Centrality measure and visualization technique for multiple-parent nodes of earthquakes based on correlation-metric

Yuki Yamagishi[1,2*], Kazumi Saito[2], Kazuro Hirahara[2] and Naonori Ueda[2]

*Correspondence:
yamagishi.yuki@sist.ac.jp;
yuki.yamagishi.ks@a.riken.jp

[1] Faculty of Informatics,
Shizuoka Institute of Science
and Technology, Fukuroi, Japan
[2] Center for Advanced
Intelligence Project, RIKEN, Kyoto,
Japan

Abstract

In this paper, we address the problem of earthquake declustering, and propose a *k*-nearest neighbors approach based on the selection of multiple-parent nodes with respect to each of the given earthquakes, which can be regarded as a natural extension of the conventional correlation-metric method based on the selection of a single-parent node. Based on this approach, we develop a centrality measure that exploits link weight assigned by a logarithmic-distance scheme and a technique of individually visualizing each set of child nodes with respect to given target earthquakes. For experimental evaluation, we used an earthquake catalog covering Japan and selected 24 earthquakes that caused considerable damage or casualties. We first show that our proposed centrality measure using a logarithmic-distance scheme can rank these 24 major earthquakes higher than four link-weighting schemes (i.e., uniform, magnitude, inverse-distance, and normalized-inverse-distance weighting) and conventional single-parent selection. We then show that unlike the conventional approach to simultaneously visualizing all the events in the catalog, our proposed technique can produce a naturally interpretable classification result for these 24 major earthquakes, by individually visualizing each set of the first to *k*-th child nodes with different colored markers plotted in the directly interpretable spatio and temporal metrics. As a consequence, we confirm that our approach based on multiple-parent selection is vital and promising.

Keywords: Correlation-metric, Multiple-parent selection, Centrality measure, Visualization

Introduction

In seismology, the relationships between earthquakes collected in an extensive catalog must be unveiled. In particular, earthquake declustering, which classifies an earthquake as foreshock, mainshock, or aftershock (van Stiphout et al. 2012), is essential for applications such as earthquake prediction and seismic activity modeling. The key in earthquake declustering lies in identifying pairs of strongly interacting earthquakes.

In this paper, we address the problem of earthquake declustering, and propose a *k*-nearest neighbors approach based on the selection of a multiple-parent node with respect to each of the given events, which can be regarded as a natural extension of the

Yamagishi *et al. Applied Network Science* (2023) 8:14

Page 2 of 16

conventional correlation-metric method based on the selection of a single-parent node. Specifically, we construct a network containing multiple-parent nodes, with each node (vertex), link (edge), and weight corresponding to an earthquake (event), the interaction between two earthquakes, and interaction strength, respectively. In addition, we aim to find empirical regularities and explain basic properties of the resulting complex networks. To this end, we can use various techniques developed for large-scale complex networks, such as centrality analysis and community extraction. Finding regularities can unveil structures and trends and lead to new knowledge and insights underlying interactions between earthquakes.

Based on this proposed approach, we first derive a centrality measure that exploits link weight assigned by a logarithmic-distance scheme. Considering different link-weighting schemes, we evaluated the networks constructed using the proposed approach regarding ranking accuracy given by the centrality measure. Next, we also developed a technique of individually visualizing each set of child nodes with respect to given target events. This method visualizes the set of the first to k-th child nodes with different colors in the directly interpretable spatio and temporal metrics. In our experiments, unlike the conventional approach, we show that our technique is expected to uncover the types of major events by individually visualizing each set of child nodes. For experimental evaluation of the centrality measure and the visualization technique, we used an earthquake catalog covering Japan and selected 24 earthquakes that caused considerable damage or casualties.

Below, we summarize the contributions of this paper compared to the conference version (Yamagishi et al. 2022). As the methodological differences, we (1) formalized our proposed algorithms based on a k-nearest neighbors approach as a network construction method, a centrality measure, and a visualization technique and (2) proposed a technique of individually visualizing the set of the child nodes for given target earthquakes, unlike the conventional approach to simultaneously visualizing all the events in the catalog. As the experimental differences, we (1) showed that our proposed visualization technique can produce a naturally interpretable classification result for some major earthquakes, and (2) demonstrated that our visualization results using different colored markers for the first to k-th child nodes plotted in the directly interpretable spatio and temporal metrics can uncover some remarkable characteristics of these major earthquakes.

The remainder of this paper is organized as follows. In section "Related work", we describe conventional algorithms for earthquake declustering. Section "Proposed method" details the proposed network construction, centrality measure, and visualization technique. Section "Experimental evaluation" reports the experimental results using the earthquake catalog and an analysis of the proposed method. Finally, we draw conclusions and present future research directions in section "Conclusion".

Related work

Zaliapin et al. (2008; 2013a; b; 2016; 2020) have shown the effectiveness of using nearest neighbor (single-parent) earthquake selection and the correlation metric described below. On the other hand, Yamagishi et al. (2020; 2021b; a) used single-parent earthquake selection and the mean shift algorithm to experimentally demonstrate the

Yamagishi *et al. Applied Network Science* (2023) 8:14

Page 3 of 16

limitations of this approach. To overcome the limitations of existing approaches, we aim to enhance unweighted single-parent networks by describing weighted multiple-parent ones. In this section, we first introduce fundamental algorithms of seismicity declustering, and then review studies on link-based declustering and k-nearest neighbors (kNN) algorithm that rely on networks with k parent nodes.

Fundamental declustering algorithms

Seismicity declustering has been widely studied (van Stiphout et al. 2012) as a method to separate a catalog into subsets based on a specific relationship (e.g., foreshocks, mainshocks, and aftershocks) between earthquakes. Most of these declustering algorithms are based on a deterministic spatio-temporal window method (Knopoff and Gardner 1972; Gardner and Knopoff 1974) or stochastic model (Kagan and Jackson 1991; Zhuang et al. 2002), which suitably represents large earthquakes characterized by a series of aftershocks. Remarkably, the basis of several stochastic declustering methods is the epidemic-type aftershock sequence model (Ogata 1988, 1998) using likelihood analysis and considering space, time, and magnitude.

The window method is widely recognized as a straightforward way of identifying mainshocks and aftershocks. The origins of this technique can be traced to the proposal of window lengths and durations by Knopoff and Gardner (1972; 1974). Subsequently, Uhrhammer proposed alternative window parameter settings (Uhrhammer 1986), and Molchan and Dmitrieva conducted comparative experiments (Molchan and Dmitrieva 1992). Meanwhile, the algorithm devised by Reasenberg (1985), known as the cluster method, assumes an interaction zone centered on each earthquake. This method is based on the prior research of Savage (1972), and Molchan and Dmitrieva (1992) succinctly summarize the work of Reasenberg in their paper. As an alternative to the deterministic declustering methods above, the concept of probabilistic separation appeared in the research conducted by Kagan et al. (1991). Zhuang et al. (2002; 2004; 2006) suggested the stochastic declustering method, also called stochastic reconstruction, to bring such a probabilistic treatment into practice based on the ETAS (epidemic-type aftershock sequence) model (Ogata 1988, 1998). The generalization of stochastic declustering proposed by Marsan and Lengline (2008; 2010) is model-agnostic and can employ any (additive) seismicity model.

Link-based declustering algorithms

In other studies, Frohlich and Davis (1990; 1991) proposed the single-link cluster analysis based on a spatio-temporal metric between two earthquakes. As another cluster analysis with links, Baiesi and Paczuski (2004) proposed a simple spatio-temporal correlation metric to connect earthquakes, and Zaliapin et al. (2008) defined the rescaled distance and time. Such methods directly consider a tree of earthquakes, with an earthquake being a parent (i.e., foreshock) of subsequent earthquakes or a child (i.e., aftershock) having a single earlier earthquake. The parent can be identified as the nearest neighbor using the proximity function of the spatio-temporal metric based on the Gutenberg–Richter law (1954), which relates the magnitude and frequency of aftershocks, and modified Omori's law (1894; 1961; 1995), which relates the time after the mainshock and occurrence rate of aftershocks. This metric follows a bimodal distribution related to the background seismicity

Yamagishi *et al. Applied Network Science* (2023) 8:14

Page 4 of 16

and aftershocks, and methods to obtain two separate distributions are being actively studied Aden-Antoniów et al. (2022). Furthermore, the correlation metric is promising because it resembles the epidemic-type aftershock sequence model. Using this metric, Zaliapin and Ben-Zion (2013a; b) determined statistical properties of earthquake clusters describing bursts and swarms. They found a relationship between the predominant cluster and heat flow in a seismic region.

*k*NN algorithm

Yamagishi et al. (2020; 2021b) grouped earthquakes by link disconnection based on their average magnitudes using link-based declustering. By selecting a parent node from the earthquakes that occurred before a child node, single-parent earthquake selection, which is equivalent to the nearest neighbor graph (Preparata and Shamos 1985), was guaranteed to be the minimum spanning tree (or minimum weight spanning tree) (Kruskal 1956; Prim 1957). Proximity graphs have well-known extensions. For example, the *k*NN graph (Altman 1992; Eppstein et al. 1997) naturally extends the nearest neighbor graph, which is a *k*NN graph with $k = 1$. In addition, the relative neighborhood graph (Toussaint 1980) and the Gabriel graph (Gabriel and Sokal 1969) naturally extend the minimum spanning tree. However, the relative neighborhood and Gabriel graphs have $O(N^3)$ time complexity. Hence, the *k*NN graph is fast and relatively simple in concept and implementation, despite having $O(N^2)$ time and space complexities. In the *k*NN approach, a similarity or distance metric is used to find the $k \geq 1$ nearest neighbors (parents) of each child node. Thus, a weighted network (Wasserman and Faust 1994) using a similarity or distance metric is a natural extension with high expected performance.

Proposed method

Let $\mathcal{D} = \{(\mathbf{x}_i, t_i, m_i) \mid 1 \leq i \leq N\}$ be a set of observed earthquakes, where $\mathbf{x}_i, t_i$ and m_i are the location vector, time, and magnitude of earthquake *i*, respectively. Every earthquake (event) represents a single point in a spatio-temporal dimension, like in representative declustering methods (van Stiphout et al. 2012). We assume that the earthquakes are ordered from the oldest to the most recent (i.e., $t_i < t_j$ if $i < j$). We describe the proposed method based on *k*NN, and it consists of three components: (1) constructing a network of observed earthquakes (events) in $\mathcal{D}$ by selecting multiple-parent nodes per event; (2) a centrality measure per event, where a weight to each link is assigned over the constructed network based on a link-weighting scheme; and (3) a visualization technique per set of child nodes with respect to given target earthquakes.

Network construction

Among the available seismicity declustering algorithms, we consider the correlation metric (Baiesi and Paczuski 2004). For every pair of earthquakes *i* and *j* such that $i < j$, spatio-temporal metric $n(i, j)$ is defined as

$$n(i, j) = (t_j - t_i) \|\mathbf{x}_i - \mathbf{x}_j\|^{d_f} 10^{-b \, m_i}, \tag{1}$$

where d_f is the fractal dimension set to $d_f = 1.6$, *b* is the parameter of the Gutenberg–Richter law (1954) set to $b = 0.95$, and the spatial and temporal metrics are expressed in kilometers and seconds, respectively. Earthquake *j* is regarded as an aftershock (child

Yamagishi *et al. Applied Network Science* (2023) 8:14

Page 5 of 16

node) of $i(j)$ if $n(i, j)$ is minimized, that is, $i(j) = \arg\min\{n(i, j) \mid 1 \leq i < j\}$. Using the correlation metric, we construct directed network $G = (\mathcal{V}, \mathcal{E})$ with a single-parent node, where $\mathcal{V} = \{1, \cdots, N\}$ and $\mathcal{E} = \{(i(j), i) \mid 2 \leq j \leq N\}$ are the sets of nodes and links, respectively.

Next, we construct a network with multiple-parent nodes. For a given k and every earthquake $j \in \{k, \cdots, N\}$, we select a set of multiple-parent nodes $\mathcal{I}_k(j)$ such that $j \geq k$ and $|\mathcal{I}_k(j)| = k$. After initializing $\mathcal{I}_k(j) \leftarrow \emptyset$, we iterate the following two operations as $|\mathcal{I}_k(j)| < k$ holds: 1) $\hat{i} = \arg\min\{n(i, j) \mid 1 \leq i < j, i \notin \mathcal{I}_k(j)\}$. and 2) update $\mathcal{I}_k(j)$ by $\mathcal{I}_k(j) \leftarrow \mathcal{I}_k(j) \cup \{\hat{i}\}$. The resulting network with multiple-parent nodes is given by $G(k) = (\mathcal{V}, \mathcal{E}(k))$, where $\mathcal{E}(k) = \{(i, j) \mid k \leq j \leq N, i \in \mathcal{I}_k(j)\}$.

Centrality measure

We enhance the constructed network by assigning a weight to every link. We introduce a scheme based on logarithmic-inverse-distance (LID) weighting to assign the following weight to link $(i, j) \in \mathcal{E}$:

$$w(i,j) = \log\left(1 + \frac{1}{n(i,j)}\right), \tag{2}$$

where $n(i, j)$ is the spatio-temporal distance computed by Eq. (1) based on the correlation metric. As variants, we consider four weighting schemes: uniform, magnitude, inverse-distance, and normalized-inverse-distance weighting. These schemes are defined by $w(i,j) = 1$, $w(i,j) = m_j$, $w(i,j) = 1/n(i,j)$, and $w(i,j) = 1/(1 + n(i,j))$, respectively, where m_j is the magnitude of earthquake j. We experimentally evaluated the effectiveness of networks constructed considering these link-weighting schemes, as detailed in section "Experimental evaluation".

We performed the evaluations using the most basic centrality measure, namely, weighted degree ranking. Let $\mathcal{J}_k(i)$ be the set of child nodes defined by $\mathcal{J}_k(i) = \{j \mid j \in \mathcal{V}, i \in \mathcal{I}_k(j)\}$. The weighted degree ranking is given by

$$c_k(i) = \sum_{j \in \mathcal{J}_k(i)} w(i,j). \tag{3}$$

Centrality value $c_k(i)$ represents the number of aftershocks (i.e., child nodes) of earthquake i for the uniform scheme and the correspondingly weighted aftershocks for the other schemes.

Visualization technique

We first revisit a visualization technique of events in $\mathcal{D}$ proposed by Zaliapin et al. (2008) in the framework of the correlation-metric (*CM*) method. More specifically, for each pair of two events i and j such that $i < j$, the rescaled distance $R(i, j)$ and time $T(i, j)$ is defined as

$$R(i,j) = \|\mathbf{x}_i - \mathbf{x}_j\|^{d_f} 10^{-b\,m_i/2}, \quad T(i,j) = (t_j - t_i)10^{-b\,m_i/2}, \tag{4}$$

where note that the spatio-temporal metric $n(i, j)$ in Eq. (1) is computed as $n(i,j) = R(i,j) \times T(i,j)$. Then, for each event $j > 1$, by defining the vertical and horizontal coordinates as $v(j) = R(i(j), j)$ and $h(j) = T(i(j), j)$, respectively, we can obtain a

Yamagishi *et al. Applied Network Science* (2023) 8:14

Page 6 of 16

Table 1 Information of 24 major earthquakes selected for evaluation

| ID	Time	Longitude	Latitude	Depth	Mag.	Name
1	10/6/2000 13:30	133.349	35.2742	8.96	7.3	2000 Tottori
2	12/18/2001 13:02	122.8152	23.8935	8.00	7.3	2001 Yonaguni Island inshore
3	9/26/2003 4:50	144.0785	41.7785	45.07	8	2003 Tokachi-oki
4	9/5/2004 19:07	136.7977	33.0332	37.58	7.1	2004 Kii Peninsula 1
5	9/5/2004 23:57	137.1413	33.1375	43.54	7.4	2004 Kii Peninsula 2
6	10/23/2004 17:56	138.8672	37.2925	13.08	6.8	2004 Chuetsu
7	3/20/2005 10:53	130.1763	33.7392	9.24	7	2005 Fukuoka
8	8/16/2005 11:46	142.2778	38.1495	42.04	7.2	2005 Miyagi
9	3/25/2007 9:41	136.686	37.2207	10.70	6.9	2007 Noto
10	7/16/2007 10:13	138.6095	37.5568	16.75	6.8	2007 Chuetsu offshore
11	6/14/2008 8:43	140.8807	39.0298	7.77	7.2	2008 Iwate-Miyagi Nairiku
12	8/11/2009 5:07	138.4993	34.7862	23.32	6.5	2009 Shizuoka
13	2/27/2010 5:31	128.68	25.9187	37.00	7.2	2010 Ryukyu Islands
14	3/9/2011 11:45	143.2798	38.3285	8.28	7.3	2011 Tohoku 1
15	3/11/2011 14:46	142.861	38.1035	23.74	9	2011 Tohoku 2
16	3/11/2011 15:08	142.7668	39.8207	32.02	7.4	2011 Iwate offshore
17	3/11/2011 15:15	141.2525	36.1208	42.70	7.6	2011 Ibaraki offshore
18	3/11/2011 15:25	144.751	37.9143	11.00	7.5	2011 Sanriku offshore
19	3/12/2011 3:59	138.5978	36.986	8.38	6.7	2011 Nagano
20	4/7/2011 23:32	141.9202	38.2042	65.89	7.2	2011 Miyagi
21	4/11/2011 17:16	140.6727	36.9457	6.42	7	2011 Fukushima
22	4/14/2016 21:26	130.8087	32.7417	11.39	6.5	2016 Kumamoto 1
23	4/16/2016 1:25	130.763	32.7545	12.45	7.3	2016 Kumamoto 2
24	11/22/2016 5:59	141.6042	37.3547	24.50	7.4	2016 Fukushima

visualization result where it is expected that the aftershocks are reasonably separated from the other background events.

Let $\Delta \mathcal{J}_k(i)$ be the set of newly-added child nodes at k defined by $\Delta \mathcal{J}_k(i) = \mathcal{J}_k(i) \backslash \mathcal{J}_{k-1}(i)$, where $\mathcal{J}_0(i) = \emptyset$. Then, with respect to a given target event i, we consider visualizing each set of child nodes $\mathcal{J}_k(i)$ by using different colored markers for each $\Delta \mathcal{J}_{k'}(i)$ where $k' \leq k$. Here note that since the value $10^{-b\,m_i/2}$ are the same for all the child nodes in $\mathcal{J}_k(i)$, we propose to use a pair of simplified coordinates defined by $v'(j) = \|\mathbf{x}_i - \mathbf{x}_j\|$ and $h'(j) = t_j - t_i$.

Experimental evaluation

We constructed a dataset using an earthquake catalog that contains source parameters collected by the Japan Meteorological Agency [1] and covering all Japan, as in Yamagishi et al. (2021b, 2021a). The dataset contains $N = 104,343$ earthquakes that occurred from October 1, 1997 to December 31, 2016 with minimum magnitude $m_{\min} = \min_{1 \leq i \leq N}\{m_i\}$ and maximum depth of 3.0 and 100 km, respectively. In addition, we considered 24 major earthquakes that caused considerable damages or casualties in Japan, as in Yamagishi et al. (2021a). Table 1 lists the information of the 24 major earthquakes. Hereafter,

---
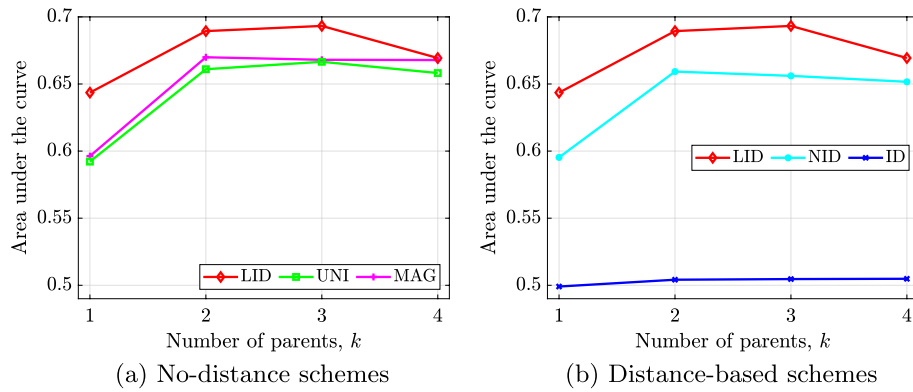
Fig. 1 Ranking accuracy using different weighting schemes

we simply refer to the major earthquake with $ID = i$ as earthquake i, and the j-th event in our earthquake catalog as event j.

Evaluation of centrality measure

In our first part of the experiments, we evaluated our proposed centrality measure, i.e., weighted degree ranking of the selected 24 major earthquakes, which should be ranked highly. Let $\mathcal{A}$ be the set containing the major earthquakes and $\mathcal{B}_{G(k)}(h)$ the set of the top h events weighted over constructed network $G(k)$. For a given $h \in \{1, \cdots, N\}$, we compute the precision and recall as $P_{G(k)}(h) = |\mathcal{A} \cap \mathcal{B}_{G(k)}(h)|/h$ and $R_{G(k)}(h) = |\mathcal{A} \cap \mathcal{B}_{G(k)}(h)|/|\mathcal{A}|$, respectively. For network $G(k)$, the area under the curve for the precision–recall curve is given by
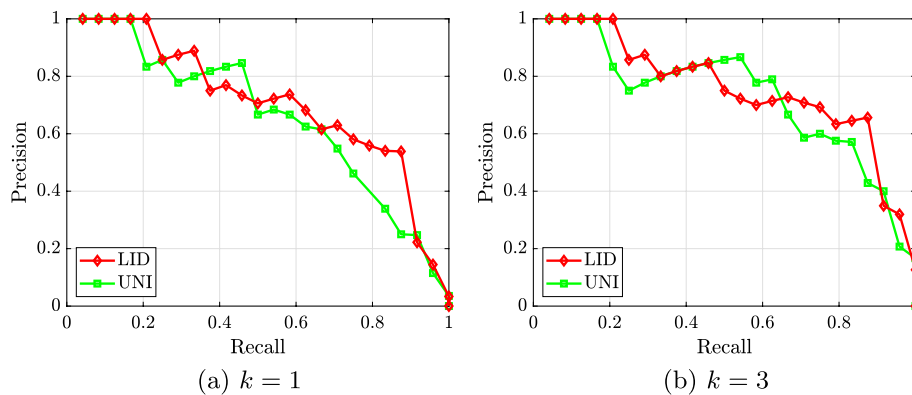
$$AUC_{G(k)} = \sum_{h=1}^{N-1} \frac{P_{G(k)}(h) + P_{G(k)}(h+1)}{2}(R_{G(k)}(h+1) - R_{G(k)}(h)), \tag{5}$$

where tie breaks are performed adequately.

Figure 1 shows the evaluation results for the k-th parent nodes considering the proposed LID link weighting, no-distance schemes of uniform (UNI) and magnitude (MAG) weighting (Fig. 1a), and distance-based schemes of inverse-distance (ID) and normalized-inverse-distance (NID) weighting (Fig. 1b), where $k \in \{1, \cdots, 4\}$. The results show that ranking can substantially improve by considering multiple-parent nodes for all the weighting schemes except for ID weighting. The proposed LID scheme works well, and the ID scheme has the worst performance, while the other schemes show similar performance. Compared with the single-parent network without link weights in Zaliapin and Ben-Zion (2020) (i.e., UNI scheme with $k = 1$), the ranking performance improves when considering multiple-parent nodes, as shown by the UNI scheme with $k = 2$ or 3. The performance can be further improved by employing link weights, as shown by the proposed LID scheme with $k = 2$ or 3. In addition, link weighting selection is important for ranking. Hence, we analyze the high performance of some types of weighted networks with multiple-parent nodes.

Figure 2 shows the results based on weighted degree centrality for the 24 major earthquakes and the k-th parent nodes using UNI (Fig. 2a) and LID (Fig. 2b)

Yamagishi *et al. Applied Network Science* (2023) 8:14

Page 8 of 16

(a) UNI scheme   (b) LID scheme
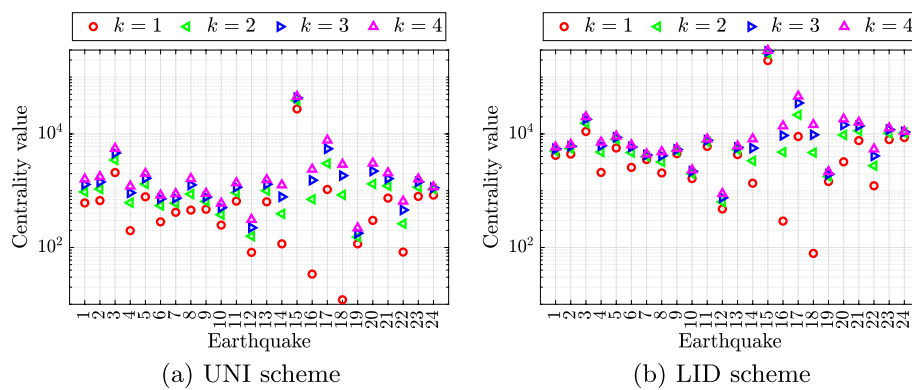
Fig. 2 Weighted degree centrality values

(a) $k = 1$ (b) $k = 3$

Fig. 3 Examples of precision–recall curves

weighting, where $k \in \{1, \cdots, 4\}$. The centrality values slightly increase with number k of parents. For earthquake 18 (2011 Sanriku offshore), the centrality for $k = 1$ is substantially smaller than the values for $k > 1$. Earthquake 15 (2011 Tohoku 2) had magnitude 9.0, which is the largest magnitude registered in the evaluated earthquake catalog. Earthquakes 15 and 18 occurred at close periods in a specific region. Thus, for most aftershocks of these earthquakes, the parent node was earthquake 15 for $k = 1$, and earthquake 18 was the next parent node for $k > 1$. This can explain the centrality of earthquake 18 increasing rapidly when $k = 2$. Similar trends can be observed for earthquakes 4, 14, 16, 20, and 22. These trends support the use of networks with multiple-parent nodes.

Figure 3 shows examples of precision–recall curves for LID and UNI weighting with $k = 1$ (Fig. 3a) and $k = 3$ (Fig. 3b). Regarding the performance of the constructed networks, LID weighting outperforms UNI weighting. On the other hand, as UNI weighting has a high performance for a small region with rank *h*, we may need to explore more sophisticated weighting schemes. Nevertheless, constructing weighted networks with multiple-parent nodes seems promising.

Yamagishi *et al. Applied Network Science* (2023) 8:14

Page 9 of 16

Table 2 Classification labels for the 24 major earthquakes

Type	id	Mag.	Name	Type	id	Mag.	Name
A1	1	7.3	2000 Tottori	A2	4	7.1	2004 Kii Peninsula 1
	2	7.3	2001 Yonaguni Island inshore		5	7.4	2004 Kii Peninsula 2
	3	8	2003 Tokachi-oki		6	6.8	2004 Chuetsu
	7	7	2005 Fukuoka		14	7.3	2011 Tohoku 1
	8	7.2	2005 Miyagi		22	6.5	2016 Kumamoto 1
	9	6.9	2007 Noto		23	7.3	2016 Kumamoto 2
	10	6.8	2007 Chuetsu offshore	B	15	9	2011 Tohoku 2
	11	7.2	2008 Iwate-Miyagi Nairiku		16	7.4	2011 Iwate offshore
	12	6.5	2009 Shizuoka		17	7.6	2011 Ibaraki offshore
	13	7.2	2010 Ryukyu Islands		18	7.5	2011 Sanriku offshore
	19	6.7	2011 Nagano		20	7.2	2011 Miyagi
	21	7	2011 Fukushima				
	24	7.4	2016 Fukushima				

Evaluation of visualization technique

In our second part of the experiments, we evaluated our proposed visualization technique by using the 24 major earthquakes shown in Table 1 as our target ones. Here recall that as the typical visualization results, the conventional *CM* method produces two clearly-separated clusters of events, which are interpreted as aftershocks and the other background ones, respectively. According to this characteristic, we first classified our results into two groups whether such clearly-separated clusters are observed or not. Hereafter, the group of those affirmative results is referred to as Type-A and the other one as Type-B. Next, we further classified our results of Type-A into two groups whether the cluster of aftershocks contains only the nearest neighbor child nodes in $\mathcal{J}_1(i)$ or not. Hereafter, the group of those affirmative results is referred to as Type-A1 and the other one as Type-A2. Table 2 shows the classification labels for the 24 major earthquakes where the numbers of elements in Type-A1, Type-A2, and Type-B are 13, 6, and 5 in this order.

In order to more closely examine our results, we selected some samples belonging to each group as sown in Fig. 4. Here we depict these events belonging to $\Delta\mathcal{J}_1(i)$, $\Delta\mathcal{J}_2(i)$, $\Delta\mathcal{J}_3(i)$, and $\Delta\mathcal{J}_4(i)$ by markers of red circle, green left-pointing triangle, blue right-pointing triangle, and magenta upward triangle, in this order. Note also that we employ the scales of 'km' and 'day' with respect to the horizontal and vertical axes, respectively. Here it should be emphasized that although most visualization results are omitted, they were reasonably comparable to the corresponding results of each type shown in Fig. 4.

In Fig. 4a, we show the visualization result of earthquake 1, as an example of Type-A1. As expected, we can observe two clearly-separated clusters of events, which are interpreted as aftershocks and the other background ones, respectively, where the former and latter clusters spread horizontally around the bottom part and diagonally around the upper-right part, respectively, and we can see that the former cluster contains almost only the events in $\Delta\mathcal{J}_1(i)$ depicted by red circle. Similarly, in Fig. 4b, we show the result of earthquake 4, as an example of Type-A2. Again, we can observe two clearly-separated clusters of aftershocks and the other background events, respectively, but the former cluster contains not only the events in $\Delta\mathcal{J}_1(i)$

Yamagishi *et al. Applied Network Science* (2023) 8:14

Page 10 of 16

(a) Type-A1 (earthquake 1)

(b) Type-A2 (earthquake 4)

(c) Type-B (earthquake 15)

(d) Type-B (earthquake 18)
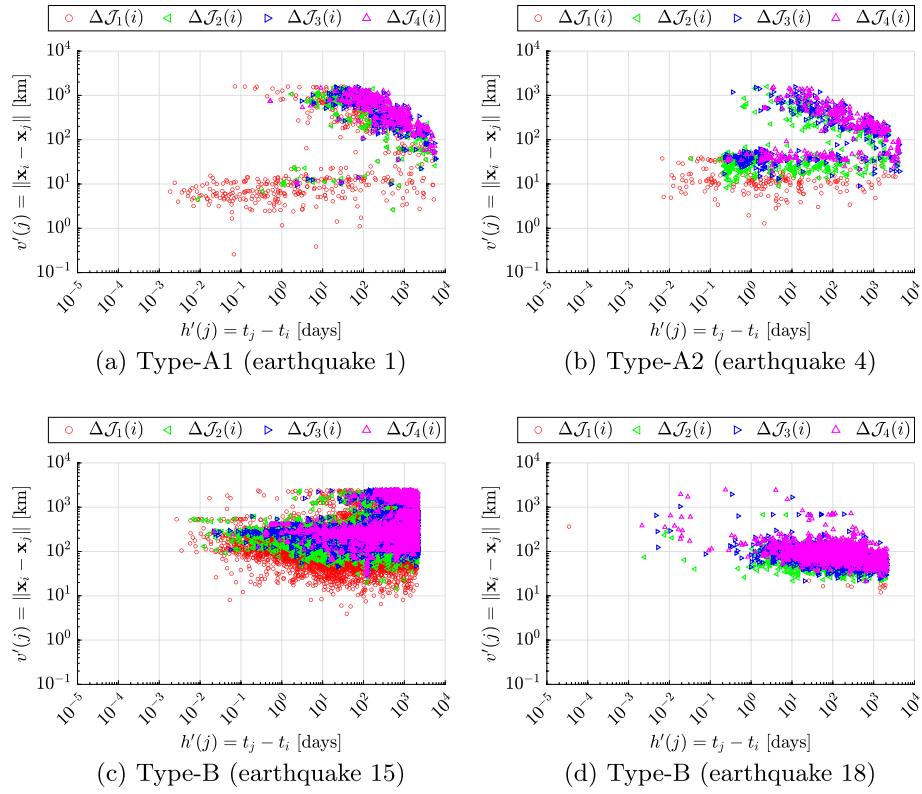
Fig. 4 Sample visualization results of each type

but also those in $\mathcal{J}_4(i)\setminus\mathcal{J}_1(i) = \Delta\mathcal{J}_2(i) \cup \Delta\mathcal{J}_3(i) \cup \Delta\mathcal{J}_4(i)$ depicted by three types of triangles. On the other hand, in Fig. 4a, d, we show the results of earthquakes 15 and 18, as examples of Type-B. Evidently, we cannot observe the two clearly-separated clusters of events in the case of these examples, unlike the examples shown in Fig. 4a, b.

Below we summarize the derivations from our visualization results. First, we cannot observe the two clearly-separated clusters of events in the visualization results for the four earthquakes in Type-B, i.e., the 2011 Tohoku earthquake with magnitude 9.0 and the other three ones that occurred after it within a quite short time-period. Second, child nodes in $\mathcal{J}_4(i)\setminus\mathcal{J}_1(i)$ play interesting roles to classify our visualization results, i.e., as for the clusters of aftershocks in Type-A, only the nearest neighbor child nodes belonging to $\Delta\mathcal{J}_1(i)$ appeared in the case of Type-A1, while those belonging to $\mathcal{J}_4(i)\setminus\mathcal{J}_1(i)$ appeared frequently in the case of Type-A2. In addition, as one advantage of our proposed visualization technique over the existing one proposed by Zaliapin et al. (2008), we can straightforwardly interpret each coordinate of visualized events as spatio-temporal distances from the target earthquakes in terms of the scales of 'km' and 'day'. For instance, from Fig. 4a, b, we can see that the clusters of aftershocks spread around the distance of 10 km. We believe that these experimental results suggest the vitality of our proposed technique based on multiple-parent nodes.
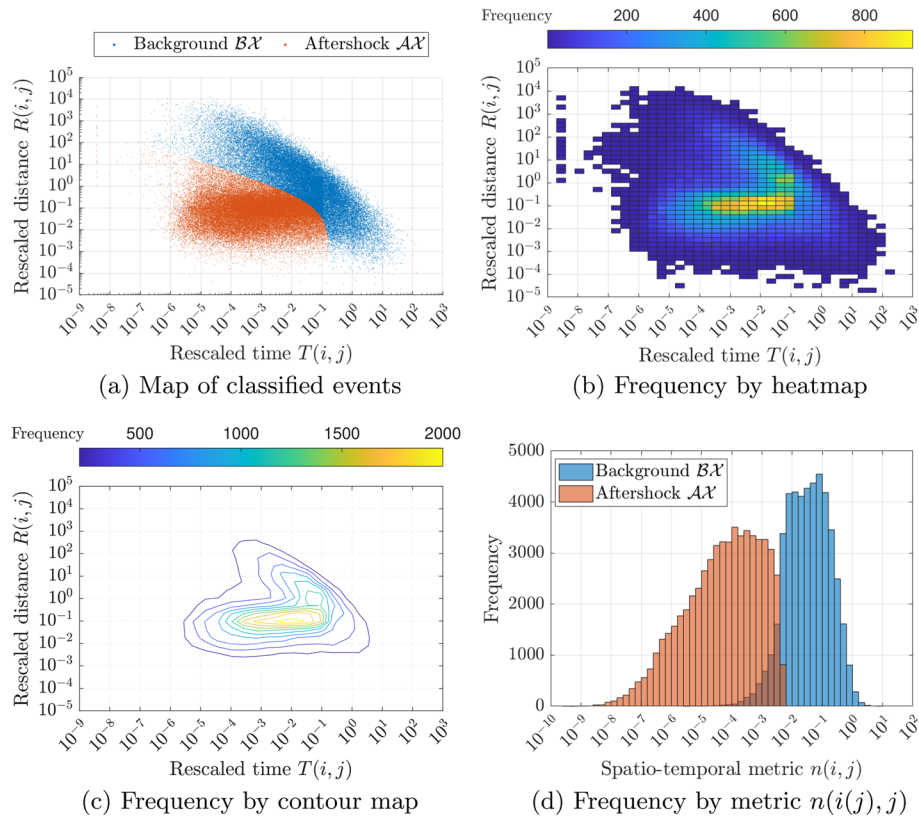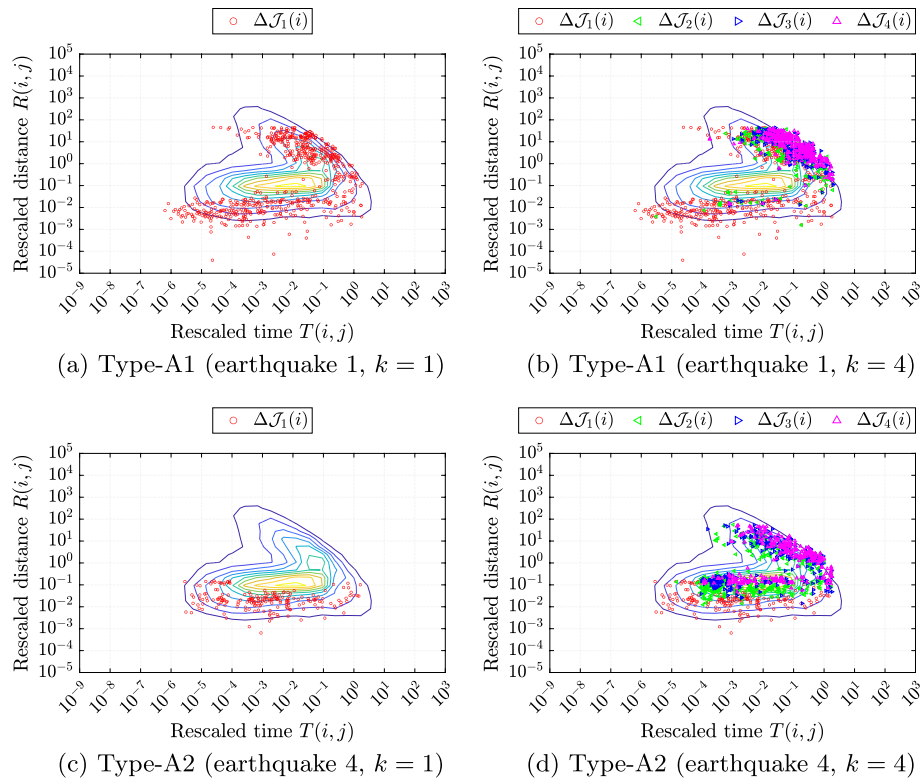
Fig. 5 Visualization results by the *CM* method

Analysis of visualization results

In this section, we analyze our visualization results in comparison to those obtained by the conventional *CM* method. As mentioned earlier, the *CM* method first computes a 2-dimensional vector $(v(j), h(j))$ for every event $j > 1$, and then divides $\mathcal{X} = \{(v(j), h(j)) \mid 2 \leq j \leq N\}$ into the aftershock and background components, $\mathcal{AX}$ and $\mathcal{BX}$, i.e., $\mathcal{X} = \mathcal{AX} \cup \mathcal{BX}$, by applying a Gaussian mixture clustering procedure, where recall that $v(j) = R(i(j), j)$ and $h(j) = T(i(j), j)$, $R(i, j)$ and $T(i, j)$ are the rescaled distance and time described in Eq. (4), and the event $i(j) = \arg\min\{n(i, j) \mid 1 \leq i < j\}$ is the first parent node of j. Here, it should be mentioned that according to the standard machine learning approach, we classify the observed events into $\mathcal{AX}$ and $\mathcal{BX}$ by selecting the class with the largest posterior probability, rather than determining some threshold value to the metric $n(i, j) = T(i, j) \times R(i, j)$.

Figure 5 shows several types of visualization results obtained by the *CM* method for our dataset. Recall that the fractal dimension d_f is set to $d_f = 1.6$, and the parameter b of the Gutenberg-Richter law is set to $b = 0.95$ (Baiesi and Paczuski 2004). In Fig. 5a, we plot all the events belonging to aftershock and background components as blue and brown points, where the formers and latters spread horizontally around the bottom part and diagonally around the upper-right part, respectively. As naturally expected, we can recognize a bimodal distribution with respect to these classified events. Here, we can also affirmatively confirm this bimodality by using the heat map in Fig. 5b and the
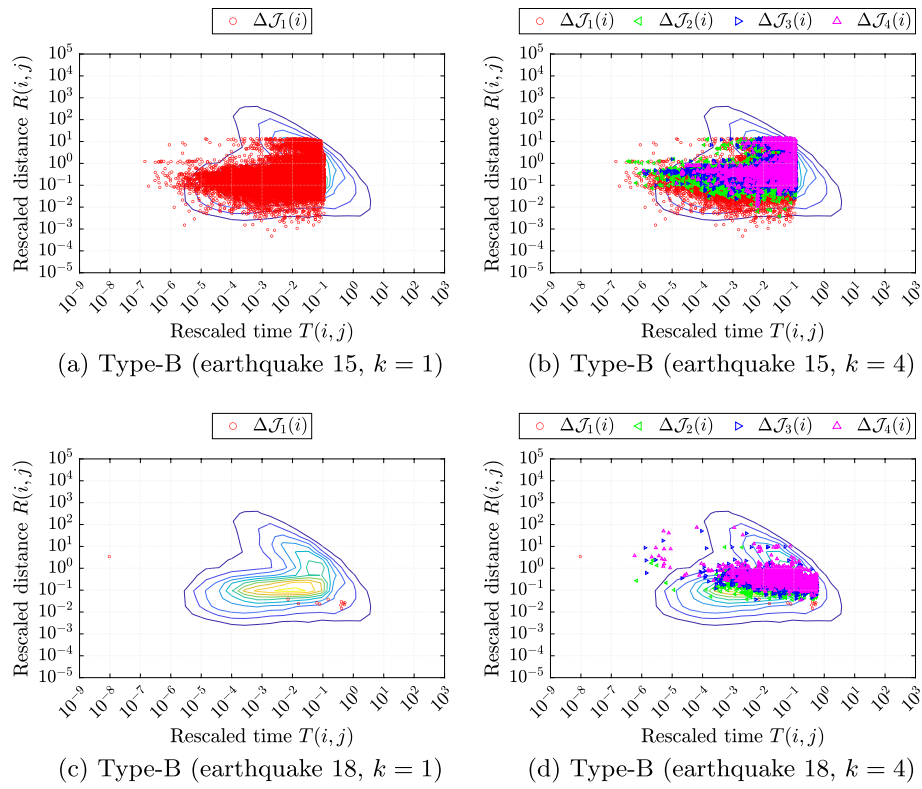
Fig. 6 Visualization results by variant method (Type-A)

contour map in Fig. 5c. In Fig. 5d, we show the distribution of these classified events based on the spatio-temporal metric, i.e., $n(i(j), j)$. From this distribution, it can be naturally indicated that even by using some threshold method based on the spatio-temporal metric $n(i(j), j)$, we can also obtain a similar classification result.

For an analysis purpose, we construct a variant of our visualization method. More specifically, for a given target earthquake i, in our framework of focusing on the child node (event) $j \in \mathcal{J}_k(i)$, we plot them by using the rescaled distance and time, i.e., $v(j) = R(i,j)$ and $h(j) = T(i,j)$, as coordinates of vertical and horizontal axes, instead of our direct metric $v'(j) = \|\mathbf{x}_i - \mathbf{x}_j\|$ and $h'(j) = t_j - t_i$. Figure 6 shows the visualization results by this variant method with respect to the Type-A earthquakes, together with the contour map by the original *CM* method shown in Fig. 5c. Specifically, by using earthquakes 1 and 4, we show the visualization results corresponding to Fig. 4a, b, respectively, where the cases of $k = 1$ and $k = 4$ are separately, as shown in the pairs of Fig. 6a, b, and c, d. From these visualization results, we can confirm that in the case of the Type-A1 earthquakes, the events in $\mathcal{J}_1(i)$ locate in the regions of both aftershock and background (Fig. 6a), while those in $\mathcal{J}_k(i)$ with $k \geq 2$ locate only in the region of background (Fig. 6b). On the other hand, in the case of the Type-A2 earthquakes, the events in $\mathcal{J}_1(i)$ locate only in the region of aftershock (Fig. 6c), while those in $\mathcal{J}_k(i)$ with $k \geq 2$ locate in the regions of both aftershock and background (Fig. 6d).

Figure 7 shows the visualization results by this variant method with respect to the Type-B earthquakes, together with the contour map by the original *CM* method

Fig. 7 Visualization results by variant method (Type-B)
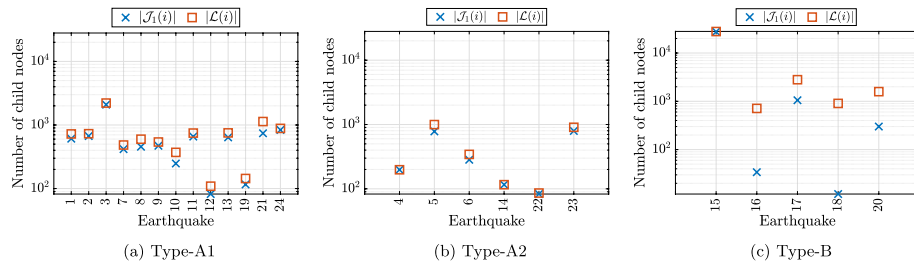
Fig. 8 Comparison of the numbers of child nodes of $\mathcal{J}_1(i)$ and $\mathcal{L}(i)$

shown in Fig. 5c. Specifically, by using earthquakes 15 and 18, we show the visualization results corresponding to Fig. 4c, d, respectively, where the cases of $k = 1$ and 4 are also separately as shown in the pairs of Fig. 7a, b, and c, d. From these visualization results, we can see that in the case of earthquake 15, the child nodes in $\mathcal{J}_k(i)$ locate in the regions of both aftershock and background (Fig. 7a, b). On the other hand, in the case of earthquake 18, the number of the child nodes in $\mathcal{J}_1(i)$ is quite small, they locate in some spherical region of aftershock (Fig. 7c), while those in $\mathcal{J}_k(i)$ with $k \geq 2$ locate around some region between aftershock and background (Fig. 7d).

According to Zaliapin et al. (2020), for the sake of discussing our obtained results, we define the attractive domain of given target earthquake i as $\mathcal{L}(i) = \{j \mid i = \arg\min\{n(i', j) \mid i \leq i' < j\}\}$, i.e., $\mathcal{L}(i)$ is the set of child nodes for

Yamagishi *et al. Applied Network Science* (2023) 8:14

Page 14 of 16

earthquake i when excluding any earthquake or event that occurred before it. Then we compare the numbers of first child nodes and elements in the attractive domain, i.e., $|\mathcal{J}_1(i)|$ and $|\mathcal{L}(i)|$, where $|\mathcal{J}_1(i)| \leq |\mathcal{L}(i)|$ from these definitions. Figure 8 shows the comparison of the numbers of child nodes of $\mathcal{J}_1(i)$ and $\mathcal{L}(i)$. Actually, in our catalog, the pairs of these numbers for earthquakes 1, 4, 15, and 18 were 609 and 723, 198 and 198, 27,555 and 27,855, and 12 and 904, in this order. Namely, in the case of earthquake 18 of Type-B, we can naturally suppose that most events in the attractive domain became the first child nodes of earthquake 15, and thus the number of the first child nodes of earthquakes with 18 became quite small (Fig. 8c). Recall that earthquake 15 had a magnitude 9.0, which is the largest magnitude in our catalog. Conversely, we can suppose that the first child nodes of earthquake 15 must be formed by some complex union of the attractive domains of the other many earthquakes. This conjecture can partly explain the reason why the clearly-separated clusters did not appear in the case of earthquake 15, unlike those observed in the Type-A earthquakes.

Conclusion

In this paper, as a natural extension of the conventional correlation-metric (*CM*) method, we proposed a k-nearest neighbors approach based on the selection of multiple-parent nodes with respect to each of the given earthquakes and addressed the problem of earthquake declustering. According to this approach, we proposed a centrality measure that exploits link weight assigned by a logarithmic-distance scheme and a technique of individually visualizing each set of child nodes with respect to given target earthquakes.

We performed evaluations using an earthquake catalog covering Japan and selected 24 major earthquakes that caused considerable damage or casualties. In short, we could confirm that our approach based on multiple-parent selection is vital and promising in terms of evaluations based on centrality measure and visualization. More specifically, the performance of our proposed centrality measure based on a logarithmic distance was better than those of four different link-weighting schema, i.e., uniform, magnitude, inverse-distance, and normalized-inverse-distance, as well as the conventional one that considers single-parent selection. In addition, by applying our visualization technique, we could obtain a naturally interpretable classification result for these 24 major earthquakes, by individually visualizing each set of the first to k-th child nodes with different colored markers plotted in the directly interpretable spatio and temporal metrics.

In future work, we will evaluate our proposed centrality measures and visualization technique by using a variety of datasets. In addition, we attempt to clarify the reason why these major earthquakes exhibit different characteristics in our visualization results.

Availability of data and materials

The dataset analyzed during the current study is available in Japan Meteorological Agency (https://www.data.jma.go.jp/svd/eqev/data/bulletin/eqdoc_e.html).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors have no competing interest to declare that are relevant to the content of this article.

References

Aden-Antonióv F, Frank WB, Seydoux L (2022) An adaptable random forest model for the declustering of earthquake catalogs. J Geophys Res: Solid Earth 127(2):e2021JB023254

Altman NS (1992) An introduction to kernel and nearest-neighbor nonparametric regression. Am Stat 46(3):175–185

Baiesi M, Paczuski M (2004) Scale-free networks of earthquakes and aftershocks. Phys Rev E, 69

Davis SD, Frohlich C (1991) Single-link cluster analysis, synthetic earthquake catalogues, and aftershock identification. Geophys J Int 104(2):289–306

Eppstein D, Paterson MS, Yao FF (1997) On nearest-neighbor graphs. Discrete Comput Geom 17:263–282

Frohlich C, Davis SD (1990) Single-link cluster analysis as a method to evaluate spatial and temporal properties of earthquake catalogues. Geophys J Int 100(1):19–32

Gabriel KR, Sokal RR (1969) A new statistical approach to geographic variation analysis. Syst Biol 18(3):259–278

Gardner JK, Knopoff L (1974) Is the sequence of earthquakes in Southern California, with aftershocks removed, Poissonian? Bull Seismol Soc Am 64(5):1363–1367

Gutenberg B, Richter C (1954) Seismicity of the earth and associated phenomena. Princeton University Press, Princeton, New Jersey, 2nd edition, 310

Kagan Y, Jackson D (1991) Long-term earthquake clustering. Geophys J Int 104:117–133

Knopoff L, Gardner JK (1972) Higher seismic activity during local night on the raw worldwide earthquake catalogue. Geophys J Int 28:311–313

Kruskal JB (1956) On the shortest spanning subtree of a graph and the traveling salesman problem. Proc Am Math Soc 7(1):48–50

Marsan D, Lengliné O (2008) Extending earthquakes' reach through cascading. Science 319(5866):1076–1079

Marsan D, Lengliné O (2010) A new estimation of the decay of aftershock density with distance to the mainshock. J Geophys Res: Solid Earth, 115

Molchan GM, Dmitrieva OE (1992) Aftershock identification: methods and new approaches. Geophys J Int 109:501–516

Ogata Y (1988) Statistical models for earthquake occurrences and residual analysis for point processes. J Am Stat Assoc 83(401):9–27

Ogata Y (1998) Space-time point-process models for earthquake occurrences. Ann Inst Stat Math 50:379–402

Omori F (1894) On the aftershocks of earthquakes. J Coll Sci 7:111–200

Preparata FP, Shamos MI (1985) Computational geometry: an introduction. texts and monographs in computer science, Springer

Prim RC (1957) Shortest connection networks and some generalizations. Bell Syst Tech J 36(6):1389–1401

Reasenberg P (1985) Second-order moment of central California seismicity, 1969–1982. J Geophys Res 90:5479–5495

Savage WU (1972) Microearthquake clustering near Fairview Peak, Nevada, and in the Nevada seismic zone. J Geophys Res 77:7049–7056

Toussaint GT (1980) The relative neighbourhood graph of a finite planar set. Pattern Recognit 12(4):261–268

Uhrhammer RA (1986) Characteristics of northern and central California seismicity. Earthq Notes 57(1):21–37

Utsu T (1961) A statistical study on the occurrence of aftershocks. Geophys Mag 30:521–605

Utsu T, Ogata Y, Sam R (1995) Matsu'ura: the centenary of the Omori formula for a decay law of aftershock activity. J Phys Earth 43(1):1–33

van Stiphout T, Zhuang J, Marsan D (2012) Seismicity declustering. Community Online Resource for Statistical Seismicity Analysis

Wasserman S, Faust K (1994) Social Network analysis: methods and applications. structural analysis in the social sciences. Cambridge University Press, Cambridge

Yamagishi Y, Saito K, Hirahara K, Ueda N (2020) Spatio-temporal clustering of earthquakes based on average magnitudes. In: Complex networks and their applications IX - volume 1, proceedings of the ninth international conference on complex networks and their applications. Studies in computational intelligence, vol 943, pp 627–637. Springer

Yamagishi Y, Saito K, Hirahara K, Ueda N (2021) Magnitude-weighted mean-shift clustering with leave-one-out bandwidth estimation. In: Proceedings of the 18th Pacific Rim international conference on artificial intelligence

Yamagishi *et al. Applied Network Science* (2023) 8:14

Page 16 of 16

Yamagishi Y, Saito K, Hirahara K, Ueda N (2021) Spatio-temporal clustering of earthquakes based on distribution of magnitudes. Appl Netw Sci

Yamagishi Y, Saito K, Hirahara K, Ueda N (2022) Constructing weighted networks of-earthquakes with -multiple-parent nodes based on-correlation-metric. In: Complex networks and their applications X. pp 487–498. Springer, Cham

Zaliapin I, Ben-Zion Y (2016) A global classification and characterization of earthquake clusters. Geophys J Int 207(1):608–634

Zaliapin I, Ben-Zion Y (2013) Earthquake clusters in southern California I: identification and stability. J Geophys Res: Solid Earth 118(6):2847–2864

Zaliapin I, Ben-Zion Y (2013) Earthquake clusters in southern California II: classification and relation to physical properties of the crust. J Geophys Res: Solid Earth 118(6):2865–2877

Zaliapin I, Ben-Zion Y (2020) Earthquake declustering using the nearest-neighbor approach in space-time-magnitude domain. J Geophys Res: Solid Earth 125(4)

Zaliapin I, Gabrielov A, Keilis-Borok V, Wong H (2008) Clustering analysis of seismicity and aftershock identification. Phys Rev Lett 101(1):1–4

Zhuang J (2006) Multi-dimensional second-order residual analysis of space-time point processes and its applications in modelling earthquake data. J R Stat Soc 68(4):635–653

Zhuang J, Ogata Y, Vere-Jones D (2002) Stochastic declustering of space-time earthquake occurrences. J Am Stat Assoc 97(458):369–380

Zhuang J, Ogata Y, Vere-Jones D (2004) Analyzing earthquake clustering features by using stochastic reconstruction. J Geophys Res 109

Publisher's Note