Prioritizing disaster mapping tasks for online volunteers based on information value theory

Yingjie Hu[1], Krzysztof Janowicz[2], Helen Couclelis[2]

[1]Department of Geography, University of Tennessee, Knoxville, TN 37996

[2]Department of Geography, University of California, Santa Barbara, CA 93106

Abstract

In recent years, online volunteers have played important roles in disaster response. After a major disaster, hundreds of volunteers are often remotely convened by humanitarian organizations to map the affected area based on remote sensing images. Typically, the affected area is divided using a grid-based tessellation, and each volunteer can select one grid cell to start mapping. While this approach coordinates the efforts of volunteers, it does not differentiate the priorities of different cells. As a result, volunteers may map grid cells in a random order. Due to the spatial heterogeneity within the disaster-affected area, different cells may contain geographic information that is of higher or lower value to emergency responders. Ideally, cells that potentially contain more valuable information should be assigned higher priority for mapping. This paper presents an analytical framework for prioritizing the mapping of cells based on the values of information contained in these cells. Our objective is to provide guidance for online volunteers so that potentially more important cells are mapped first. We present a novel method that is based on information value theory (IVT) and focus on road networks. We apply this method to a number of simulated scenarios and to a real disaster mapping case from the 2015 Nepal earthquake.

Keywords: disaster response, information value theory, crisis mapping, volunteered geographic information.

1 Introduction

"Is that road up there passable?" – Wohltman (2010), after the Haiti earthquake.

When a major disaster strikes, there is an urgent need for geographic information about the affected area. Emergency responders want to know who needs help, where they are, and how to get there (Zook et al., 2010). While many government agencies maintain their own geographic data, such data often provide only limited help, since existing roads may be blocked, bridges may have collapsed, and other critical infrastructure may be out of order. As a result, governments, organizations, and individuals often need to collect up-to-date information that reflects the status of the environment after the disaster.

In recent years, online volunteers have been actively involved in disaster response. Thanks to information and communication technologies (ICT), people throughout the world can

contribute to relief efforts without having to be physically present in the affected area (Haklay and Weber, 2008; Graham, 2010). Meanwhile, humanitarian communities, such as Standby Task Force (Ziemke, 2012; Meier, 2012a) and Crisis Mapper (Shanley et al., 2013), play important roles in organizing volunteers and coordinating their efforts. The support from both technologies and social organizations has greatly facilitated the involvement of online volunteers in disaster response.

One important way by which online volunteers contribute to relief efforts is by mapping the disaster-affected areas based on remote sensing images. For example, after the 2010 Haiti earthquake, a Web crisis mapping platform was quickly established and geographic information was collected with the help from thousands of online volunteers (Meier, 2011). A similar collaborative mapping effort was organized by the Humanitarian OpenStreetMap Team (HOT)[1] after the 2015 Nepal earthquake (Poiani et al., 2016; Haworth, 2016; Hu and Janowicz, 2015). During the mapping process, volunteers not only create new information but also help verify the existing geographic data, such as the connectivity and physical conditions of roads, which can be invaluable to emergency responders. As a result, the importance of online volunteers and the collaborative mapping paradigm have been increasingly recognized by the disaster relief community (Meier, 2015) and the general public (NPR, 2015).

As many online volunteers can be involved in collaborative mapping, one common approach to coordinating their efforts is to divide the affected area into a number of cells using a grid-based tessellation. Each volunteer can then select one grid cell to perform the mapping task. Figure 1 is a screenshot taken during the collaborative online mapping effort after the Nepal earthquake.

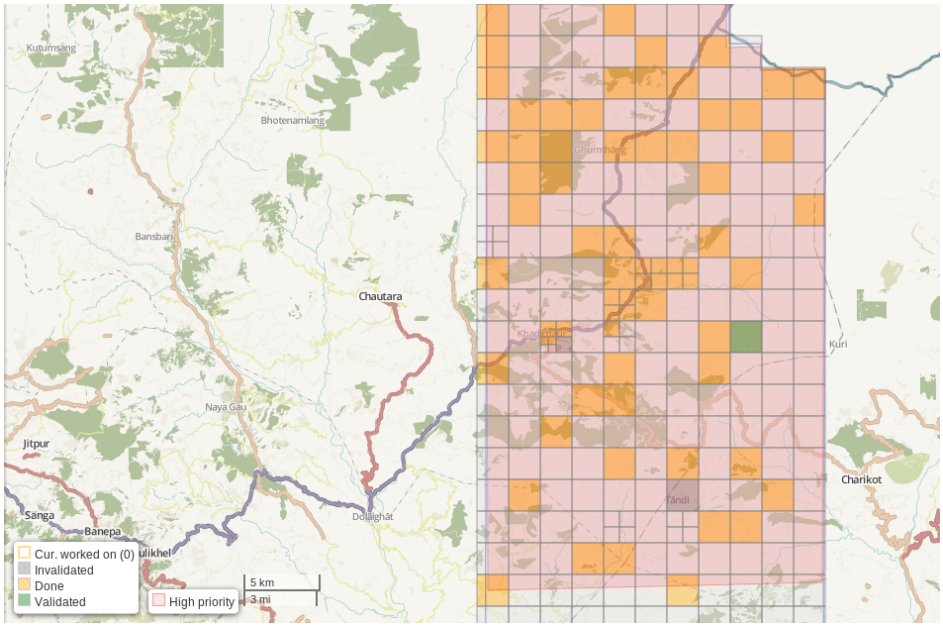


Figure 1: Collaborative online mapping after the Nepal earthquake of April 25th, 2015.

The different colors of the cells represent different mapping status. The orange color represents the cells that have been mapped, and green represents the cells that have been

[1] https://hotosm.org/

not only mapped but also validated. While this approach can help avoid editing conflicts and duplications, it does not differentiate the priorities of different grid cells. Consequently, volunteers may map these cells in a random order, as seems to be the case in the figure.

In any geographical area, different grid cells are likely to contain geographic information that is of differing value to emergency responders. For example, one cell may contain a road segment that can be used as the major access to the disaster affected area, whereas another cell may contain only a small road far away from populated areas. For an emergency responder who needs to plan relief trips to the affected area, the road connectivity information contained in the former cell will have higher value than that in the latter cell. Accordingly, it would be helpful if online volunteers can be guided to first map those cells that potentially contain more important information.

In this paper, we propose an analytical framework for prioritizing grid-based disaster mapping tasks. Our objective is to provide online volunteers with guidance on the priorities of cells, so that cells containing more important information can be mapped first. The first 72 hours after a disaster are widely considered to be the golden period for disaster response (Fiedrich et al., 2000; Comfort et al., 2004; Ochoa and Santos, 2015). Helping emergency responders obtain the most important geographic information as soon as possible can greatly facilitate urgent rescue tasks. While a grid cell may contain many different types of valuable geographic information, this research focuses on road network connectivity information in combination with population distribution. These are among the most important criteria used in disaster response for planning relief routes. The foundation of our analytical framework is information value theory (IVT), which originated in economics and artificial intelligence but, to the best of our knowledge, has not been applied to disaster response to date.

The research contributions of this work are as follows:

- We introduce a novel analytical framework for ranking the mapping priorities of grid cells overlaid on a disaster-affected area based on the value of road connectivity information contained in each cell.

- We present heuristics for implementing the proposed framework, which can significantly reduce the computing time while still achieving satisfactory accuracy.

- We provide a comparison between the priority ranking generated by the proposed framework and the actual order in which online volunteers mapped the affected area in a real-world disaster.

The remainder of this paper is organized as follows. Section 2 briefly introduces information value theory and reviews related work on relief efforts that involve the participation of online volunteers. Section 3 presents the methodological details of our analytical framework. Section 4 discusses heuristics for enhancing the computational efficiency of our framework. Section 5 compares the priority ranking generated by the framework with a population-based ranking and examines the ranking variations using four simulated disaster scenarios. The framework is then applied to data from the actual disaster mapping case of the 2015 Nepal earthquake. Section 6 summarizes our work and points out future research directions.

2 Related work

In this section, we provide some background on information value theory which serves as the foundation of our analytical framework. In addition, we review past disaster relief efforts that involve online volunteers.

2.1 Information value theory

Information value theory was originally proposed in 1966 (Howard, 1966). In contrast to Shannon and Weaver's information theory (1949) that measures the *amount* of information using *bits*, IVT quantifies the *value* of information as a function of its potential to assist decision making. The core idea of IVT can be summarized in equation (1):

$$V(I) = U(d') - U(d) \tag{1}$$

where I is an information item, and $V(I)$ represents the value of this information. d is the initial decision before I has been obtained, and d' is the modified decision after I has been obtained. U is a *utility function*, and $U(d)$ represents the utility of decision d. *Utility function* is a key component of decision theory, which quantifies the benefit of a decision using a numeric value (Russell and Norvig, 2010). A decision will be assigned a high utility value if it achieves the preferred outcome, and has a lower utility value if the outcome is less preferred. As can be seen from equation (1), the value of an information item is calculated as the utility difference before and after the information has been obtained. Depending on the magnitude of the utility difference, information items can have different values.

IVT has been applied to a variety of studies, such as investment analysis (Chen et al., 2001) and clinical trials (McFall and Treat, 1999). In these applications, monetary values have been used to represent the value of information. Recently, Hu et al. (2015) proposed a spatiotemporal approach for quantifying the value of information. Their method integrates IVT with time geography (Hägerstrand, 1970) and employs space-time prisms to examine the value of information with regard to the daily tasks of an individual. In this work, IVT will be employed and extended for ranking the priorities of disaster mapping cells.

2.2 Disaster relief with the participation of online volunteers

The idea of involving online volunteers in disaster response is related to the concepts of *crowdsourcing* (Howe, 2006) and *volunteered geographic information* (VGI) (Goodchild, 2007). While often lacking professional trainings, online volunteers fill a critical gap by providing timely geographic information of satisfactory quality (Goodchild and Glennon, 2010).

One early example that involves many online volunteers is the 2010 Haiti earthquake (Forrest, 2010). More than $4,000$ online volunteers from different countries around the world have participated in this response (Heinzelman and Waters, 2010). Volunteers contributed to the relief efforts in a variety of ways: they helped collect and integrate information from different sources (Crooks and Wise, 2011), translate the text messages sent by the local people (Meier and Munro, 2010; Hester et al., 2010), monitor social media platforms to extract information (Rogstadius et al., 2013), and collaboratively map the affected area based on remote sensing images (Liu and Ziemke, 2013). In addition to the Haiti earthquake,

online volunteers were also involved in the response to the 2011 crisis in Libya (Burns, 2014), the 2012 Hurricane Sandy in the U.S. (Meier, 2012b), and the 2013 Typhoon Haiyan in the Philippines (Humanitarian OpenStreetMap Team, 2013).

During the 2015 Nepal earthquake, online volunteers played an active and important role by mapping the affected area. According to a study conducted by Poiani et al. (2016), 4,287 new users registered for OpenStreetMap (OSM) within one day of the earthquake, and the average number of daily edits increased from around 20 to around 1,030. While this number decreased after two weeks, it increased again after the 7.3-magnitude aftershock on May 12th, 2015. Motivated by these online mapping examples, this research aims at providing guidance for online volunteers by helping to prioritize the mapping cells.

3 Analytical framework

In this section, we present the methodological details of our analytical framework. We first formalize the problem targeted by this research and then present the framework.

3.1 Problem statement

We address the problem of prioritizing online mapping tasks in the context of disaster response. Specifically, the disaster-affected area has been divided into cells using a grid-based tessellation. Online volunteers can choose a cell to start the mapping task based on remote sensing images. While each cell can contain a variety of geographic information, this study focuses on road network information, which is frequently required in disaster response, e.g., for evacuation.

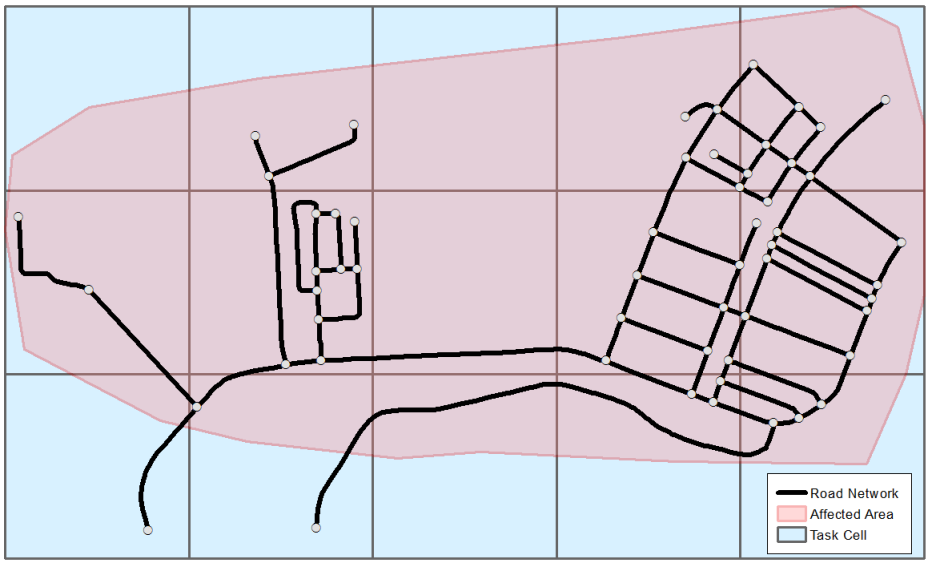


Figure 2: Our running example illustrating some of the main types of data used.

Figure 2 illustrates a simple example that will be used in the analysis. We consider four types of datasets for this problem.

- **Existing road network.** This dataset contains the road network in the affected area before the disaster (see the black lines and gray nodes in Figure 2). The information provided by online volunteers can help verify the connectivity of the roads *after* the disaster. We use a directed weighted graph $G(V, E)$ to represent the road network, with V representing the nodes and E representing the edges.

- **Disaster affected area.** This dataset contains the emergency planning zone (EPZ), which delineates the general region affected by the disaster (the pink region in Figure 2). Such an area is often provided by disaster experts based on the specific type of disaster, and we use A to represent this area.

- **Population distribution.** This dataset is used to estimate the number of people who may need help after the disaster. Block-level population data can often be obtained from local authorities. LandScan (Dobson et al., 2000) is a global-scale high-resolution population distribution dataset, which can also be used in many situations. A typical approach to integrating population data with road network is to aggregate population to the nodes of the road network (Cova and Church, 1997; Zhang et al., 2015). We use Pop_i to represent the population aggregated to node v_i.

- **Task grid cells.** These are the cells overlaid on the affected area to coordinate the mapping tasks (see the light blue squares in Figure 2). The sizes of the cells are defined by aid agencies. For example, in the Nepal earthquake, the sizes of cells were defined by the Humanitarian OpenStreetMap Team. We use C to represent the set of grid cells, and use c_k to represent one particular cell. Each cell is associated with a normalized disaster severity s_k in $[0, 1]$, with 0 indicating that the cell is not affected by the disaster, and 1 indicating that the cell is maximally affected. s_k can be estimated based on the intensity of the disaster within the cell. The disaster intensity data can be acquired from government agencies, such as the U.S. Geological Survey (USGS).

With these input data and their mathematical notations defined, we state the problem as follows: *given a road network $G(V, E)$, the disaster-affected area A, the affected population estimate at each node Pop_i, and a grid-based tessellation C, rank the mapping priorities of the cells $c_1, c_2, ..., c_n$ in C based on the values of the road connectivity information in these cells.*

3.2 Method

3.2.1 Potential target decision

To measure the value of information, we need a *target decision*, the decision for which the information will be used. However, most disaster mapping platforms are established for the general purpose of collecting information rather than for supporting a specific decision. Here, we propose the concept of *potential target decision*, that is, the implicit decision that is likely to be assisted by the information.

After a major disaster, road information is essential for emergency responders in order to plan trips for sending rescue teams and relief resources to the affected area. Each trip can be

considered as starting from the outside of the affected area, moving on the road network, and eventually arriving at one of the affected nodes. Selecting suitable routes is important for ensuring the timely success of these potential relief trips. Thus, we consider choosing good routes for these relief trips as the potential target decisions for road connectivity information. It is worth noting that relief trips do not always have to start from outside and go into the affected area, since there can also be relief efforts from responders within that area. In this study, however, we focus on potential relief trips performed by non-local (e.g., international) humanitarian organizations, as these are often the ones that establish the online mapping platforms.

To find the potential relief trip routings, we divide the nodes of the road network into two groups based on the affected area A: those within A and those that are immediately outside (a buffer of 100 meters have been used in this study to identify these nodes outside). We define the former as *affected nodes* (represented as a set V_a), and the latter as *entrance nodes* (represented as a set V_e). A potential relief trip from an entrance node to an affected node is represented as t_h. An emergency responder may need to plan many of these trips by deciding the suitable routes to take (e.g., those with shortest travel time). Thus, a potential target decision d can be formalized as deciding the routes for the potential relief trips to all the affected nodes:

$$d : \text{deciding the routes for relief trips } \{t_1, t_2, ..., t_{|V_a|}\} \tag{2}$$

3.2.2 Utility function

With the potential target decision d formalized, we continue to develop a utility function for quantifying the benefit of d. Since d is an aggregation (or mathematically a *set*) of the potential relief trips, we define the utility of d as the sum of the utilities of individual trips:

$$U(d) = \sum_h U(t_h) \tag{3}$$

where t_h is a potential relief trip to node v_h, $U(t_h)$ is the utility of t_h, and $U(d)$ is the utility of the potential target decision.

Each trip t_h connects an affected node v_h to an entrance node with the lowest travel cost. Let Pop_h be the corresponding population at v_h, c_k be the cell that v_h is located in, and s_k be the normalized disaster severity assigned to c_k. The utility of t_h can be estimated as:

$$U(t_h) \propto Pop_h \cdot s_k \tag{4}$$

This simple utility function models the utility of a potential trip as proportional to the estimated number of people who are affected by the disaster and who may need help. If an affected node has a large population and has been severely affected, then the corresponding relief trip has a high utility value to the emergency responders.

After a disaster, there exists uncertainty as to the connectivity of the roads. Consequently, some planned trips could fail due to the unknown connectivity. We use p_h to represent the probability that a trip t_h can succeed. The expected utility of t_h is:

$$EU(t_h) = p_h \cdot U(t_h) \qquad p_h \in [0, 1] \tag{5}$$

For each trip t_h, we use C_h to represent the set of cells that contain or partially contain this trip. Let c_k represent one cell in C_h, and let l_{hk} represent the length of the part of trip t_h contained in cell c_k. Each trip has an origin, a destination, and a number of middle nodes (the road intersections passed by the trip). Figure 3 shows a simple trip spanning two grid cells.

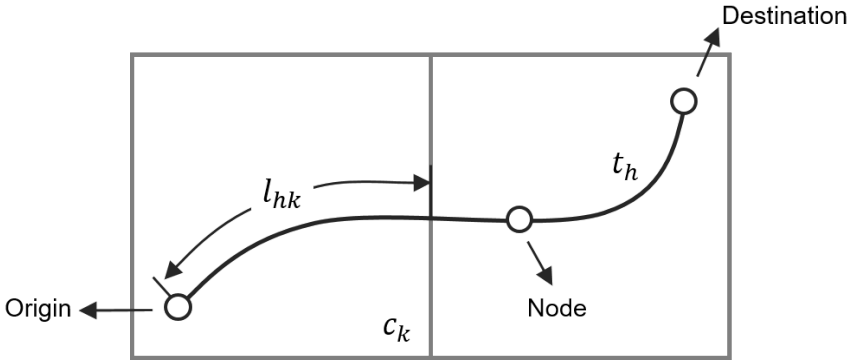


Figure 3: A simple trip spanning two grid cells.

Within the cell c_k, we assume a value p_u is available, which indicates the probability that a road with a unit length l_u (e.g., $1,000$ meters) is still passable after the disaster. p_u can be estimated based on the normalized disaster severity s_k in the cell. For a trip segment with length l_{hk} in the cell c_k, its probability of remaining passable p_{hk} can be calculated as:

$$p_{hk} = p_u^{\frac{l_{hk}}{l_u}} \qquad p_u \in [0, 1] \tag{6}$$

If l_{hk} is equivalent to the unit length l_u, then $p_{hk} = p_u$; if $l_{hk} > l_u$, then $p_{hk} < p_u$; and if $l_{hk} < l_u$, then $p_{hk} > p_u$. This modeling strategy captures the fact that longer roads are more likely to be affected by the disaster, such as being blocked by collapsed buildings.

With p_{hk}, the probability that trip t_h can be successful (i.e., p_h) is modeled as the joint probability of p_{hk}:

$$p_h = \prod_k p_{hk} \tag{7}$$

where p_{hk} represents the probability that the segment of trip t_h in the cell c_k remains passable.

By combining equations (5) and (7), we can derive equation (8) for calculating the expected utility of trip t_h.

$$EU(t_h) = \prod_k p_{hk} \cdot U(t_h) \tag{8}$$

Subsequently, the expected utility of the potential target decision d can be calculated as the sum of $EU(t_h)$:

$$EU(d) = \sum_h EU(t_h) \tag{9}$$

3.2.3 Information value

Let d be the potential target decision made before the road connectivity information in the cell c_k has been obtained, and d' be the decision after. Using IVT, the value of road connectivity information in c_k can be calculated by:

$$V(c_k) = EU(d') - EU(d) \tag{10}$$

where $V(c_k)$ represents the value of road connectivity information in c_k. By combining equation (9) and (10), we can get:

$$V(c_k) = \sum_h EU(t'_h) - \sum_h EU(t_h) \tag{11}$$

where t'_h represents the modified trip after the road connectivity information in c_k has been obtained, and t_h represents the trip before. If t_h is not (partially) contained by c_k, obtaining information in c_k does not influence the planning of t_h. As a result, we have $t'_h = t_h$ and $EU(t'_h) = EU(t_h)$. If t_h is (partially) contained by c_k, t'_h can be different from t_h depending on the obtained road connectivity information, thereby $EU(t'_h)$ can be different from $EU(t_h)$.

When the road connectivity information in c_k has been obtained, there can be many different situations. Let n be the number of road segments in c_k, and let $p_{k1}, p_{k2}, ..., p_{kn}$ be the probabilities that these road segments are passable. There can be situations, such as: all road segments are passable (with a probability of $q_1 = \prod_{i=1}^{n} p_{ki}$); all road segments are passable except the first one (with a probability of $q_2 = \prod_{i=2}^{n} p_{ki} \cdot (1 - p_{k1})$); and many other possible situations (we use q_i to represent the probability of a situation). In total, given n road segments in c_k, there can be 2^n situations (equation (12)). Figure 4 shows the 8 different situations when a cell contains three road segments.

$$C_n^0 + C_n^1 + ... + C_n^n = 2^n \tag{12}$$

Given one particular cell, we can find all possible situations and calculate their corresponding probabilities. For each relief trip (partially) contained by this cell, we can re-plan its route in any of these situations, and re-calculate the expected utility. Consider the situation shown in Figure 4(a). If a potential relief trip uses segments e_1, e_2, then there is no need to change its original route since both segments are passable in this situation. The expected utility of the relief trip t_h becomes: $EU(t'_{h\_i}) = (p_{h1} \cdot ... p_{h(k-1)} \cdot 1 \cdot 1 \cdot p_{h(k+1)} \cdot ...) \cdot U(t_h)$, where p_{hn} $(n \neq k)$ represents the probabilities that the road segments in other cells are passable, and the two 1s are the probabilities that the two road segments used by the trip within the cell are passable. Since e_1, e_2 are passable in this situation, their probabilities are 1.

In some other situations, such as in Figure 4 (d), a different route should be planned based on the updated road network. The expected utility of trip t_h in this situation then becomes: $EU(t'_{h\_i}) = \prod_{k'} p_{hk'} \cdot U(t_h)$, where k' represents the indexes of a new set of cells that (partially) contain the new route.

By summing up the expected utilities t'_{h_i} in all possible situations, we can calculate the expected utility of t'_h after the road connectivity information in c_k has been received:

$$EU(t'_h) = \sum_i q_i \cdot EU(t'_{h\_i}) \tag{13}$$

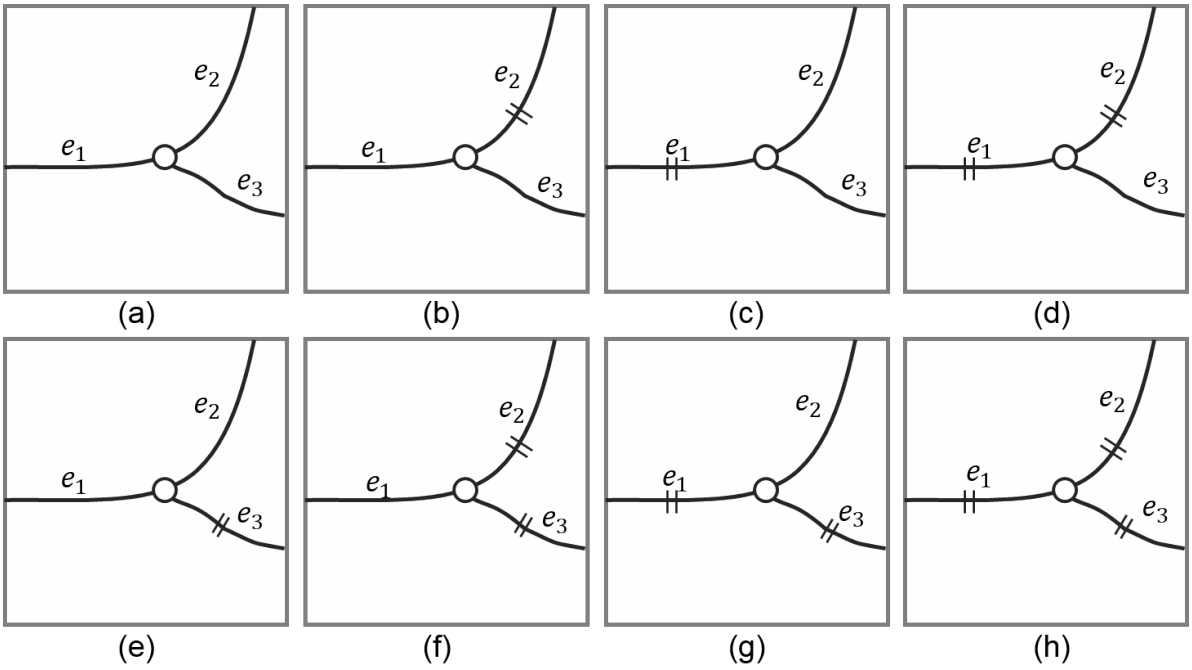


Figure 4: Eight possible situations when a cell contains three road segments.

where q_i is the probability of the situation, and $EU(t'_{h\_i})$ is the expected utility of the relief trip in this situation. We further define I_{hk} as an indicator variable that indicates whether the trip t_h is (partially) contained by the cell c_k.

$$I_{hk} = \begin{cases} 1 & \text{if } t_h \text{ is (partially) contained by } c_k \\ 0 & \text{if } t_h \text{ is not (partially) contained by } c_k \end{cases}$$

With I_{hk}, we can reduce the computational complexity of our method by calculating the expected utility changes only for the trips that are (partially) contained by c_k. Thus, the value of information in c_k can be calculated as:

$$V(c_k) = \sum_h [EU(t'_h) I_{hk}] - \sum_h [EU(t_h) I_{hk}] \tag{14}$$

We can further transform the above equation into:

$$V(c_k) = \sum_h [(EU(t'_h) - EU(t_h)) I_{hk}] \tag{15}$$

Thus, the road network information in the cell c_k has value when it can improve the expected utilities of the potential relief trips that are (partially) contained by c_k.

3.3 Summary

Figure 5 summarizes the workflow for ranking the priorities of grid cells.

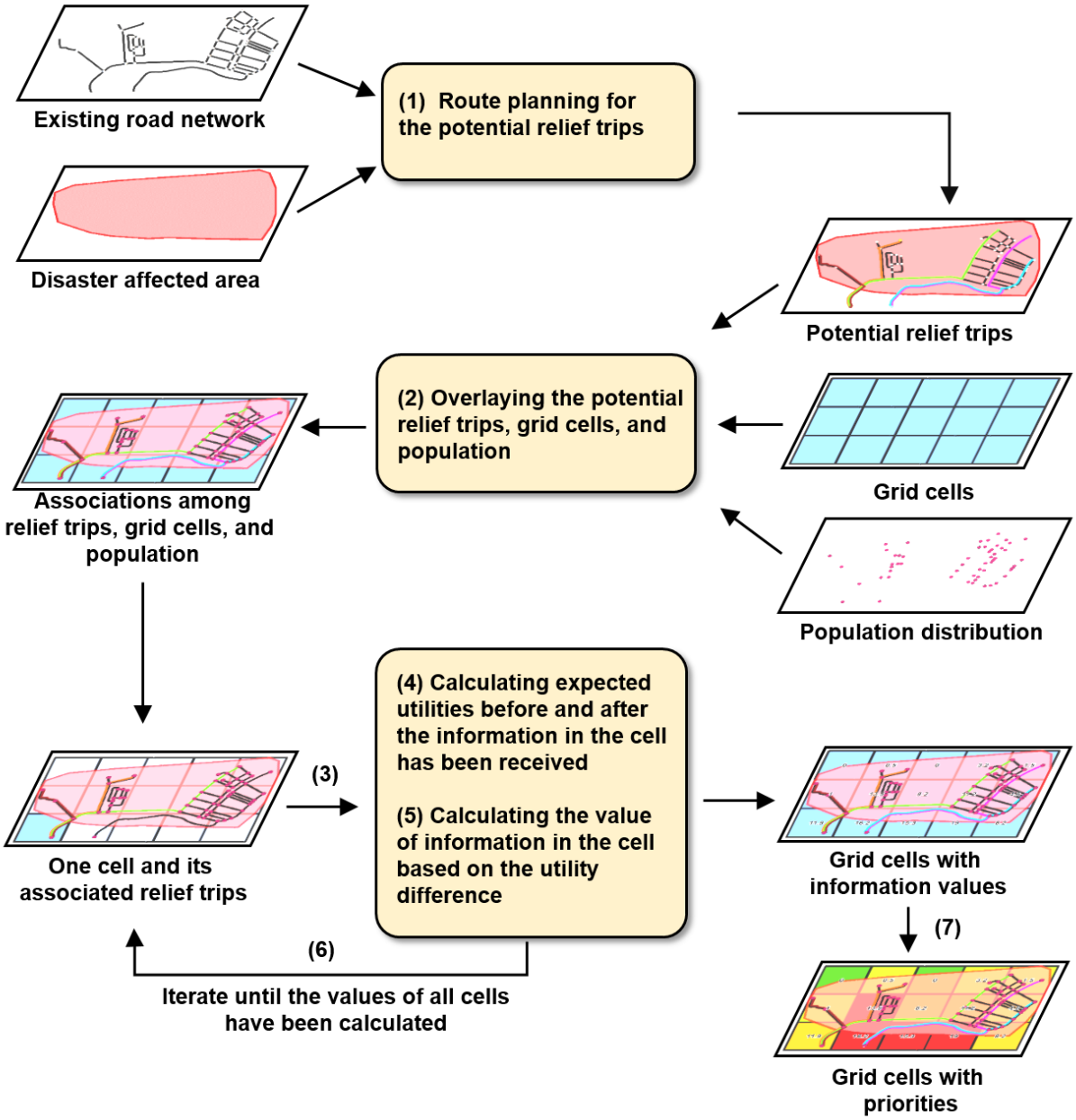


Figure 5: The workflow for ranking the priorities of disaster mapping cells.

In step (1), we plan routes for potential relief trips based on the existing road network and the disaster-affected area. Shortest path algorithms, such as Dijkstra's algorithm or A^*, can be used to find the routes with the lowest costs to the affected nodes. Step (2) overlays the potential relief trips (from step (1)), pre-defined grid cells, and the node-based population distribution, and builds associations among the three. Each relief trip will be associated with the grid cells it will pass by, as well as the affected population at the destination node. Meanwhile, each grid cell will be associated with the potential relief trips which are (partially) contained by this cell. In step (3), we select the first grid cell, and identify the potential

relief trips that are associated with this cell. Step (4) calculates the expected utilities of the associated trips before and after the information in the cell has been obtained, and then step (5) calculates the value of the information in the cell based on the utility difference. Step (6) repeats step (3) to (5) on all grid cells, and calculates the information value for each grid cell. Step (7) assigns priorities to grid cells based on their information values. These priorities can then help guide online volunteers for mapping.

4 Implementation and heuristics

As can be seen from equation (12), the proposed framework results in a computational complexity that is exponential to the total number of road segments in a cell. Such a high complexity can largely limit the use of the framework, especially in disaster response when time is of critical importance. In this section, we propose heuristics that can significantly reduce the computational complexity while still achieving satisfactory results.

4.1 A relief-trajectory-based heuristic

4.1.1 Method

This heuristic is based on the trajectories of the potential relief trips. We illustrate it using an example shown in Figure 6.

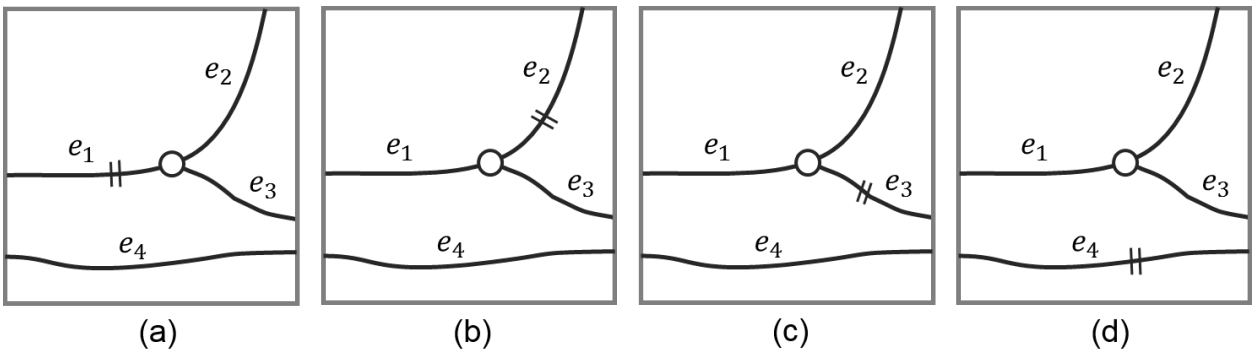


Figure 6: An example for illustrating the trajectory-based heuristic.

In this example, there are 4 road segments, and there is one relief trip that passes through e_1, e_2. In total, there are $2^4 = 16$ possible situations with different road connectivity conditions, and 4 of these situations are shown in the figure. Our original method will calculate the expected utilities of the potential relief trip in all these situations, which leads to an exponential complexity.

By re-examining the 4 situations, we can see that in sub figure (a) and (b), the obtained information can help emergency responders re-plan the relief trip which originally needs to go through e_1, e_2, thereby helping avoid the potential failure and improving the expected utility. In contrast, in sub figure (c) and (d), the emergency responders do not need to change the original route, since the disconnection of e_3 or e_4 do not directly affect the relief

trip. Since information value is calculated based on the utility improvement, we can focus on the segments whose disconnection can lead to re-routing. For the other road segments, we can use their existing probabilities to approximate their connectivity. In this example, we can reduce the number of situations from 2^4 to 2^2. More generally, we can reduce the number of situations from 2^n to 2^k, where n is the number of road segments in a cell, and k is the number of segments that may be used by the relief trip. While this heuristic still has an exponential time complexity, we can largely reduce the total amount of time, since 2^k can be much smaller than 2^n.

However, this heuristic also brings the possibilities of underestimating or overestimating the information value. Consider the example in Figure 7, and let us assume that the original relief trip will go through e_1, e_2, and the second best route is e_1, e_3, if e_2 become impassable. The heuristic can underestimate the information value in the situation in sub figure (a). This is because we have used the probability of e_3 (represented as p_3) to estimate its connectivity, but e_3 is actually connected and therefore has a probability of 1 of being passable. Since in most situations $p_3 < 1$, the heuristic will underestimate the information value. In contrast, the heuristic can overestimate the information value in situation shown in Figure 7(b). Our heuristic will use e_1, e_3 as the alternative route, but e_3 is, in fact, disconnected and therefore a different route has to be selected. Since e_1, e_3 is the second best route, our heuristic will overestimate the utility improvement, thereby leading to an overestimation of the information value.

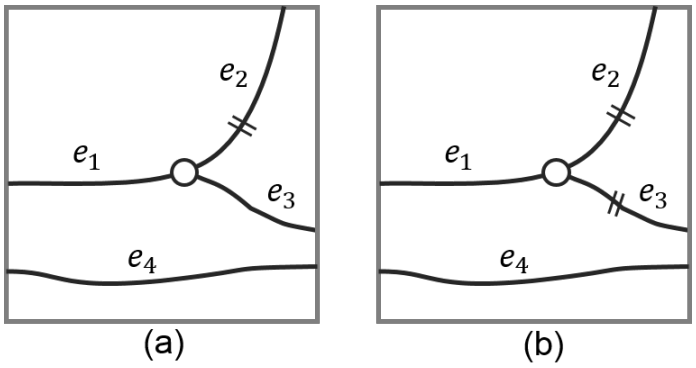


Figure 7: An example illustrating the underestimation and overestimation by the heuristic.

The pseudocode of our implementation is shown in algorithm 1. We start by a *for loop* that iterates through all the grid cells C (line 1). For each grid cell c_k, we identify the relief trips (partially) contained by this cell, calculate the expected utility for each trip before the information has been obtained (line 5), and also calculate the expected utility after (line 6-13). We then sum up the utility improvement from each of these trips as the value for c_k (line 14). The algorithm will finish when the values of all the cells have been calculated.

4.1.2 Evaluation

We evaluate our heuristic from two aspects: computational efficiency and result accuracy. We implemented our original method which has high computational complexity but can generate

Algorithm 1 Calculating the information values for grid cells

Input: road network with weighted nodes (affected population), grid cells with normalized disaster severities, potential relief trips

Output: information value for each cell

for each c_k in C **do**
  information value $v_k = 0$
  identify the set of relief trips T_k (partially) contained by c_k
  for each t_h in T_k **do**
    $EU(t_h)$ = the expected utility of the trip without information
    E_h = the set of road segments used by t_h within c_k
    Z_h = all possible situations based on E_h
    $EU'(t_h) = 0$
    for each z_i in Z_h **do**
      q_i = the probability of situation z_i
      t'_{h_i} = the new best route in this situation
      $EU'(t_h) = EU'(t_h) + q_i EU(t'_{h\_i})$
    end for
    $v_k = v_k + (EU'(t_h) - EU(t_h))$
  end for
  output v_k as the value of c_k
end for

the most accurate result. By comparing the performance of the heuristic-based implementation with the original method, we can quantitatively evaluate the reduced computing time as well as the accuracy change.

The evaluation experiment has been conducted using four road networks with different degrees of complexity (Figure 8). From sub figure (a) to (d) the complexity of the road network (in terms of the number of nodes and edges) increases.

We run both the original method and the heuristic-based method on the same computer which has 4 virtual CPUs based on Intel Xeon CPUs E5-2695 v2 with 2.40 GHz and 16 GB virtual RAM. Each experiment has been run 100 times, and the average computing time has been calculated. We measure the amount of time saved by the heuristic as:

$$TimeSaved = \frac{T_o - T_h}{T_o} \tag{16}$$

where T_o is the computing time used by the original method, and T_h is the computing time used by the heuristic-based method. We also evaluate the quality of the result from the heuristic-based method using the normalized root-mean-square deviation (NRMSD). It measures how much the result from the heuristic-based method deviates from the most accurate result from the original method. NRMSD is calculated as follows:

$$NRMSD = \frac{RMSD}{y_{max} - y_{min}} = \frac{\sqrt{\sum_{i=1}^{n} \frac{(\hat{y_i} - y_i)^2}{n}}}{y_{max} - y_{min}} \tag{17}$$

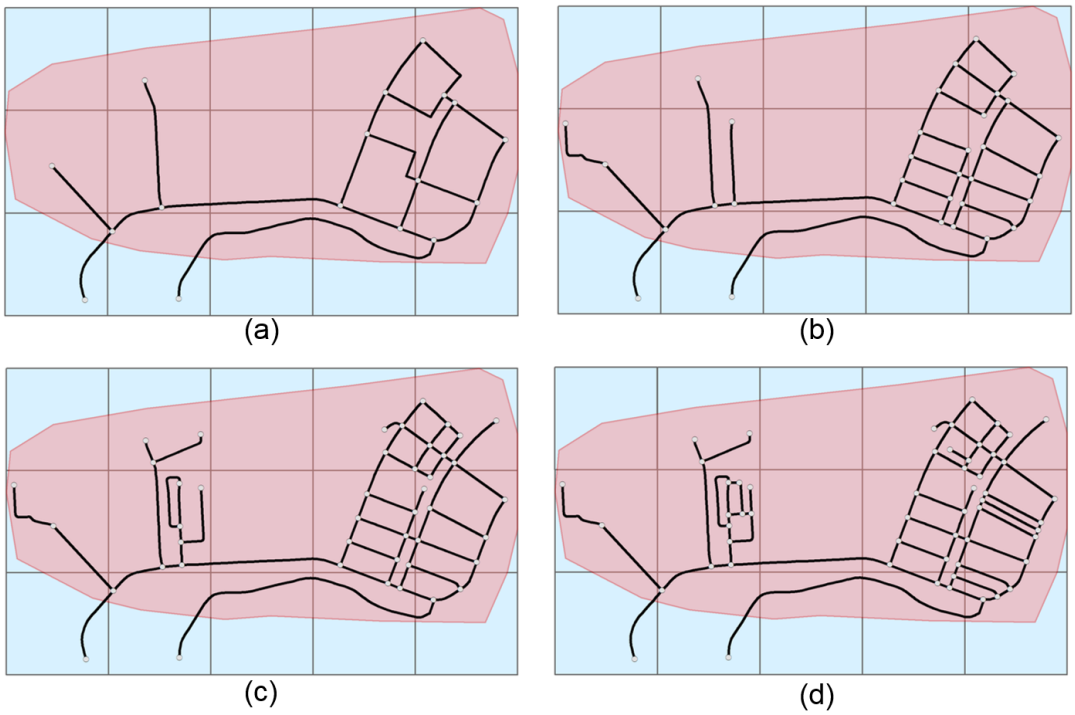


Figure 8: Four road networks with different degrees of complexity.

where n is the number of cells, $\hat{y}_i$ is the information value calculated by the heuristic-based approach, and y_i is the value calculated by the original method. The result of our experiments has been summarized in Table 1. We also plot out the computing times in Figure 9.

Table 1: Summary of the experiment result.

	Dataset (a)	Dataset (b)	Dataset (c)	Dataset (d)
Nodes	17	33	44	56
Edges	21	43	57	74
NRMSD	1.57%	1.22%	1.43%	1.42%
TimeSaved	44.44%	97.97%	99.37%	99.94%

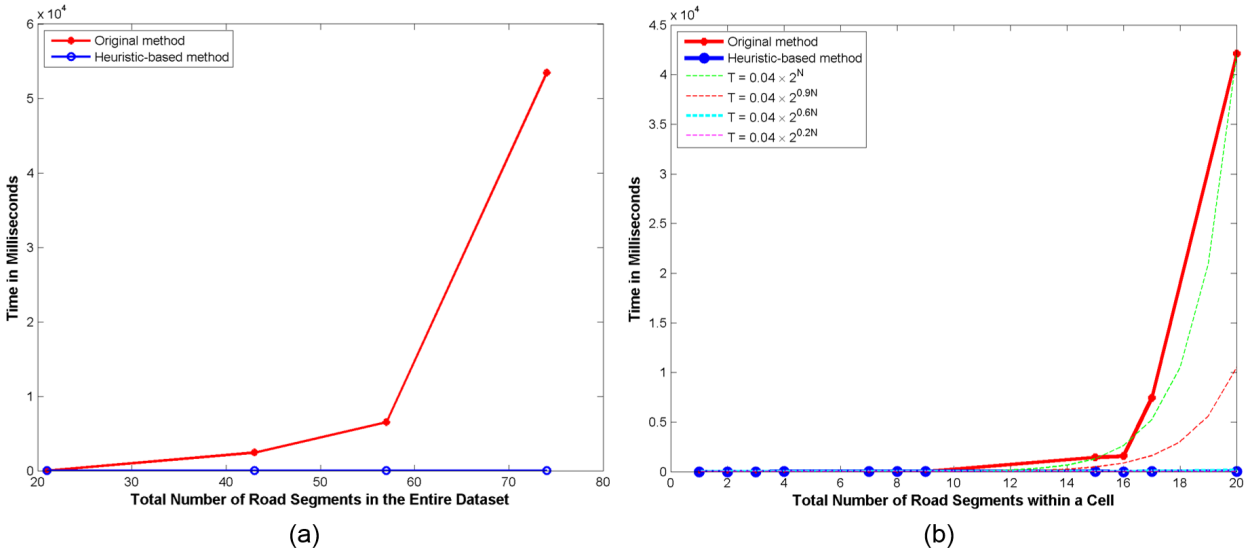


Figure 9: Comparison of the computational times between the original method and the heuristic-based approach: (a) computing time for entire datasets; (b) computing time for individual cells based on dataset (d).

It can be seen that the heuristic-based approach can significantly reduce the amount of computing time while slightly reducing the accuracy of the result. Figure 9(a) shows the computing times of the two methods based on the four entire datasets, and a large difference can be observed with the increase of the total number of road segments. Figure 9(b) shows the computing time for individual cells based on dataset (d). As discussed previously, our heuristic can reduce the computational complexity of one cell from 2^n to 2^k ($k \leqslant n$). To examine the effects of different values of k on reducing the computing time, we compare our experimental results with four standard mathematical models in the form of $T = C \times 2^k$, where T is the computing time, C is a constant, and $k = N, 0.9N, 0.6N$ and $0.2N$ respectively (N represents the total number of road segments in a cell). It can be observed that with a slight decrease of k (e.g., from N to $0.9N$), the computing time T decreases dramatically. In fact, when $k = 0.6N$, the curve is almost flat in the current extent of N. In our experiment, $k = 0.21N$, and therefore we can observe a large decrease of the computing time.

4.2 A road-hierarchy-based heuristic

We also propose another heuristic for improving the computational efficiency based on the road network hierarchy. Consider the example shown in Figure 10 in which (a) represents the original road network while (b) includes only the major road segments. A large difference in the numbers of road segments can be observed in the example cell shown in sub figures (a) and (b). From a perspective of disaster response, reaching the major road intersections and segments may be considered as reaching the general neighborhood affected by disaster. From a computational perspective, removing the minor road segments can exponentially reduce the computing time. Therefore, if the input road network data contain road hierarchy information, we can run the proposed framework on only the major road segments to increase

the computational efficiency while still achieving a reasonable result.

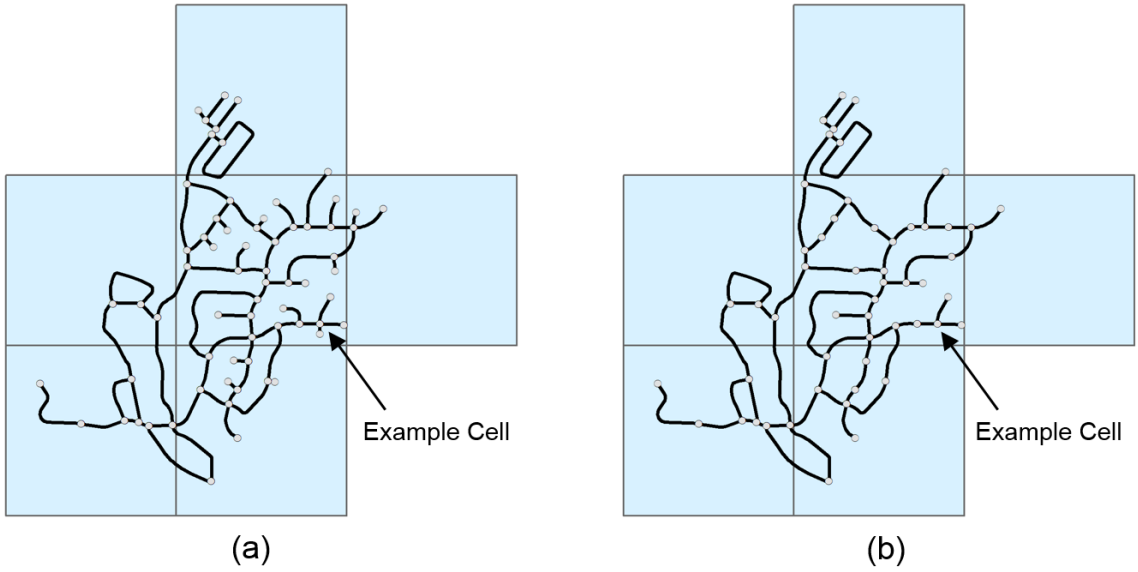


Figure 10: An example for illustrating the road-hierarchy-based heuristic.

5 Experiments

In this section, we conduct experiments based on simulated disaster scenarios as well as a real disaster mapping case from the 2015 Nepal earthquake.

5.1 Simulations

We perform two groups of simulations based on the dataset from the running example (the road data is a fragment taken from the road network of the county of Santa Barbara, California, USA). In the first group, we compare the results from our proposed framework with an approach based on the affected population. In the second group, we examine the variations of the information values when population and disaster severity are distributed differently.

5.1.1 Comparison with a population-based ranking

An intuitive approach for ranking the priorities of the disaster mapping cells is based on the estimated number of people affected in each cell. Therefore, we compare the priority ranking from our framework with that from the population-based approach.

In this simulation, we assign 100 units of population to each node of the road network, and the normalized disaster severity is set to 0.1 for each cell. We first apply our framework to the data, and the result is shown as in Figure 11. The colors of the roads indicate the number of the potential trips on each road segment. It can be seen that the two major access

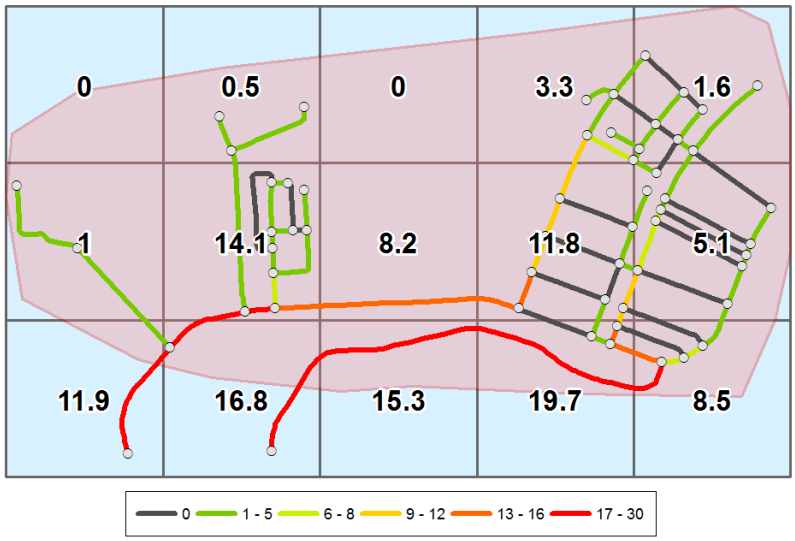
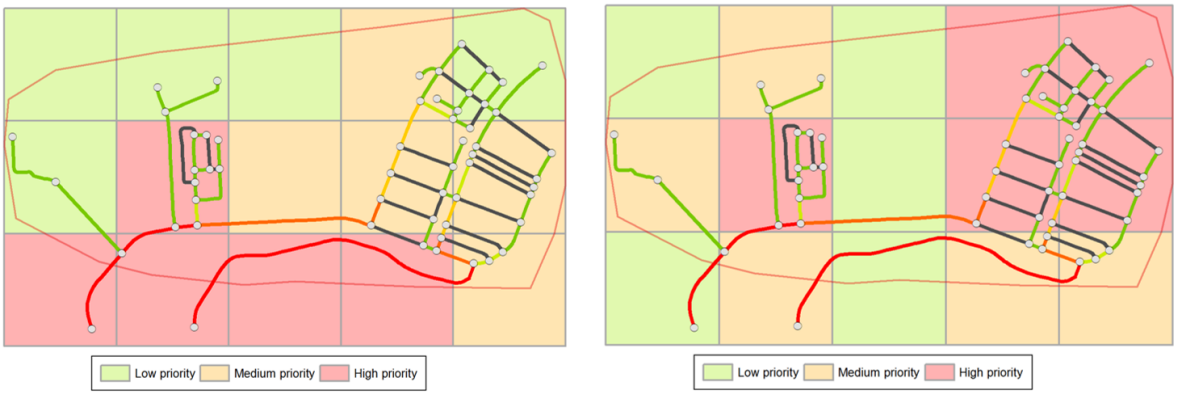


Figure 11: The value of road connectivity information in each cell.

roads starting from the lower left corner have been heavily used by the potential relief trips. The number on each grid cell represents the value of the road information contained in this cell. We can further classify the cells into *high priority*, *medium priority*, and *low priority* based on their information values, and Figure 12 (a) shows the ranked results. It is worth noting that the ranking result does not always have to be classified into 3 categories. Since the proposed framework has provided numeric value for each cell, emergency responders can classify the cells into other suitable numbers of categories depending on the specific needs. For the population-based approach, we aggregate the affected population of each node to the grid cell that contains this node. We then classify the grid cells into the same 3 categories (Fig. 12 (b)).

By comparing the sub figures (a) and (b), we can see an interesting difference which will be elaborated in the following text. To enhance the clarity of our discussion, we will refer to a grid cell using the coordinates based on its row and column numbers in the format of $(row, column)$. The grid cell at the upper left corner will be $(1, 1)$, and the cell at the lower right corner will be referred to as $(3, 5)$.

Cell $(3, 1)$ has been considered as *high priority* by our approach but *low priority* by the population-based approach. This result is understandable since this cell does not contain any affected node, and therefore has low affected population. However, this cell contains one important road segment that can be used as a major access point to the affected nodes. Knowing the connectivity of this road segment can greatly facilitate the planning of about half of the potential relief trips. Therefore, this cell has been assigned high priority by our framework. On the contrary, cell $(1, 5)$ has been considered as *low priority* by our approach but *high priority* by the population-based approach. This cell contains 8 affected nodes, and therefore has been ranked as important by the population-based approach. However, for the relief trips reaching to the 8 affected nodes, 5 trips have only small segments in this cell (3 have relatively longer segments). This means knowing the road connectivity in this



(a) Proposed approach (b) Population-based approach

Figure 12: Comparison of the ranking results between the proposed framework and the population-based approach.

cell will not largely improve the routes of these trips, since the major parts of these trips are in other cells. Thus, this cell has been assigned lower priority by our framework. It can be seen that the population-based approach ranks the cells more from a perspective of risk assessment, i.e., understanding which areas may need help. Our approach is from a perspective of relief trip planning, i.e., how to reach the affected areas. Both rankings have their values in disaster response.

5.1.2 Ranking variations in different disaster scenarios

Affected population and disaster severity are two important factors on the basis of which we calculate the values of information. Accordingly, different information values may be generated based on different spatial distributions of population and disaster severity, which may lead to different priority rankings. In the following, we examine this ranking variation by simulating four different scenarios as summarized in Table 2. By applying our framework to the four scenarios, we obtain the results as shown in Figure 13.

Table 2: A summary of the four simulated scenarios.

		Population distribution	
		Homogeneous	Heterogeneous
Disaster severity	Homogeneous	Scenario 1	Scenario 2
	Heterogeneous	Scenario 3	Scenario 4

In scenario 1, each node has been assigned 100 units of population, and the disaster severity for each cell is set as 0.1. The obtained information values are the same as Figure 11 discussed in Section 5.1.1. In scenario 2, three nodes (enlarged in the figure) are assigned $1,000$ units of population, while the other conditions remain the same as in scenario 1. It can be seen that the cells that (partially) contain the relief trips to these three nodes, such as cell $(1,2), (2,1), (2,2)$ and $(3,1)$, have an increase in their information value compared with

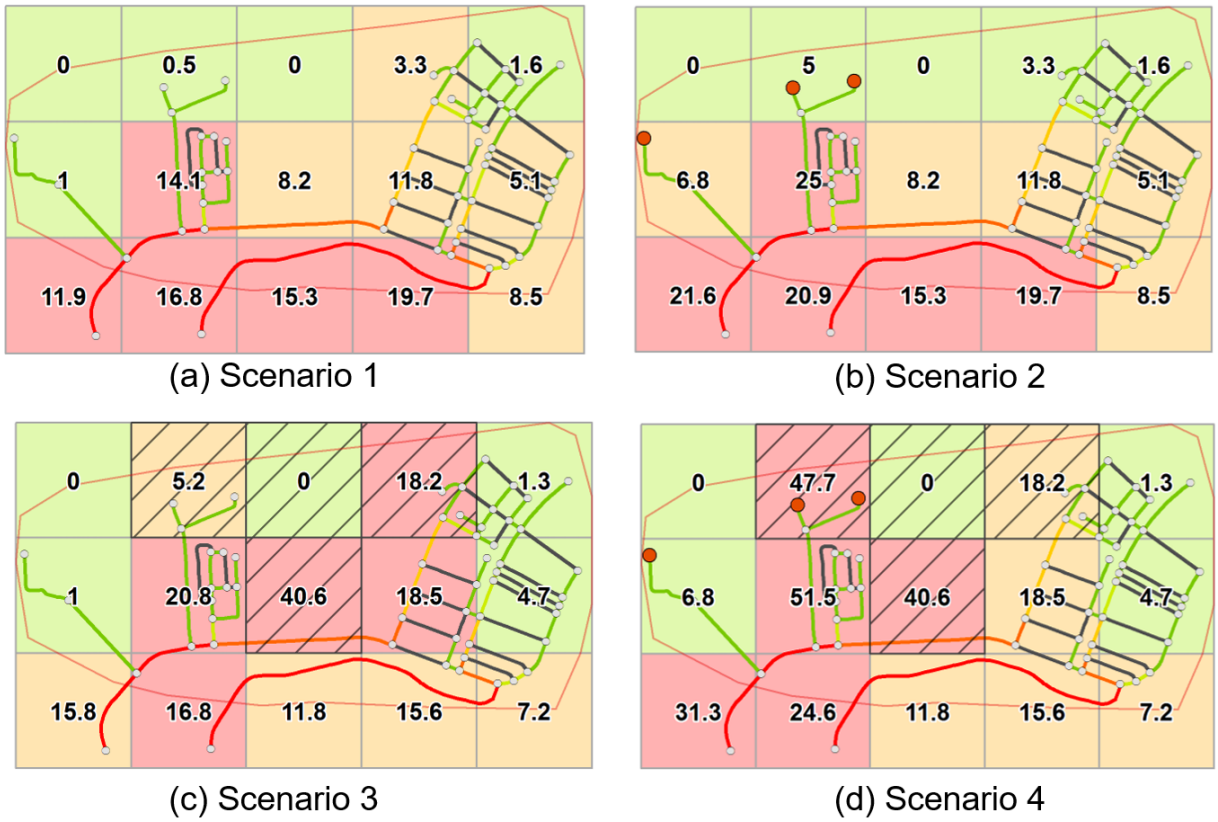


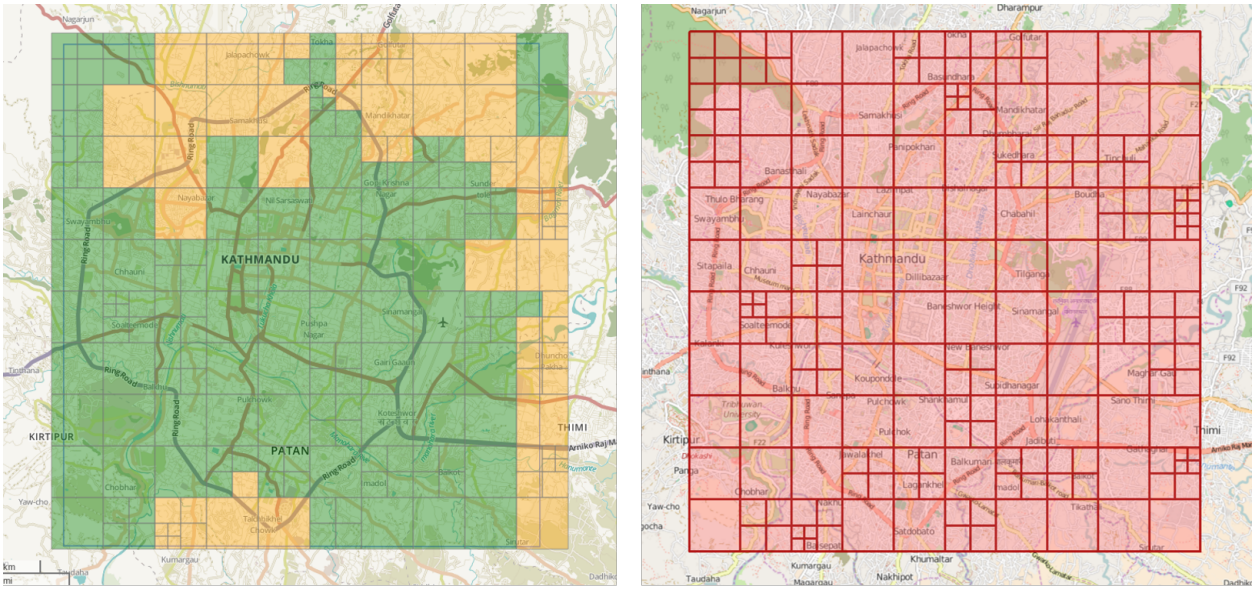
Figure 13: Different priority rankings in the four simulated scenarios.

scenario 1. This is because knowing the road connectivity information in these cells can help increase the probabilities of success for these important trips that serve large numbers of people.

In scenario 3, a higher normalized disaster severity value, 0.3, has been assigned to four cells (highlighted in the figure), and other conditions remain the same as in scenario 1. The increased disaster severity has two effects: first, it increases the uncertainty of the road connectivity within the severely affected cells; second, it increases the number of affected people at the nodes within these cells (e.g., in cells $(1, 2)$ and $(1, 4)$, each node has 30 (100×0.3) units of affected people). It can be seen that the cells (e.g., $(2, 3)$) that (partially) contain the trips passing (or going to) the severely affected areas have an increase in their information value compared to scenario 1. This is because knowing the road connectivity information in these cells can reduce the uncertainty and therefore improve the expected utility to a larger degree. In scenario 4, we combine the conditions from scenario 2 and 3. It can be observed that cell $(1, 2)$ has a large increase in its information value, due to the overlapping effects of large population and high disaster severity.

5.2 A case study based on a real mapping project from the 2015 Nepal earthquake

We conduct an experiment based on a real online mapping case which took place in the aftermath of the 2015 Nepal earthquake and was organized by the Humanitarian OpenStreetMap Team. This mapping project focused on the capital city of Nepal, Kathmandu, which was heavily affected by the earthquake. Figure 14 (a) is a screenshot of this real example, and we reproduced the same grid tessellation as shown in Figure 14 (b).



(a) Real mapping case (b) Reproduced grid tessellation

Figure 14: The real online mapping case for Kathmandu and the reproduced grid tessellation.

5.2.1 Dataset

Existing road network. The road network data for Kathmandu was downloaded from *Planet.osm*[2], which maintains a full history of the OpenStreetMap (OSM) data. As OSM data in Nepal has been greatly enhanced after the earthquake (Poiani et al., 2016), we retrieved the OSM data dump on April 20, 2015 (the latest data before the earthquake) to reproduce the initial situation before the mapping started. The retrieved OSM data covers the entire planet, and its size, after unzipping, is about 1.2 Tb. We used a bounding rectangle to clip out the data in the study area, and extracted the road network from the clipped OSM data. The extracted road data contain $21,232$ line features (Figure 15 (a)).

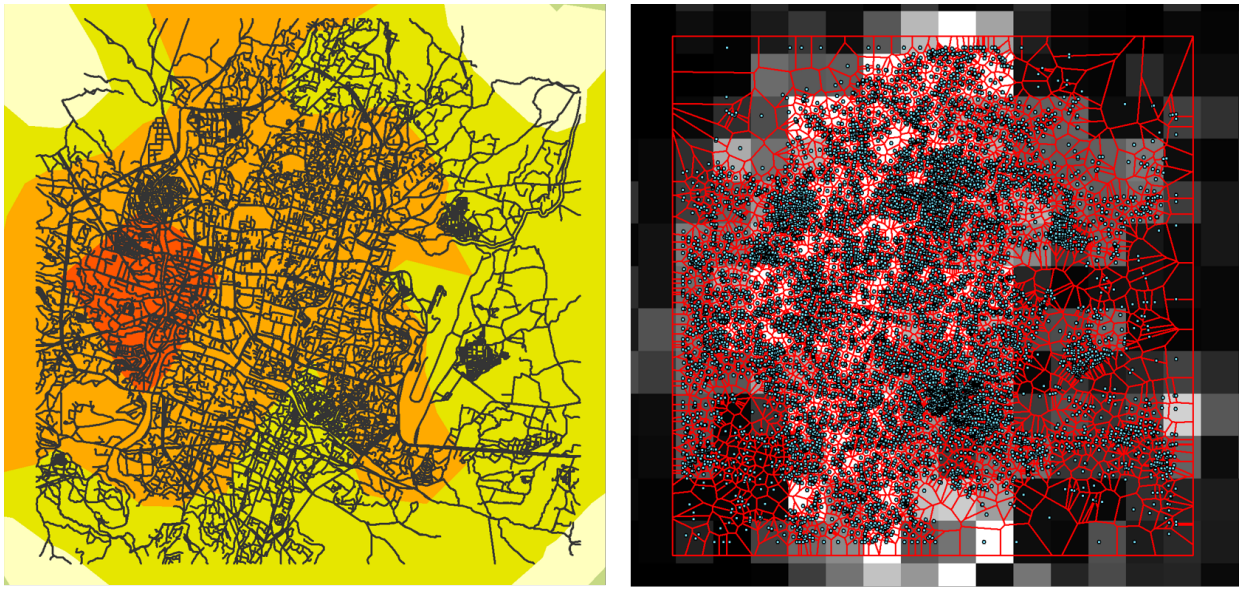
Disaster affected area. The disaster affected area in this case is larger than the target region to be mapped. Thus, we consider the entire target region as the affected area.

Population distribution. We downloaded 2014 LandScan data in the study area, which is the most recent LandScan data available. Compared with district-level population data,

[2] `http://planet.openstreetmap.org/planet/full-history/2015/`

LandScan has higher spatial resolution (about 1 km). We used a Voronoi-based method to aggregate the LandScan data to the nodes of the road network (Figure 15 (b)).

Task grid cells. We manually recreated the grid tessellation based on the real scenario. We also downloaded the earthquake intensity data from USGS[3], which has been visualized as the background map in Figure 15 (a). It can be seen that the west part of Kathmandu was affected more severely than the east part. We aggregated the intensity values to grid cells, and normalized the data into the range of $[0, 1]$.



(a) Road network and the earthquake intensities

(b) Voronoi polygons and LandScan data

Figure 15: Road network, earthquake intensity, LandScan data, and voronoi polygons generated based on road nodes.

In addition to the above datasets, we also retrieved the dates and times when online volunteers finished mapping each grid cell. In this way, we can compare the priority ranking generated by our framework with the actual order in which online volunteers mapped these cells.

5.2.2 Results and discussion

In total, there are 208 grid cells that need to be ranked. We ran our heuristic-based implementation on the prepared datasets. We also conducted a small-scale parallel computing using 4 separate program threads, each of which calculated the information value for 50 cells (the 4th thread handled 58 cells). The experiment took 141 minutes and the total time used by the 4 threads was 273 minutes. It is worth noting that the grid-based tessellation provides a natural basis for parallel computing, and in an urgent scenario, we can in principle distribute this work to as many as 208 computers, so that each computer only needs to work on one cell.

[3] `http://earthquake.usgs.gov/earthquakes/eventpage/us20002926#impact_shakemap`

Based on the derived information values, we classified the cells into 5 ranks based on quantiles. We also classified the actual order in which online volunteers mapped these cells into 5 ranks. The result is visualized in Figure 16. The red and orange colors indicate that these cells should have been (or were) mapped first, whereas yellow and the two green colors represent cells that should have been (or were) mapped later. For reference, we also show the road network in Figure 16 (a).

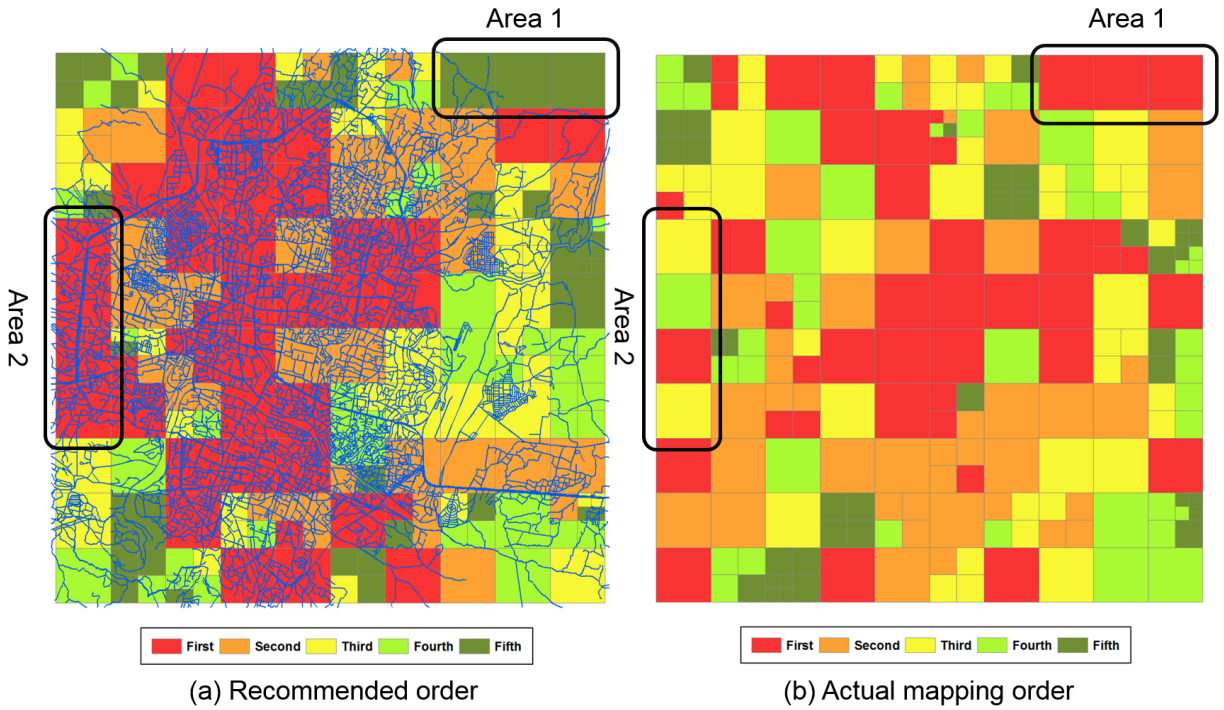


Figure 16: Rankings generated by the proposed framework and the actual mapping order by online volunteers.

Two different orders can be observed. Our framework produces a ranking whose spatial configuration is closely related to the road network structure, population distribution, and disaster severity (e.g., the west part is more severely affected). The actual mapping order from online volunteers shows some relations to the population distribution (e.g., a number of cells at the center of the grid tessellation were mapped at the early stage). Strong differences can be observed in *Area 1* and *Area 2*. In *Area 1*, three cells, which were ranked as low priority by our framework, were mapped first by online volunteers. The result from our framework is understandable, since this area has only a few roads and low population based on the LandScan data. In *Area 2*, three out of four cells that were ranked as high priority by our methods have been mapped in a relatively later stage by online volunteers. We can observe a dense road network in these three cells, and roads in these three cells are also important for accessing the severely affected area based on the USGS data. The reasons that online volunteers mapped cells in this particular order need further investigation. Our framework provides one possible way to rank the priorities of the grid cells.

To quantify the difference between the two rankings, we use Spearman's correlation

coefficient. A ρ of 0.333 ($p < 0.001$) has been observed, which indicates a weak similarity between these two rankings. This weak similarity could be explained by the cells, such as those in the central area of the grid tessellation, whose priorities have been ranked as high by our framework and which have also been mapped early by the online volunteers.

In addition, we perform Global Moran's I to examine the spatial autocorrelations of these two rankings. Queen's case has been used to define the neighborhood for Moran's I. A value of 0.358 ($p < 0.001$) has been observed for the ranking from online volunteers, while a value of 0.445 ($p < 0.001$) has been observed for the ranking generated by our framework. This result is interesting, as it shows our priority ranking has a higher degree of spatial autocorrelation than the actual mapping order of volunteers. Although a higher spatial autocorrelation does not necessarily mean a better ranking, our ranking might better reflect the effect of the earthquake, which is a spatially continuous phenomenon. This result also indicates that the order in which online volunteers map the affected area is not entirely random, but does have a small degree of spatial autocorrelation.

6 Conclusion and future work

In recent years, online volunteers have been actively involved in disaster response. One important contribution has been to map the disaster-affected areas based on remote sensing images. Typically, the affected area is divided into a number of cells using a grid-based tessellation. Online volunteers can select a cell to start the mapping process. However, this approach does not differentiate the priorities of the grid cells and thus volunteers may map these cells in a more or less random order.

In this research, we proposed an analytical framework based on information value theory for quantifying the value of information contained in these grid cells to emergency responders. Our objective is to provide guidance for online volunteers so that the cells that contain more important information can be mapped first. Specifically, we focused on road network connectivity information which has been frequently used in disaster response for planning relief trips. We described the details of our framework, and designed and implemented heuristics to enhance its computational efficiency. We also performed simulations to examine the ranking variations under different scenarios, and applied our method to an experiment based on a real mapping case from the 2015 Nepal earthquake. While in this paper the priority ranking has been proposed for online volunteers, the approach could also be used by agency professionals when ground-based surveying and mapping are necessary after a disaster.

This research, however, has its limitations that can be addressed in future work. First, so far we have adopted simple methods to estimate the uncertainty of road connectivity (based on the length of the road). Many other factors, such as the texture and width of the roads, and the structural characteristics of buildings, can also influence the uncertainties. Therefore, more accurate modeling strategies could be developed to include these factors into the current framework. However, strategies seeking higher accuracy may come at the cost of higher computational complexity. If all we need is a priority ranking, simpler methods may already produce a ranking that is as good as that of the more complicated models. Second, the current framework ranks the priorities of grid cells based on only the road network infor-

mation. Since a cell can contain many other types of geographic information (e.g., locations of shelters), we can extend this framework to accommodate these additional information types by identifying the corresponding potential target decisions. Third, the capability of the framework to re-prioritize grid cells on the fly may need further examination. In online disaster mapping, volunteers may respond quickly and fresh geographic information may be collected that could change the priority ranking generated previously. Such a situation requires our framework to re-prioritize the grid cells based on the newly available information. While our framework allows parallel computing and thereby should be able to complete the reprioritization in a short time period, further experiments are necessary to examine and quantify the performance. Despite these and other potential shortcomings, we hope that this research will make a modest contribution towards improving the efficiency of disaster response.

References

Burns, R., 2014. Moments of closure in the knowledge politics of digital humanitarianism. Geoforum 53, 51–62.

Chen, C. J., Chen, S., Su, X., 2001. Is accounting information value-relevant in the emerging chinese stock market? Journal of International Accounting, Auditing and Taxation 10 (1), 1–22.

Comfort, L. K., Ko, K., Zagorecki, A., 2004. Coordination in rapidly evolving disaster response systems the role of information. American Behavioral Scientist 48 (3), 295–313.

Cova, T. J., Church, R. L., 1997. Modelling community evacuation vulnerability using gis. International Journal of Geographical Information Science 11 (8), 763–784.

Crooks, A., Wise, S., 2011. Modelling the humanitarian relief through crowdsourcing, volunteered geographical information and agent-based modelling: a test case-haiti. In: Proceedings of the 11th International Conference on GeoComputation. pp. 183–187.

Dobson, J. E., Bright, E. A., Coleman, P. R., Durfee, R. C., Worley, B. A., 2000. Landscan: a global population database for estimating populations at risk. Photogrammetric engineering and remote sensing 66 (7), 849–857.

Fiedrich, F., Gehbauer, F., Rickers, U., 2000. Optimized resource allocation for emergency response after earthquake disasters. Safety science 35 (1), 41–57.

Forrest, B., 2010. Technology saves lives in haiti. Forbes.com .

Goodchild, M. F., 2007. Citizens as sensors: the world of volunteered geography. GeoJournal 69 (4), 211–221.

Goodchild, M. F., Glennon, J. A., 2010. Crowdsourcing geographic information for disaster response: a research frontier. International Journal of Digital Earth 3 (3), 231–241.

Graham, M., 2010. Neogeography and the palimpsests of place: Web 2.0 and the construction of a virtual earth. Tijdschrift voor economische en sociale geografie 101 (4), 422–436.

Hägerstrand, T., 1970. What about people in regional science? Papers in regional science 24 (1), 7–24.

Haklay, M., Weber, P., 2008. Openstreetmap: User-generated street maps. Pervasive Computing, IEEE 7 (4), 12–18.

Haworth, B., 2016. Emergency management perspectives on volunteered geographic information: Opportunities, challenges and change. Computers, Environment and Urban Systems 57, 189 – 198.
URL http://www.sciencedirect.com/science/article/pii/S0198971516300175

Heinzelman, J., Waters, C., 2010. Crowdsourcing crisis information in disaster-affected Haiti. US Institute of Peace.

Hester, V., Shaw, A., Biewald, L., 2010. Scalable crisis relief: Crowdsourced sms translation and categorization with mission 4636. In: Proceedings of the first ACM symposium on computing for development. ACM, p. 15.

Howard, R. A., 1966. Information value theory. Systems Science and Cybernetics, IEEE Transactions on 2 (1), 22–26.

Howe, J., 2006. The rise of crowdsourcing. Wired magazine 14 (6), 1–4.

Hu, Y., Janowicz, K., 2015. Prioritizing road network connectivity information for disaster response. In: Proceedings of the 1st Workshop on Emergency Management. EM-GIS 2015. ACM, New York, NY, USA, pp. 1–4.

Hu, Y., Janowicz, K., Chen, Y., 2015. Task-oriented information value measurement based on space-time prisms. International Journal of Geographical Information Science 30 (6), 1228–1249.

Humanitarian OpenStreetMap Team, O., 2013. Openstreetmap and yolanda: A report from manila.
URL https://hotosm.org/updates/2013-12-05_openstreetmap_and_yolanda_a_report_from_manila

Liu, S., Ziemke, J., 2013. From cultures of participation to the rise of crisis mapping in a networked world. The participatory cultures handbook , 185–196.

McFall, R. M., Treat, T. A., 1999. Quantifying the information value of clinical assessments with signal detection theory. Annual review of psychology 50 (1), 215–241.

Meier, P., 2011. New information technologies and their impact on the humanitarian sector. International review of the Red Cross 93 (884), 1239–1263.

Meier, P., 2012a. Crisis mapping in action: How open source software and global volunteer networks are changing the world, one map at a time. Journal of Map & Geography Libraries 8 (2), 89–100.

Meier, P., 2012b. What was novel about social media use during hurricane sandy. URL http://irevolution.net/2012/10/31/hurricane-sandy/

Meier, P., 2015. Digital humanitarians: how big data is changing the face of humanitarian response. CRC Press.

Meier, P., Munro, R., 2010. The unprecedented role of sms in disaster response: Learning from haiti. SAIS Review of International Affairs 30 (2), 91–103.

NPR, 2015. Virtual volunteers use twitter and facebook to make maps of nepal. URL http://www.npr.org/sections/goatsandsoda/2015/05/05/404438272/virtual-volunteers-use-twitter-and-facebook-to-make-maps-of-nepal

Ochoa, S. F., Santos, R., 2015. Human-centric wireless sensor networks to improve information availability during urban search and rescue activities. Information Fusion 22, 71–84.

Poiani, T. H., dos Santos Rocha, R., Degrossi, L. C., de Albuquerque, J. P., 2016. Potential of collaborative mapping for disaster relief: A case study of openstreetmap in the nepal earthquake 2015. In: Proceedings of the Hawaiian International Conference on System Sciences. p. 11.

Rogstadius, J., Vukovic, M., Teixeira, C., Kostakos, V., Karapanos, E., Laredo, J. A., 2013. Crisistracker: Crowdsourced social media curation for disaster awareness. IBM Journal of Research and Development 57 (5), 4–1.

Russell, S. J., Norvig, P., 2010. Artificial intelligence: a modern approach. Prentice hall Upper Saddle River, NJ.

Shanley, L., Burns, R., Bastian, Z., Robson, E., 2013. Tweeting up a storm: the promise and perils of crisis mapping. Available at SSRN 2464599 .

Shannon, C. E., Weaver, W., 1949. The mathematical theory of communication. University of Illinois Press, IL.

Wohltman, S., 2010. Haitian earthquake emphasizes danger of a split geo community. URL http://geosquan.blogspot.com/2010/01/haitian-earthquake-emphasizes-danger-of.html

Zhang, P., Zhang, H., Guo, D., 2015. Evacuation shelter and route selection based on multi-objective optimization approach. In: Proceedings of the 1st Workshop on Emergency Management. EM-GIS 2015. ACM, New York, NY, USA, pp. 1–5.

Ziemke, J., 2012. Crisis mapping: The construction of a new interdisciplinary field? Journal of Map & Geography Libraries 8 (2), 101–117.

Zook, M., Graham, M., Shelton, T., Gorman, S., 2010. Volunteered geographic information and crowdsourcing disaster relief: A case study of the haitian earthquake. World Medical & Health Policy 2 (2), 7–33.