE* **6**, e29640 (2011).
- Fournet, J. & Barrat, A. Epidemic risk from friendship network data: an equivalence with a non-uniform sampling of contact networks. *Sci. Rep.* **6**, 24593 (2016).
- Meloni, S. *et al.* Modeling human mobility responses to the large-scale spreading of infectious diseases. *Sci. Rep.* **1**, 62 (2011).
- Ryan, S. J., Jones, J. H. & Dobson, A. P. Interactions between Social Structure, Demography, and Transmission Determine Disease Persistence in Primates. *PLoS ONE* **8**, e76863 (2013).
- Wen, T.-H. & Chin, W.-C.-B. Incorporation of Spatial Interactions in Location Networks to Identify Critical Geo-Referenced Routes for Assessing Disease Control Measures on a Large-Scale Campus. *Int. J. Environ. Res. Public Health* **12**, 4170–4184 (2015).
- Chan, J., Holmes, A. & Rabadan, R. Network Analysis of Global Influenza Spread. *PLoS Comput. Biol.* **6**, e1001005 (2010).
- Gómez, J. M. & Verdú, M. Network theory may explain the vulnerability of medieval human settlements to the Black Death pandemic. *Sci. Rep.* **7**, 43467 (2017).
- Wang, Y. *et al.* Global analysis of an SIS model with an infective vector on complex networks. *Nonlinear Anal. Real World Appl.* **13**, 543–557 (2012).
- Wang, Y., Cao, J., Alofi, A., Al-Mazrooei, A. & Elaiw, A. Revisiting node-based SIR models in complex networks with degree correlations. *Phys. Stat. Mech. Its Appl.* **437**, 75–88 (2015).
- Sun, G.-Q. Pattern formation of an epidemic model with diffusion. *Nonlinear Dyn.* **69**, 1097–1104 (2012).
- Li, L. Patch invasion in a spatial epidemic model. *Appl. Math. Comput.* **258**, 342–349 (2015).
- Sun, G.-Q., Jusup, M., Jin, Z., Wang, Y. & Wang, Z. Pattern transitions in spatial epidemics: Mechanisms and emergent properties. *Phys. Life Rev.* **19**, 43–73 (2016).
- Sun, G.-Q. Mathematical modeling of population dynamics with Allee effect. *Nonlinear Dyn.* **85**, 1–12 (2016).

43. Sun, G.-Q., Wu, Z.-Y., Wang, Z. & Jin, Z. Influence of isolation degree of spatial patterns on persistence of populations. *Nonlinear Dyn.* **83**, 811–819 (2016).
44. Sun, G.-Q., Wang, C.-H. & Wu, Z.-Y. Pattern dynamics of a Gierer–Meinhardt model with spatial effects. *Nonlinear Dyn.* **88**, 1385–1396 (2017).
45. Abler, R., Adams, J. S. & Gould, P. *Spatial organization; the geographer's view of the world.* (Prentice-Hall, 1971).
46. Cohen, J. & Tita, G. Diffusion in Homicide: Exploring a General Method for Detecting Spatial Diffusion Processes. *J. Quant. Criminol.* **15**, 451–493 (1999).
47. Mantel, N. The detection of disease clustering and a generalized regression approach. *Cancer Res.* **27**, 209–220 (1967).
48. Knox, E. G. & Bartlett, M. S. The Detection of Space–Time Interactions. *Appl. Stat.* **13**, 25–30 (1964).
49. Kulldorff, M. & Hjalmars, U. The Knox Method and Other Tests for Space–Time Interaction. *Biometrics* **55**, 544–552 (1999).
50. Wen, T.-H., Lin, M.-H. & Fang, C.-T. Population Movement and Vector-Borne Disease Transmission: Differentiating Spatial–Temporal Diffusion Patterns of Commuting and Noncommuting Dengue Cases. *Ann. Assoc. Am. Geogr.* **102**, 1026–1037 (2012).
51. Kulldorff, M., Athas, W. F., Feurer, E. J., Miller, B. A. & Key, C. R. Evaluating cluster alarms: a space-time scan statistic and brain cancer in Los Alamos, New Mexico. *Am. J. Public Health* **88**, 1377–1180 (1998).
52. Wen, T.-H., Tsai, C.-T. & Chin, W.-C.-B. Evaluating the role of disease importation in the spatiotemporal transmission of indigenous dengue outbreak. *Appl. Geogr.* **76**, 137–146 (2016).
53. Ciofi degli Atti, M. L. *et al.* Mitigation Measures for Pandemic Influenza in Italy: An Individual Based Model Considering Different Scenarios. *PLoS ONE* **3**, e1790 (2008).
54. Balcan, D. *et al.* Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc. Natl. Acad. Sci. USA.* **106**, 21484–9 (2009).
55. Sun, G.-Q., Wang, S.-L., Ren, Q., Jin, Z. & Wu, Y.-P. Effects of time delay and space on herbivore dynamics: linking inducible defenses of plants to herbivore outbreak. *Sci. Rep.* **5**, srep11246 (2015).
56. Li, L., Jin, Z. & Li, J. Periodic solutions in a herbivore–plant system with time delay and spatial diffusion. *Appl. Math. Model.* **40**, 4765–4777 (2016).
57. Shang, C.-S. *et al.* The Role of Imported Cases and Favorable Meteorological Conditions in the Onset of Dengue Epidemics. *PLoS Negl. Trop. Dis.* **4**, e775 (2010).
58. Yu, H.-L., Yang, S.-J., Yen, H.-J. & Christakos, G. A spatio-temporal climate-based model of early dengue fever warning in southern Taiwan. *Stoch. Environ. Res. Risk Assess.* **25**, 485–494 (2011).
59. Shu, P.-Y. *et al.* Molecular Characterization of Dengue Viruses Imported Into Taiwan during 2003–2007: Geographic Distribution and Genotype Shift. *Am. J. Trop. Med. Hyg.* **80**, 1039–1046 (2009).
60. Huang, J.-H. *et al.* Molecular Characterization and Phylogenetic Analysis of Dengue Viruses Imported into Taiwan during 2008–2010. *Am. J. Trop. Med. Hyg.* **87**, 349–358 (2012).
61. Chang, S.-F., Huang, J.-H. & Shu, P.-Y. Characteristics of dengue epidemics in Taiwan. *J. Formos. Med. Assoc.* **111**, 297–9 (2012).
62. Yang, C.-F., Hou, J.-N., Chen, T.-H. & Chen, W.-J. Discriminable roles of *Aedes aegypti* and *Aedes albopictus* in establishment of dengue outbreaks in Taiwan. *Acta Trop.* **130**, 17–23 (2014).
63. World Health Organization. Epidemiology, burden of disease and transmission. In *Dengue: Guidelines for Diagnosis, Treatment, Prevention and Control* 1–21 (World Health Organization, 2009).
64. Hsu, C.-I. & Tsai, Y.-C. An Energy Expenditure Approach for Estimating Walking Distance. *Environ. Plan. B Plan. Des.* **41**, 289–306 (2014).
65. Leys, C., Ley, C., Klein, O., Bernard, P. & Licata, L. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *J. Exp. Soc. Psychol.* **49**, 764–766 (2013).
66. Kan, C.-C. *et al.* Two clustering diffusion patterns identified from the 2001–2003 dengue epidemic, Kaohsiung, Taiwan. *Am. J. Trop. Med. Hyg.* **79**, 344–352 (2008).
67. Wen, T.-H. *et al.* Spatial–temporal patterns of dengue in areas at risk of dengue hemorrhagic fever in Kaohsiung, Taiwan, 2002. *Int. J. Infect. Dis.* **14**, e334–e343 (2010).
68. Schmidt, W.-P. *et al.* Population Density, Water Supply, and the Risk of Dengue Fever in Vietnam: Cohort Study and Spatial Analysis. *PLoS Med.* **8**, e1001082 (2011).
69. Souris, M. *et al.* Poultry Farm Vulnerability and Risk of Avian Influenza Re-Emergence in Thailand. *Int. J. Environ. Res. Public Health* **11**, 934–951 (2014).
70. Sharp, B. L. *et al.* Seven years of regional malaria control collaboration—Mozambique, South Africa, and Swaziland. *Am. J. Trop. Med. Hyg.* **76**, 42–47 (2007).
71. Laxminarayan, R. Trans-boundary commons in infectious diseases. *Oxf. Rev. Econ. Policy* **32**, 88–101 (2016).
72. Hsueh, Y.-H., Lee, J. & Beltz, L. Spatio-temporal patterns of dengue fever cases in Kaoshiung City, Taiwan, 2003–2008. *Appl. Geogr.* **34**, 587–594 (2012).
73. Khalid, B. & Ghaffar, A. Dengue transmission based on urban environmental gradients in different cities of Pakistan. *Int. J. Biometeorol.* **59**, 267–283 (2015).
74. Stoddard, S. T. *et al.* The Role of Human Movement in the Transmission of Vector-Borne Pathogens. *PLoS Negl. Trop. Dis.* **3**, e481 (2009).

## Acknowledgements

The research was supported by the grants of the Ministry of Science and Technology in Taiwan (MOST 104-2627-M-002-020; MOST 105-2627-M-002-018). The funder had no role in the study design, data collection and analysis or in the preparation of the manuscript.

## Author Contributions

T.H.W. and I.H.W. conceived the experiments. W.C.B.C. and I.H.W. conducted the experiments. W.C.B.C., and T.H.W. analyzed the results. W.C.B.C., T.H.W., and C.E.S. wrote the paper. All authors reviewed the manuscript and approved the content.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-017-12852-z>.

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017