Article

Towards Development of a Real-Time Point Feature Quality Assessment Method for Volunteered Geographic Information Using the Internet of Things

Sepehr Honarparvar [1,*], Mohammad Reza Malek [1], Sara Saeedi [2] and Steve Liang [2]

[1] Faculty of Geodesy and Geomatics Engineering, K.N. Toosi University of Technology, Tehran 19697-64499, Iran; mrmalek@kntu.ac.ir

[2] Department of Geomatics Engineering, University of Calgary, Calgary, AB T2N 4V8, Canada; ssaeedi@ucalgary.ca (S.S.); steve.liang@ucalgary.ca (S.L.)

* Correspondence: sep.honarparvar@mail.kntu.ac.ir; Tel.: +98-21-88877555

Abstract: One of the most important challenges of volunteered geographic information (VGI) is the quality assessment. Existing methods of VGI quality assessment, either assess the quality by comparing a reference map with the VGI map or deriving the quality from the metadata. The first approach does not work for a real-time scenario and the latter delivers approximate values of the quality. Internet of Things (IoT) networks provide real-time observations for environment monitoring. Moreover, they publish more precise information than VGI. This paper introduces a method to assess the quality of VGI in real-time using IoT observations. The proposed method filters sensor observation outliers in the first step. Then it matches sensors and volunteers' relationships in terms of location, time, and measurement type similarity using a hypergraph model. Then the quality of matched data is assessed by calculating positional and attribute accuracy. To evaluate the method, VGI data of the water level and quality in Tarashk–Bakhtegan–Maharlou water basin is studied. A VGI quality map of the data is assessed by a referenced authoritative map. The output of this step is a VGI quality map, which was used as a reference to check the proposed method quality. Then this reference VGI quality map and the proposed method VGI quality map are compared to assess positional and attribute accuracy. Results demonstrated that 76% of the method results have less than 20 m positional error (i.e., difference with the reference VGI quality map). Additionally, more than 92% of the proposed method VGI data have higher than 90% attribute accuracy in terms of similarity with the reference VGI quality map. These findings support the notion that the proposed method can be used to assess VGI quality in real-time.

Keywords: internet of things; volunteered geographic information; data quality assessment; wireless sensor network; VGI quality assessment

1. Introduction

The significant advances of Web2.0 technologies and mobile device proliferation have enabled none professionals to collect and map spatial data. This phenomenon has been described with different terminologies: volunteered geographic information (VGI) [1], neogeography [2], crowdsourcing [3], citizen science [4], and user-generated content [5]. Although they have shared the same concept in most research, some studies conclude that they deliver different ideas [6]. VGI is defined as the collaborative acquisition of geographical information and local knowledge by "volunteers". Additionally, Goodchild argued VGI as the data collected by "Citizens who act as sensors" [1]. VGI falls in the active and spatial categories [7]. Van Exel et al. asserted that "VGI is social information with spatial dimension" [8]. Capineri et al. claimed that VGI is citizen science in case volunteers tag and annotate spatial data [9]. Haklay et al., [10] introduced "geographic citizen

science" as a specific type of VGI when collecting location information is essential for an activity (e.g., reporting observations of different bird species). In this paper, we used the latter definition, and the proposed method focuses on this type of VGI. In the case study discussed in this paper, volunteers are collecting point-based water quality information along with location information.

VGI enjoys substantial advantages over governmental or authorized databases: it is low price, up-to-date, and gathers the local information that could have never been collected otherwise [11]. Despite having such huge advantages, VGI suffers from a lack of quality control [12]. The data contributed by volunteers does not necessarily comply with existing spatial data standards [11]. Therefore, techniques for assessing VGI quality are different from traditional methods of assessing spatial data quality [13–15].

1.1. Related Works

In this research domain, two general approaches are defined for evaluating the quality of VGI. In the first approach, the quality is assessed by comparing VGI to a reference map and generates quantitative results [16] (referenced approach). The reference map includes information, which is assumed to have higher accuracy. Sometimes the acceptable accuracy is interpreted as the difference between produced data and the perfect or ideal data (internal quality) and in the other case, it is indicated as the capability of data to satisfy users' needs (external quality). The second approach does not use any reference map to assess the quality of VGI (non-referenced approach). In this approach, various methods are used to assess the quality by considering information about the history of volunteers' contributions, volunteers' location, or environment [17].

The referenced approach often uses authoritative data and quantitative calculations to assess the quantitative quality of VGI [18]. OpenStreetMap (OSM) is one of the most popular VGI platforms to study VGI quality [19]. OSM has been widely investigated using the "famous five" elements of VGI quality [20,21]. They include positional accuracy, thematic accuracy, temporal accuracy, logical consistency, and completeness. After comparing the OSM map of England and Ordnance Survey Meridian data, Haklay [22] demonstrated that 80% of OSM features have a good representation of the authoritative data in terms of positional accuracy. In another research, Zielstra and Zipf [23] found out that 60% of OSM map features in Germany are almost fitted for use. In a study on France OSM maps, Girres and Guillaume [24] found out that 77% of OSM features have positional accuracy of less than 6 m. Another study by Zhang and Malczewski [19] on Canada OSM roads data revealed that 60% of features have 25 m positional accuracy while their attribute accuracy was variant between 0.52 and 0.9. calculated 25 m of positional accuracy and 52%–90% of attribute accuracy for 60% of Canada OSM roads. All the research investigates specific datasets (i.e., OSM) for a specific area. These studies show the uncertainty of VGI by comparing OSM with a high accuracy map. However, the most important challenge of the referenced quality assessment approaches is map matching between the authoritative map and the OSM map. Koukoletsos [18] employed a vector-based method to find corresponding roads between two maps. He defined a method to break lines into segments and then, measured geometry and orientation similarities between them. Afterward, he evaluated attribute similarity between roads that are selected as corresponding roads in the previous step. His method provides solutions for line-based features and an authoritative map is required. In another approach, Mohammadi and Malek [12] introduced the location-orientation rotary descriptor as a raster-based algorithm to find corresponding street segments between OSM and the authoritative map. The method only works for raster datasets and only considers geometrical and positional aspects for matching. An authoritative map is required. Assessing the quality of features in these studies depends on the geometry type of feathers. For instance, in point-based features, Euclidean distance determines the accuracy [17] while for line-based features, the intersected area of line buffers of corresponding features (i.e., the reference map feature and the VGI feature) indicates

the accuracy of the feature [15]. The referenced approach is not suitable when the reference data is not available, or not updated, or does not contain the needed information. For example in crowdsourcing disaster management applications, accessing real-time data of events is vital [23]. In another case, a recommender system for tourism needs to update trustworthy information of crowds to recommend correct items to people [17]. Since in these use-cases, authoritative information is not accessible, the referenced approach cannot be used to validate or filter out VGI based on quality elements.

In the second approach, the VGI quality is estimated based on the inherent characteristics of data. Therefore, the approach harnesses other information derived from volunteers and the environment data to evaluate VGI quality. Van Exel et al. [24] employed collective intelligence and volunteers' history to evaluate the quality. However, the volunteer's history does not guarantee the recent contribution quality. Barron et al. [25] designed a framework to assess the quality based on volunteers' contributions history by comprising six groups of 25 indicators of volunteers' history. They have not tested the method with other VGI dataset. Indicators are calculated using metadata not data. Additionally, indicators are calculated using metadata not data. Senaratne et al. [26] explored volunteers' line of sight as an important indicator to verify data credibility. This method requires the digital elevation model (DEM) or digital terrain model (DTM). Jilani et al. [27] used a machine-learning algorithm to find topological and geometrical characteristics for road semantic classes. Then, they identified the VGI quality by matching and comparing the semantic class of the road specified by a volunteer and the topological or geometrical class of the same road. This solution specifically works for road network VGI data. Albuquerque assessed the quality of volunteer's information about incidences based on a risk map and probability of an incidence on the location of the volunteer [28]. The estimated VGI quality in that paper only reflects the probability of quality. Hung et al. [29] detected VGI clusters and volunteers' location relationships using spatial analysis. In this method, information located inside those clusters has more quality. All these quality assessment methods are considered as an estimation for the VGI quality; as they do not benefit from a VGI source or any reference to evaluate the accuracy. To conclude, the referenced approaches do not assess the VGI quality in real-time and the non-referenced approaches give an approximate quality value in comparison with referenced approaches because referenced approaches are calculated by the difference between the position of authoritative and VGI data while the non-referenced are not. Therefore, it is essential to find a new approach to assess the deterministic value of VGI quality in real-time.

In recent decades, IoT has been applied in various environmental applications [30,31] climate monitoring [32], flood warning [33], earthquake emergency [34,35], water quality [36], and recommender system [37]. International Telecommunication Union (ITU) stated the IoT as "A global infrastructure for the information society enabling advanced services by interconnecting (physical and virtual) things based on, existing and evolving, interoperable information and communication technologies" [38]. IoT is different from the wireless sensor network (WSN) in various aspects, such as communication protocols, security and privacy, scalability, robustness, and service quality (QoS) as described in the literature. Although they are both influenced by machines to machines (M2M) architecture, IoT uses internet protocol (IPv6) while WSN can use other communication protocols as well [39]. IoT communication protocols provide better results for real-time data observations because of less battery consumption and network latency [40]. Additionally, there are fewer deployment constraints (i.e., location and topology) in IoT networks [40]. Therefore, IoT is the better approach to distributing real-time data over the internet and building a VGI solution. In this paper, IoT implies a network of connected things (i.e., viscosity, water level, and temperature sensors) to the internet.

IoT provides real-time observations and, automatically, transmits information to other things. Unlike VGI, the information sources in IoT networks are identifiable [41]. Identifying the source of information would enable us to prevent errors in a calibration process. Sensor error sources are mostly identifiable while VGI error sources often are

unknown. Moreover, sensor data noises are lower than VGI [42] and more reliable than volunteered-generated data [43]. Therefore, the information which is obtained by an IoT system has more quality rather than VGI and can be applied to validate VGI quality [44]. Moreover, the datasets with higher quality can be considered as a reference for quality evaluation of the fewer quality data [45]. Accordingly, IoT can be a reference to evaluate the quality of VGI in real-time. Here, being real-time means that the VGI quality assessment process can be accomplished right after receiving data from volunteers and does not require post-processing. Emerging publish/subscribe has enabled us to trigger the VGI quality assessment process right after subscribing to the recent VGI data. Based on some literature, this can be considered as a near real-time process in case the process takes between 5 and 15 min [46]. Although using historical data can improve VGI quality assessment as well, it is beyond the scope of this paper and can be considered for future works. However, in this paper being real-time means that the process of a quality assessment of VGI is automatically triggered right after receiving data from volunteers and does not require post-processing. Several theories have been proposed to integrate IoT or sensor networks with VGI. Bakillah et al. [47] fused sensor data and VGI to create a more realistic transportation simulation. They discovered that the spatial accuracy of VGI is a concern and an appropriate technique for assessing the VGI quality is essential. Horita et al. [48] developed a spatial decision support system for flood risk management by combining VGI and sensor data. They showed that VGI fills the lack of data about river basins when sensors do not cover the area [48]. Fontes et al. [49] presented an approach to design a Web GIS (geographic information system) emergency response platform that integrates both sensors and VGI. The prior studies substantiate the belief that VGI fills the gaps of information in space while they neglect the VGI quality and methods of filtering them.

1.2. Objectives and Contributions

Previous works on VGI data quality assessment fall into two categories: The main goal of this paper is to demonstrate how to assess the VGI quality of point features using the IoT network in real-time. In other words, the proposed method fills the gaps in the previous works on how to use IoT to identify and remove unsuitable VGI data in real-time. The main purpose of data quality assessment using IoT sensors is to use sensors for ranking volunteers' credibility, removing fake or incorrect volunteers' reports, and use qualified VGI for adding more details to the sensor observations. In general, this study is expected to deliver the following benefits:

- This quality assessment method leads us to validate volunteers' credibility. The real-time quality assessment can assign scores to volunteers and their generated data during an automatic real-time process. Therefore, it would be a good approach to automatically check volunteers' contributions credibility and score them accordingly when they are located in the proximity of qualified IoT sensors. Moreover, it can be deployed in an integrated IoT-location based social network (LBSN) to check the reliability of volunteers' observations in the long term and automatically score volunteers based on previous contributions quality. Volunteer's score history can help to verify volunteers' observations when no other reference is available or there is not enough time to check the quality (e.g., disaster management case). Applying the proposed method help reducing fake and incorrect reports propagation in social networks by assessing the quality of volunteers' reports. Therefore, IoT can filter the wrong news in social networks.
- Quality assessment of VGI using qualified IoT sensors, as a new method can help scholars to ascertain how reliable a VGI is to generate the information.
- If we can evaluate and verify a high-quality VGI report using the matched IoT sensor, then we can add additional information provided by volunteers to the corresponding sensor. The additional information may include more descriptive details, a higher

observation rate, and extra locations. As volunteers are able to share descriptive information, they can generate more detail about observations. For example, a sensor observes a reduction of water level in a river while a volunteer report could be "water level is low. I see an otter dam 20 m away". Therefore, it can be concluded that the reason for the low water level is probably the otter dam.

The main contributions of this paper are as follows:

- It proposes a framework for automating VGI positional and attribute accuracy assessment using qualified IoT sensors
- It delivers a new matching method for finding corresponding IoT nodes and VGI using location, time, and observed property (an observed property specifies the phenomenon of an observation such as temperature, light, etc.) [50] as three similarity criteria. Existing matching algorithms such as spatiotemporal clustering approaches are only limited to time and location similarities (as an example see [51]). While, the proposed method matches sensors and VGI by considering the unlimited number of similarity criteria such as time, location, observed property, etc.

This research focused on qualified environmental sensors (i.e., environmental sensors with adequate accuracy and propagation for VGI quality assessment). We assume that the location and time of the observations are known and IoT sensors are qualified with proper accuracy, time, and spatial distribution. The scope of this paper is point-based VGI. Quality assessment of line or polygon VGI using environmental sensors does not look affordable. Environmental sensors do not measure the geometrical features of objects and do not provide any metric to evaluate the geometrical accuracy of polygons and lines. On the other hand, finding the data quality of point VGI seems affordable by the proposed matching process.

1.3. Organization of the Paper

The paper is organized as follows: Section 2 explains the proposed methodology to assess the quality of VGI using sensors. In the Section 3, the proposed method is evaluated and results are discussed in Section 4. Finally, conclusions and future works are provided in the Section 5.

2. Proposed Methodology in VGI Quality Assessment

The proposed method is using the Internet of Thing data to assess the quality of volunteered geographic information in real-time. The method is developed based on the following assumptions:

- o Sensors are environmental and located in known positions.
- o The location and time of sensors and VGI observations are known.
- o The rate of VGI data is not regular.
- o VGI data are only reported in point geometry. They include the time, location, and a short description of the environment observations.

As described in the related works section, VGI quality assessment methods are based on referenced or non-referenced approaches to estimate the data quality. In the proposed method, as environmental sensors are used for the VGI quality assessment, the IoT network observations are assumed as a reference for VGI. Therefore, similar to the reference methods of VGI quality assessment, a matching algorithm is required to find the corresponding IoT network nodes and VGI. Figure 1 illustrates the steps of the proposed VGI quality assessment method. The proposed method includes three major steps, including identifying outliers in sensor data, matching corresponding sensors and with VGI, observation nodes, and VGI quality assessment. The method is developed based on the publish/subscribe concept. Therefore, by subscribing to a new VGI message, the quality assessment process will be triggered. After the outlier detection step, the VGI data will be compared with sensors for matching. Therefore, among the candidate sensors, one of

them is considered as the matched sensor. The following subsections are details about the steps.
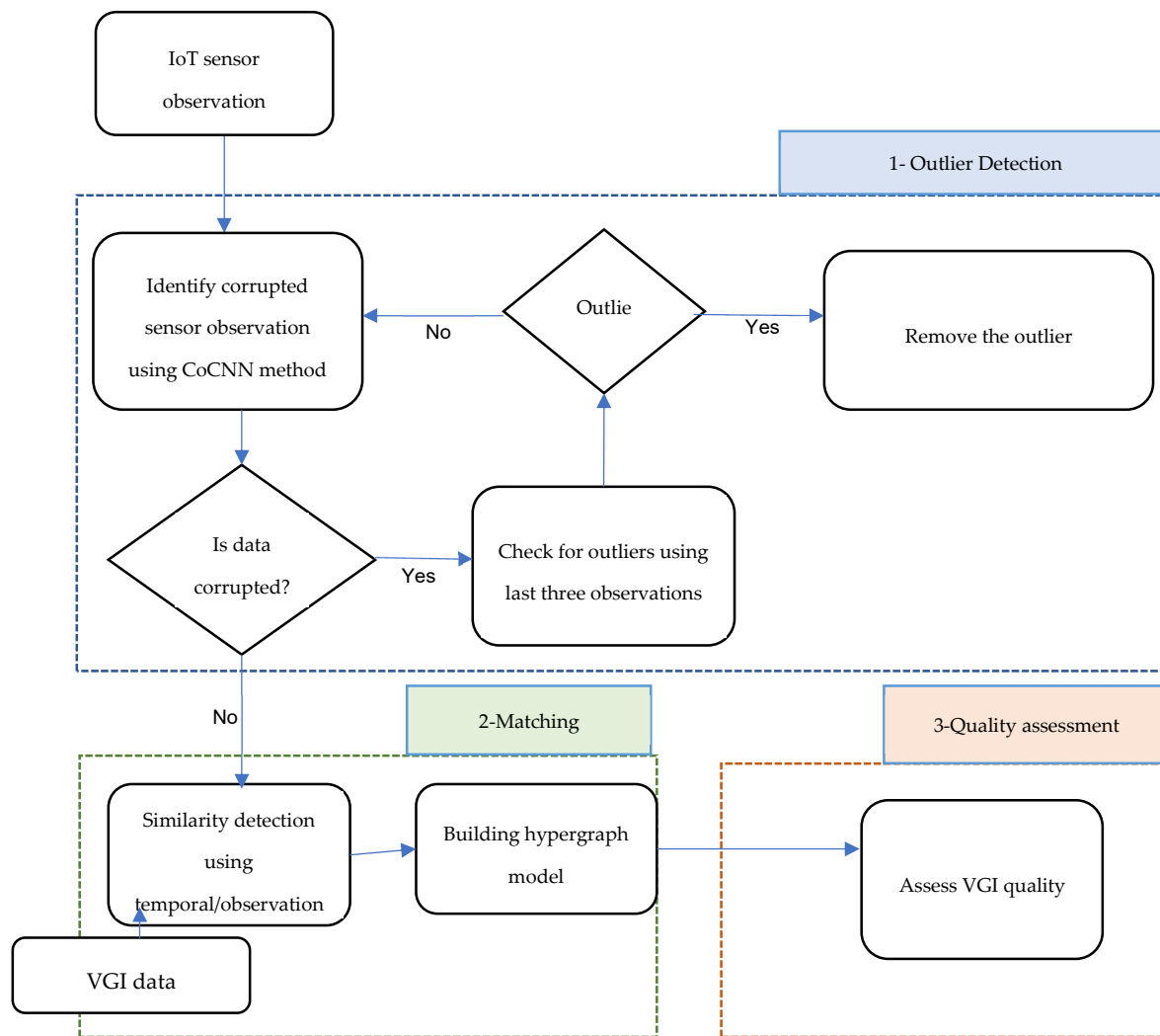
```
┌─────────────────┐
│   IoT sensor    │
│  observation    │
└─────────────────┘
         │
    ┌────────────────────────── 1- Outlier Detection ──────────────────────────┐
    │                                                                           │
    │  ┌──────────────────┐        ◇              ┌──────────────────┐          │
    │  │ Identify corrupted│  No   Outlie   Yes   │                  │          │
    │  │ sensor observation│◄─────   ◇     ─────► │ Remove the outlier│         │
    │  │ using CoCNN method│         ◇            └──────────────────┘          │
    │  └──────────────────┘                                                     │
    │         │                       ▲                                         │
    │         ▼                       │                                         │
    │      ◇ Is data ◇   Yes   ┌──────────────────┐                            │
    │      ◇corrupted?◇ ─────► │ Check for outliers│                           │
    │                          │ using last three  │                           │
    │                          │   observations    │                           │
    │         │                └──────────────────┘                            │
    └─────────┼─────────────────────────────────────────────────────────────────┘
             No
```

Figure 1. The proposed quality assessment method.

2.1. Outlier Detection in Sensor Data

Considering sensors as a reference for assessing VGI quality may cause unreliable results if sensor observations are corrupted. The observations corruption sources can be random, systematic, or blunders. Corrupted observations have an unpredictable value by the sensor, which causes random and blunder errors [52]. These errors generate outliers in observations. Then, we compared the corrupted observation data point with the last three data points for outlier detection [53]. Outliers are observations with a significant difference from other values, which indicates failure in observations [54]. However, these changes should be considered as anomalies if the difference shows ups as a shift in values [52]. In this paper, every corrupted data is firstly assumed as an outlier unless three new observations have similar values with the corrupted data [53]. Therefore, the outlier detection is run by every new sensor observation. If observations have been considered as an anomaly, all the following stages will be repeated to assess the quality with new observations.

To find outliers we followed the method proposed by [53], which are summarized as below:

o　A value would be predicted regarding the historical data. To do so the constructive neural networks (CoCNN) is applied as a semisupervised learning method. After training sample data using CoCNN, mean square error (MSE) [53] is calculated over T time intervals. The output of this step is a MSE and predicted value for the next observation.

o　Then measured value and predicted value are compared. If the difference between these two values exceeds the threshold Q, the value would be considered as an outlier. The Q threshold was specified based on the clever standard deviation (Clever SD) [55]. Therefore the Q threshold would be MSE±2.5SD.

To sum it up, in the first step, when a new sensor observation is generated, it is checked if it demonstrates a significant change in comparison to historical data. If so, it will be tagged as corrupted data and will be checked whether it is an outlier or not regarding the last three observations pattern. If it is an outlier it will be removed. Otherwise, it is considered as a correct observation in VGI assessment and sent to step 1 as historical data for future sensor measurement values checking.

2.2. Matching

The matching process matches every IoT node of the network to every volunteer data. The nodes of the IoT network can connect to others considering some constraints. As the IoT network nodes use batteries, IoT energy efficiency should be considered [19]. Therefore, one of the most critical constraints is the energy and sensors' life cycle [56]. To save energy in such networks, nodes should be in the sleep state during the time [57]. Consequently, matching the network nodes, which may hide for a while looks challenging. In sum, the matching is time and space dependent.

One of the possible matching solutions is to allocate an area for every sensor; then check if the area contains VGI locations. This is affordable as sensor locations are fixed and we can assign each VGI to one of the sensors. Therefore the environment should be partitioned into smaller polygons while there must be no overlaps or gaps between them. One of the most popular methods to create such polygons is Voronoi tessellation. It is a variational method that optimized the function of vertices of polygons [58]. Besides, it is a preferred partitioning method for computational fluid dynamics [59], partitioning ocean based on kinetic energy [60], land-use optimization [61], air pollution modeling [58], etc. Additionally, it is a solution for sensor network coverage estimation and optimization [62], retrieve information of things [63], and automating argumentations among IoT [64,65]. Nonetheless, a Voronoi polygon should include observations of one sensor to match the sensor and VGI. Then the result of the intersection of the polygons and the volunteer-generated information states whether the sensor and the VGI are corresponding or not.

As different types of sensor observations are involved in an IoT network, sensors' sleep time should be various for the network nodes. It leads sensors to asynchronously disseminate data in the IoT network. Therefore, a Voronoi diagram changes if a sensor turns into the sleep or inactive mode. Figure 2 illustrates the impact of activating a sensor on the Voronoi diagram. The sensor marked by the red circle on the left figure is "on", then, it is switched to sleep mode on the right figure. As a result, the Voronoi diagram structure changes and delivers larger polygons.
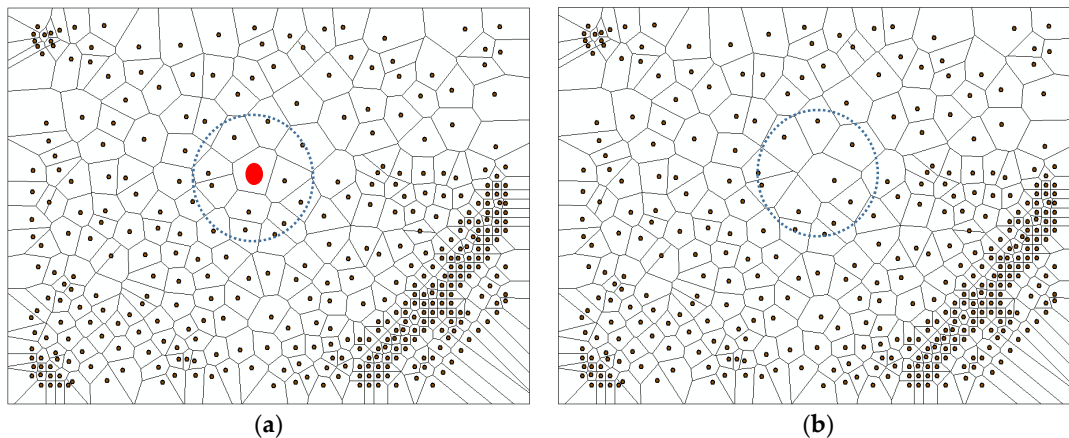
Figure 2. Voronoi diagram behavior when a sensor changes the status. (**a**) Voronoi diagram of the sensor network in active mode. (**b**) The Voronoi diagram in sleep mode.

This changing pattern is applied to several sensors that are in a different time, location, and different types of sensor observations. Consequently, the volunteer's location, time, and observed property similarity should be considered for matching. Therefore, the first step is to find spatial, temporal, and type similarities and relations. Then, a graph is built to model relations. Finally, the most similar values between the volunteers' node and other nodes are discovered and selected as the matched nodes. The following articulates steps of the matching process:

2.2.1. Step 1 (Similarities Detection)

In the method, VGI is allocated to the areas of sensors by Voronoi diagram partitioning to find spatial relationships. It makes every VGI to be located to a cell that is assigned to a sensor. Finally, a sensor can build a relationship with a VGI if the associated Voronoi polygon overlay a volunteer location. In this method, three criteria have been considered such as spatial, temporal, and observed property similarities. For every relationship, a score is considered to show how much a sensor is similar to volunteers in terms of the corresponding criteria.

To find a temporal relationship, the time intervals for active sensors (i.e., the time difference between two consecutive measurements) are measured and stored. To store volunteers' contribution time window, we assigned an estimated time interval (i.e., Δt_{VGI}) to the VGI observation for showing the duration that the volunteer spent in the location of observation. The time interval estimation (Δt_{VGI}) was calculated based on the average of known time intervals for previous sample VGI data in the application. All the time values were converted to UTC (universal coordinated time) based on the location and time zones of sensors and volunteers. The time intersection of sensor intervals with the volunteer's interval would build a relationship. For example, the red box in Figure 3 intersects (or overlaps) with sensor 1 and sensor 2. Therefore, the volunteer is linked to sensor 1 and sensor 2.
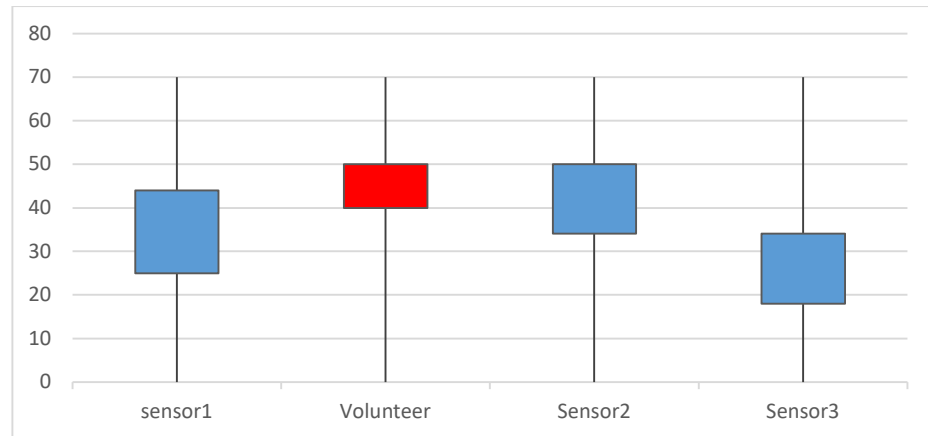
Figure 3. Sensors and volunteer temporal intersection.

To calculate the relationship score w(u,s) between a volunteer time interval (u) and a sensor time interval in (s), we used Equation (1). In the following equation, min(x) and max(x) indicate min and max values of the time interval x, and l(x) returns the length of the time interval x. It is a conditional equation to consider different situations of temporal relationships between sensors and volunteers' observations. For example, the first line of the equation considers a similar situation between sensor 1 and the volunteer in Figure 3. In this case, the difference of the maximum of the volunteer's time interval and the lower band of the sensor's time interval over the length of the volunteer time interval calculates the temporal relationship score. The more value of the score means the more temporal similarity between sensors and volunteers' observations.

$$W(u,s) = \begin{cases} \dfrac{max(s) - min(u)}{max\big(l(s), l(u)\big)} & \begin{aligned} &max(s) < max(u) \ \& \\ &min(s) < min(u) \end{aligned} \\[2em] \dfrac{max(u) - min(s)}{max\big(l(s), l(u)\big)} & \begin{aligned} &max(s) > max(u) \ \& \\ &min(s) > min(u) \end{aligned} \\[2em] \dfrac{l(s)}{l(u)} & \begin{aligned} &max(s) < max(u) \ \& \\ &min(s) > min(u) \end{aligned} \\[2em] \dfrac{l(u)}{l(u)} & \begin{aligned} &max(s) > max(u) \ \& \\ &min(s) < min(u) \end{aligned} \\[2em] 1 & \begin{aligned} &max(s) = max(u) \ \& \\ &min(s) = max(s) \end{aligned} \end{cases} \tag{1}$$

The sensor observation type is usually described by its metadata (e.g., pollution and temperature), while VGI observation type can be described linguistically by the volunteer. In this research, volunteers are assumed to share the type of observation using linguistic data rather than quantitative data. If a volunteer's type of information is similar to a sensor observed property, a relationship will be built between them. Figure 4 illustrates an example of relating sensors and volunteers in terms of observed property. For example, if a volunteer shares water pollution data, it will link with sensors, which measure water pollution.

Volunteers' observations data type is a string. To match it with corresponding sensors, a string matching algorithm was employed. The purpose of the observation type matching algorithm is to match all available IoT sensor observation types with word strings in the VGI report. All existing sensor observation types are stored in a metadata database. The database includes a dictionary of observed property names. Therefore, some sensors can connect to several strings with the same meaning. When a VGI observed property is shared, a query to search for similar strings is sent to the metadata database. Then, the connected sensors to similar texts are retrieved. The observation type matching algorithm is based on two steps: keyword extraction and keyword matching. First, words in the VGI string are cleaned up and then, they are matched with the sensor observation type values in the database. To clean the VGI string, all prepositions and punctuations are removed from the text. Then a vector of words is built. To match the values, Levenshtein is used as an "edit distance-based string-matching" algorithm (the reader may see [66–68]). Levenshtein calculates the distance between two strings by counting the number of edits that a string needs to be transformed into another one. Then the distance would be normalized to the maximum values of both strings. The volunteers' observation and sensors will be similar and be related if the normalized Levenshtein value is more than a predefined threshold, i.e., "0.75" as described in [15].

The relations are weighted because volunteers' and sensors shared data are not the same. For instance, an RGB color sensor detects dark color for water [69] while a volunteer mentions that the water is polluted. A volunteer can infer that water is polluted by considering several factors such as particles, low transparency, or observing alga in the water while the sensor only senses water transparency. Therefore, a probability function is leveraged for weighing sensors and VGI relations. Equation (2) presents how to calculate the probability based on observed property similarity. The similarity function is Jaccard index. The Jaccard index is a popular method for finding similar keywords and classification [69]. The most similarity value for P is 1 and the least one is 0.

$$P(u, s) = \frac{n(u \cap s)}{\max(n(u), n(s))} \tag{2}$$

In Equation (2), *n* function returns the numbers. The parameter *u* is a vector including items that are derived from volunteer's contributions. The s is a vector including items that are derived from sensor observations. For example, if we consider u = (transparency, particle observation, alga observation) and s = (transparency) the weight value would be 0.33.
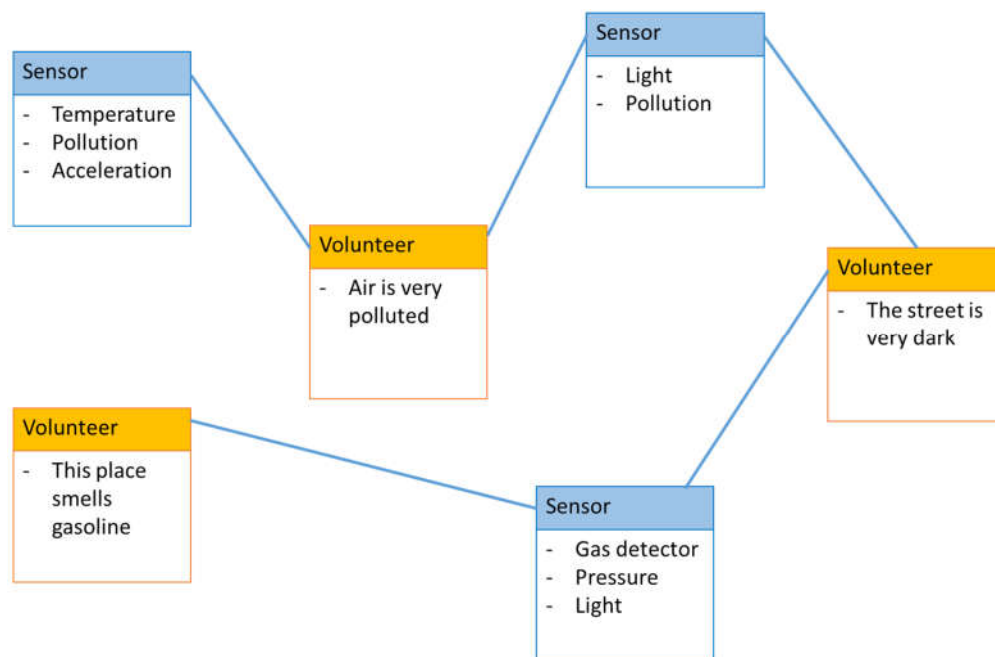
Figure 4. Relationships of sensor and volunteers based on observed property.

2.2.2. Step 2 (Building the Hypergraph Model Based on Relationships and Matching Sensors)

A graph can define the topology of a sensor network [70]. Therefore, graph theories are applied to the topology of the sensor network. In the graph, nodes are sensors and volunteers, and relationships are edges. In a simple graph, every edge connects exactly two vertices. Nevertheless, in the proposed model, every edge can connect more than two vertices. Additionally, the relationships between sensors and volunteers are complicated.

Hypergraphs are regarded as a generalization of graphs [71]. They build complex relationships models and take advantage of hyperedges which connect multiple vertices [72]. A hypergraph derives things relationships in an IoT network [73]. A schematic example of a hypergraph for the proposed method is illustrated in Figure 5. In this example, the s3 (sensor 3) is spatially connected to the user u1 while producing common observed property and share data in a similar time interval with u3, s4, and u2. e1, e2, and e3 are hyperedges of the hypergraph, which relates vertices.
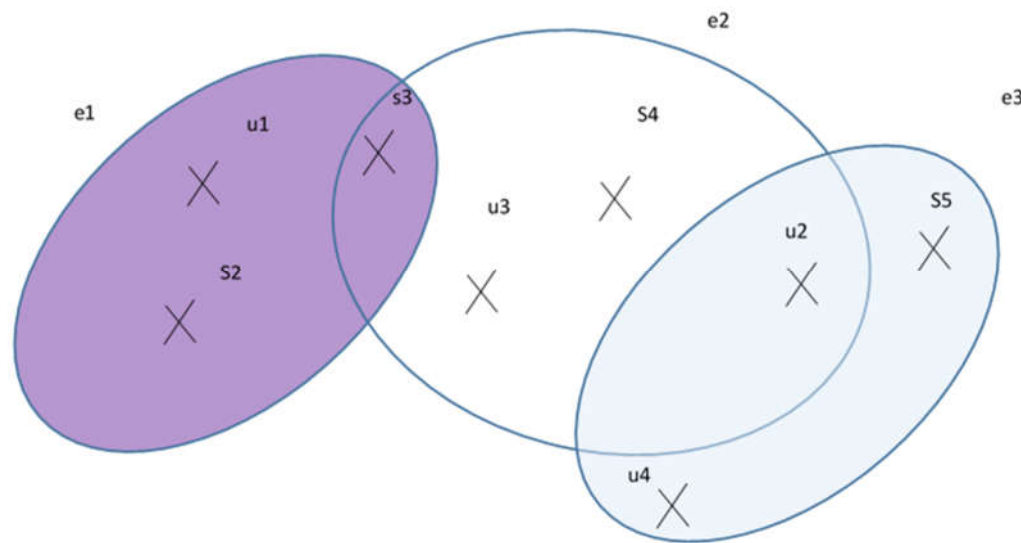
Figure 5. Hypergraph example of the proposed topology.

In a hypergraph, hyperedges are required to mathematically build relationships [74]. As hyperedges are complex and may contain several relationships, information loss is inevitable [75]. Therefore, optimization of information loss prevention is essential. The proposed model, which includes thing–thing and volunteer–things relationships is trying to reduce information loss through building relationships. To elaborate on volunteer–things relationships, the times of using a thing by a volunteer can be considered [73]. Yao et al., [75] provided a web interface to let users check and control things. The developed web service lets them count the interaction of users with things. Nonetheless, in the proposed model, users are volunteers who are assumed as information producers and relationships usage are the same as thing–thing. Consequently, deriving and optimizing thing–thing relationships is enough in this case.

To score the relationship of VGI and things, the hypergraph should become a graph. The challenge is to find the best relations while information loss is minimized. To find the best solution for making pairwise relationships, a machine learning algorithm was used. Machine learning algorithms are widely used to construct the derived graph from a hypergraph [76]. Scores of relations from a hyperedge are not available while they could be estimated under specific circumstances. Furthermore, providing a large train set is not possible. Therefore, a semisupervised learning approach is applied. Semisupervised algorithms perform high accuracy for realistic problems and benefit both supervised and unsupervised algorithms [77]. To define the train set relations, relations, which are certainly cannot-link or must-link are assumed as unlabeled and labeled relation scores [78]. The details of this algorithm are presented in the following paragraphs.

The system in [73] proposed a framework to decode thing–thing relationships from a hypergraph for a spatiotemporal context-aware IoT network. They tried to solve complex relationships by solving a relatedness ranking problem to minimize information loss in every relationship. They initially derive an incidence matrix of the hypergraph. The proposed hypergraph vertices are volunteers, things, locations, time intervals, and sensor observed properties. In the proposed method, volunteer–volunteer relations and thing–thing relations are not employed since the objective is to find only volunteer–things relations. Additionally, every sensor or volunteer connects to only one location, one-time interval, and one observed property. These rules can simplify the hypergraph construction. The incidence matrix (H) of our model is illustrated in Table 1, which is built based on the rules.

Table 1. Hypergraph incidence matrix of the proposed method.

Description	Observation Type	Location	Time	Volunteer	Thing
The Volunteer and the sensor are in the same location with the same info type at different time	IE^{SL}	LE^{SL}	0	UE^{SL}	SE^{SL}
The Volunteer and the sensor are at the same time with the same info type at a different location	IE^{ST}	0	TE^{ST}	UE^{ST}	SE^{ST}
The Volunteer and the sensor are at the same time in the same location with the same info type	IE^{SLT}	LE^{SLT}	TE^{SLT}	UE^{SLT}	SE^{SLT}
The Volunteer and the sensor are in the same location at a different time and info type	0	LE^{L}	0	UE^{L}	SE^{T}
The Volunteer and the sensor are at the same time in a different location and info type	0	0	TE^{T}	UE^{T}	SE^{T}
The Volunteer and the sensor are in the same info type at different location and time	IE^{S}	0	0	UE^{S}	SE^{S}

In Table 1, columns are vertices and rows are edges in the hypergraph. The matrix cell values are determined based on the similarity detection process. The cell values are similarity relationships between sensors and volunteers, which, also, are considered as graph edges. In other words, the edges belong to one of location, time, and observation type similarity graphs. If a vertex involves two kinds of relations, it will be involved in two graphs. In Table 1, superscripts mean different combinations of relationships. The E, U, I, S, L, and T sequentially stand for the edge, user, information, sensor, location, and time. The predicates determine which graph the edge belongs to. For instance, UE^{SLT} is the edge from the user graph in which sensor info type, time, and location relations are involved while UE^{SL} includes sensor type information and location relations. To calculate the cell values, a pessimistic approach is applied. It means a low value of each involving parameter can reduce the final value of the relationship. For instance, a zero-similarity value for time or location generates zero final value for the relationship. If we choose an optimistic approach, many unrealistic relationships will be generated. Accordingly, the multiplication of relation values specifies compound edge values. The E^{SLT} is the multiplication of location, sensor info type, and time relation values.

Then, vertices degree (D_v) and hyperedges (D_e) degree matrices are calculated by counting connected vertices and hyperedges then creating diagonal matrices by the numbers in the diameter. Finally, by a recursive function, they calculated the relatedness score of the thing k to all other things, by Equation (3), until the score difference becomes small enough in two consecutive iterations. In other words, Equation (3) provides the relatedness scores between two nodes in the hypergraph so as to provide the best corresponding sensor for every volunteer data.

$$F^{j+1} = \alpha\, D_v^{-1/2} H W D_e^{-1} H^T D_v^{-1/2} F^j + (1-\alpha)\, F^0 \tag{3}$$

While F is the relatedness score vector, W is the weight matrix of the hyperedge e, and α is the regularization parameter. H is the incidence matrix of hypergraph nodes built upon Table 2 instructions. F^0 is the initialized vector, which is obtained from the train set. The kth element of F^0 is 1 and all other values are 0. The *W* value is identical in our model since hyperedges do not have specific priorities regarding the model objective. The α value, which is a variant from 0 to 1, controls the loss function effects on relatedness scores. The loss function should be decreased during the learning process. If α is zero, the effects of all other vertices will be neglected while 1 value maximizes the adjustment of

the relatedness to all related vertices. Following the recursive equation, pairwise related-ness of vertices is derived and the simplified graph model is created. The final value of the function is a relatedness vector from the volunteer to all sensors in the network. The vector elements include ranks, which determine similarity values between the volunteer and sensors. Therefore the first rank is selected as the matched sensor for the volunteer.

2.3. Quality Assessment

VGI quality similar to the spatial data quality can be assessed by quality indicators regarding the purpose, usage, or lineage [41]. To evaluate the quality of VGI, positional accuracy, and attribute accuracy (thematic accuracy) as the quality indicators are assessed. As none of the data produced by volunteers contains the shape of features, the location normalized Euclidean difference between the sensor location and the volunteers' location is considered as positional accuracy of the point. A similar approach has been employed in [17]. In this paper, positional accuracy was calculated by the difference between sensor location and the point that the volunteers pinpoint on the map. Equation (4) describes how to assess the positional accuracy where x_v and y_v are VGI coordinates and x_s and y_s are sensor coordinates. $Positional\_Accuracy_i$ is the quality of the ith VGI data and norm and dist are normalization and distance functions. The $Positional\_Accuracy_i$ value is ra-tional since the distance between the VGI and the corresponding sensor is normalized by distances between all matched VGI and sensors. In other words, the following equation calculates the normalized distance between sensor i and the volunteered data v.

$$Positional\_Accuracy_i = \text{norm}_i\left(\sqrt{(x_v - x_s)^2 + (y_v - y_s)^2}\right) = \frac{\text{dist}(i)}{\sum_{k=1}^{n} \text{dist}(k)} \tag{4}$$

To evaluate attribute accuracy the method in [15] was used. In this paper, attribute accuracy was calculated by the difference between sensor type and the observed property reported by volunteers [17]. Therefore, for the matched VGI the Equation (5) would be considered to estimate the attribute accuracy. The s is the string of sensors measurement type, v is the string of shared VGI, and lev is the Levenshtein distance. As an example, s1 = "temperature" is the observed property of sensor1 and v1 = "temature" is the observa-tion type of volunteered data 1, which has a typo. The lev(s1,v1) would be "3" as three letters require to be dropped from s1 to be like v1.

$$\text{att. acc} = \frac{\max\big(\text{length}(s), \text{lenght}(v)\big) - \text{lev}(s, v)}{\max(\text{length}(s), \text{lenght}(v))} \tag{5}$$

3. Evaluation

To evaluate the method, an experiment is designed based on some simulated and real datasets.

3.1. Data

To implement and evaluate the proposed method, various types of sensors in an IoT architecture and VGI data with a different observed property is required. Additionally, information about location, time, and type of information producers is essential. Since volunteers can produce various types of information, our previous VGI research data in-cluded observed property [79]. This VGI data was used for evaluation in this paper as well. A summary of the research and the proposed application were provided in the Sup-plementary Documents.

In this research, volunteers shared information about water temperature, water color, and water level using predefined linguistic statements. Volunteers have gathered 1075 data records for time and geotagged locations from 7 March 2014 to 19 August 2016. An

example of the data is shown in Table 2. The study area includes the "Bakhtegan Tarasht Maharlou" basin in Fars, Iran. This region is located in Fars province and consists of both mountains and desert land cover. The region is located in Fars province includes both mountains and deserts. The area of the basin is 2.944 km², which contains 0.364 km² lakes and 740 km rivers.

Figure 6 illustrates the overall view of the study area and geospatial data layers. The geospatial data layers are described in our previous research [79]. They are produced by the Forests, Range, and Watershed Management Organization (FRW) of Iran in 2014.
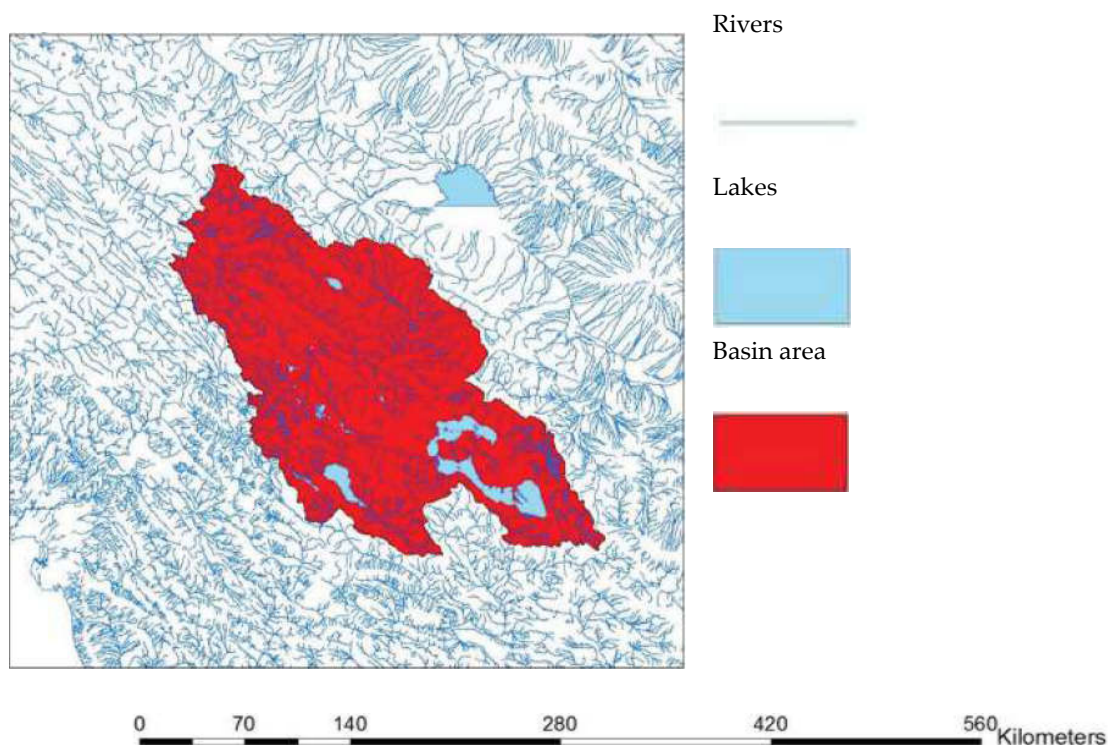

Figure 6. Bakhtegan Tashk Maharlou Basin area.

Volunteers can share their observations location by their smartphone pinpointing on the map. Then they can add attributes to the locations. To share VGI, volunteers use a mobile application (Water Resources Report) developed by authors to collect water information (water temperature, water color, and water volume) in the basin area [79]. The application is connected to the web and reports volunteers' locations and observations while there is an internet connection. When a user sends data to the server, it will be identified in terms of the data type. If the data type is recognized as a string, then a function triggers to get the qualitative corresponding values of sensor observations. The corresponding qualitative values of measurements are classified based on the predefined range of numerical values of sensor measurements. For example, if a volunteer reported "high" as the value for water level and a sensor observed 4 m of water level (which falls in the "high" category), the matching module would match the sensor and VGI observations.

Table 2. Sample volunteered geographic information (VGI) data collected in [79] research.

	Volun- teer_ID	Data_ID	Time	Date	Longitude	Latitude	Info_Type	Info_Value
1	Moh2568	Moh2568201 60716104542 5284830089	10:45:42	2016-07-16	52.848775	30.089518	color	dark
27	Sep321	Sep22452201 60506144514 5287830089	14:45:14	2016-05-06	52.878740	30.089600	level	3 m
3	Lei2542	Lei25422016 05051052145 247830189	10:52:14	2016-05-05	52.478700	30.189523	color	bright

The sensor observations were collected in the same temporal and spatial range as the VGI data. Deployed IoT sensors include the DS18B20 water temperature sensor, XL-MaxSonar-WR MB7060 water level sensor, and ASTM D445-97 viscometer. These data were previously used to obtain the quality of VGI in [80]. Table 3 illustrates the experimental results of real data using the proposed algorithm.

Table 3. The proposed method in real sensor data.

	Item	Value
1	Number of sensors	45
2	Percentage of matched data	78%
3	Average Positional accuracy	27 m
4	Average attribute accuracy	85%

However, there were several regions that no sensor observation exists while many VGI observations were recorded. Accordingly, sensor data rate throughput was low and there was a gap between VGI observations and sensors observations time. Therefore, only 65% of the data was matched with VGI data. Figure 7 illustrates the distribution of real data in comparison to VGI. Orange circles are real sensor data and purple dots are VGI data.

Figure 7. Real sensor data distribution.

Due to the low abundance of matched data, we interpolated sensor observations in time and space in the mentioned range using the simulator. To produce sensor information and model sensor network, the OMNET++ simulator was employed. The software can simulate the sensor network in real-time [81]. Additionally, it is a suitable simulator for large scales and various platforms. Last but not least, it has a C++ API to develop functions and parameters. Therefore, 35 more simulated sensors are located in the study area, which measures a combination of these observations: temperature (1.5–10 °C), viscosity (1.3–1.6 m^2/s 10^{-6}), and level of water (0.3–37 m). Sensors are connected to seven base stations with star topology (i.e., all seven sensors are connected only to one base station and the base station is the center to receive all seven sensors measurements). Sensors asynchronously push data every 1 h. The sensor coverage was set to less than 40 km. The simulated sensor observations were generated by the real sensor observations data. The period and region of data are the same as the VGI data. To test the reliability of simulation measurements, we kept 20% of real sensor observations as test data. After running the simulator for the kept real data, it revealed that the simulator generates sensor measurements with 93% accuracy.

To implement the method, the prepared sensor observations are cleaned in terms of outliers. Therefore, all real sensor observations outliers are detected and checked. Then, using the simulator, the interpolated values of sensors are generated. Sensor data includes the observation time, observation type, observation value, unit of measurement, location, and the ID of the sensor. Then VGI descriptions were analyzed and keywords were extracted using Levenstein distance. This builds info type and info value columns of VGI data mentioned in Table 2. Then using similarity matching explained in Sections 2.2.1 and 2.2.2, VGI and sensor data were matched. To do this, all VGI and sensor time values were converted to UTC format. Additionally, all coordinates were converted to the universal transverse Mercator (UTM) projection system for easier location matching. Then, related sensor attributes as the columns are added to VGI records to relate VGI and IoT records. Finally using the quality assessment method for positional and attribute accuracy in Section 2.3, the VGI quality is stored for every record.

3.2. Experimental Design

To evaluate our quality assessment method, we need to compare our results (i.e., VGI quality values from our proposed method) with other quality values calculated by the referenced approach (i.e., VGI quality which has been assessed by a reliable quality assessment method). Therefore, we have to select a reliable method to produce a reference quality map. As described in Section 1.1 we can group the existing methods for VGI quality assessment by two main categories: non-referenced and referenced ones. The first one estimates the quality of VGI using metadata such as volunteers' proficiency, history, etc. So, the low credibility led researchers to use this method in case they did not access the reference datasets [80]. Therefore, we picked up quality values that come out of the referenced method. This method assesses quantitative quality results by comparing the VGI map with a reference map (i.e., a map with high accuracy often generated by precise observations). We used the VGI quality map produced by the referenced method to compare with VGI quality map produced by our method (i.e., quality assessment using IoT). Therefore, the difference between these two maps indicates the accuracy of our method.

Figure 8 shows the evaluation method process. In the left circle, we compared the VGI map with a reference map, which had high accuracy. The results of this comparison include the attribute and positional accuracy of VGI. The normalized Euclidean distance of VGI location and corresponding observations in the reference map is considered as the positional accuracy. The Levenshtein distance of VGI data and corresponding observations in the reference map is considered as the measure of attribute accuracy. In the right circle, the VGI quality is assessed using our method. Finally, to find the accuracy of our method, the positional and attribute accuracy of the referenced approach for every data are compared with the positional and attribute accuracy of the IoT approach. The more similar values demonstrate higher accuracy of our method.
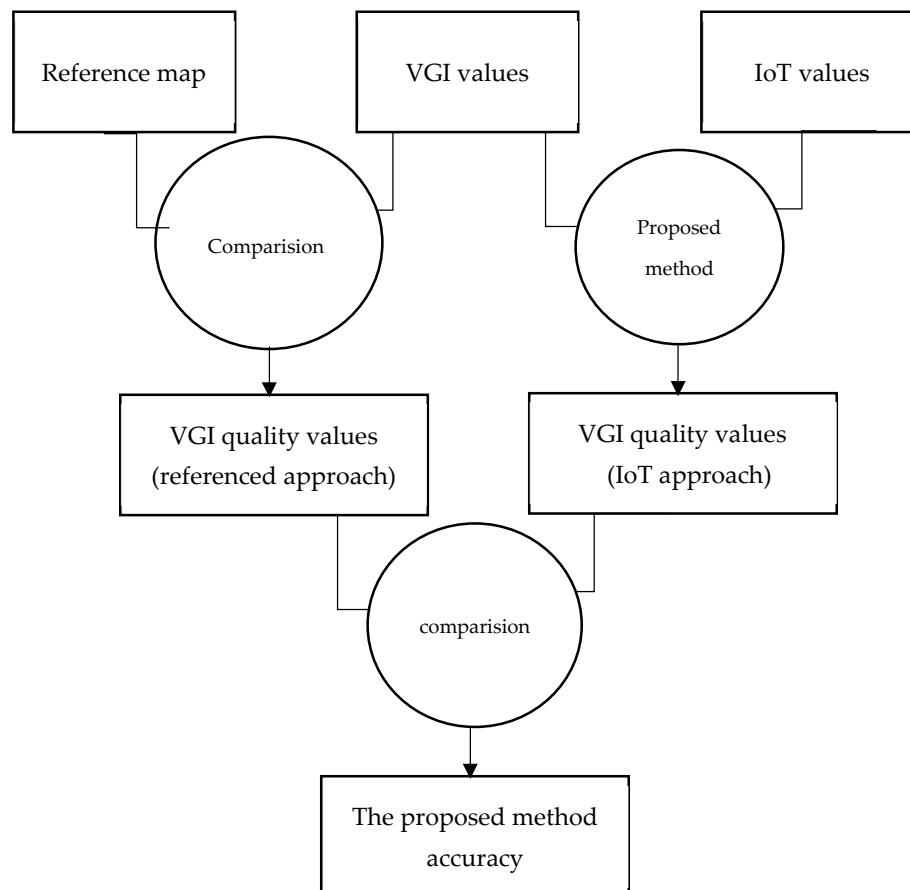
Figure 8. Evaluation flowchart.

4. Results and Discussion

Figure 9 illustrates an example of the built hypergraph model based on the VGI and sensor data. Nodes with "s" letter are sensors data and nodes with the first letter "v" are VGI data nodes. Here, the "v_Ali238" is a VGI data includes temperature report, which connects to the "s_0T2781", which is a temperature sensor, "s_VT5321", which measures temperature and viscosity, and "s_HT2781", which measures temperature and water level. "v_Hamid21" VGI data, which is a water level report, also connects to "S_HT2781".
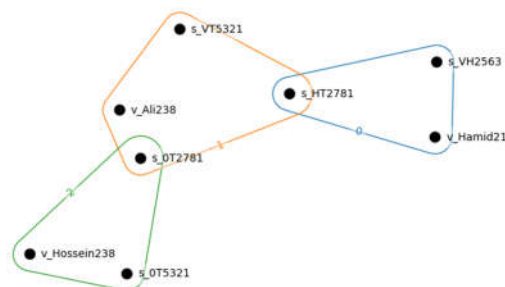

Figure 9. Hypergraph example of sensors and volunteers in the implemented method.

Regarding the evaluation approach was proposed in Figure 8, the difference between the reliable VGI quality map (i.e., the map obtained from comparing VGI values with an authoritative map) and the proposed VGI quality map (i.e., the VGI quality assessed by the proposed method) could be considered as the error of our proposed VGI quality assessment method. Figure 10 shows the difference between the quality map of the referenced approach and the proposed method quality map. The IoT_Net layer includes sensor locations. Errors are Euclidean distance between VGI corresponding locations of these two maps. Other layers are positional error in [0–10], (10–20], (20–30], (30–40], and (40–∞) meters categories. Results demonstrated that 33% of data have less than 10 m error, 43% of them have values between 10 and 20 m, 18% of errors lie in the 20–30 m range, and 16% have more than 30 m error.
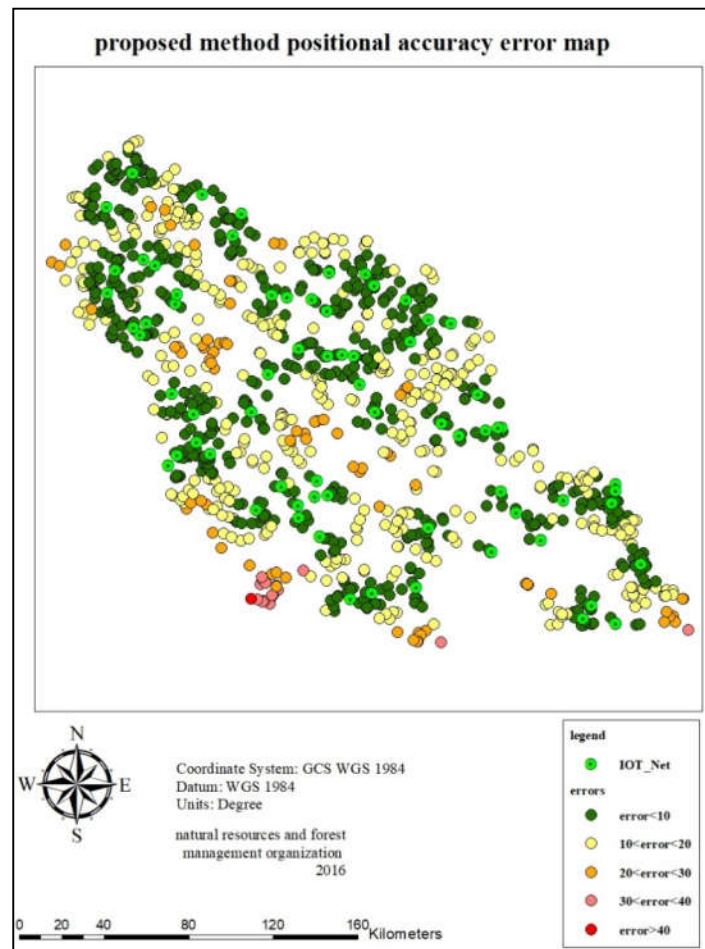
Figure 10. Positional error map of the proposed method.

Considering the map in Figure 10 revealed that there is an environmental pattern of error distribution in the area. Having observed the area, we found that there would be a relationship between area elevation and distance to the road with the most frequent errors (i.e., 10–30 m). Figure 11a,b illustrate the maps of the digital elevation model (DEM) and road accessibility map including the errors. Figure 11a shows elevation in the grayscale range from 739 m (the lowest height with black color) to 3917 m (the highest elevation with white color). Figure 11b shows a road distance map from 0 m (the closest distance with yellow color) to 120,635 m (the farthest distance with blue color). Yellow points are the errors (10–30) meters that lied in sensor proximity (10 km) and green circles are sensors.
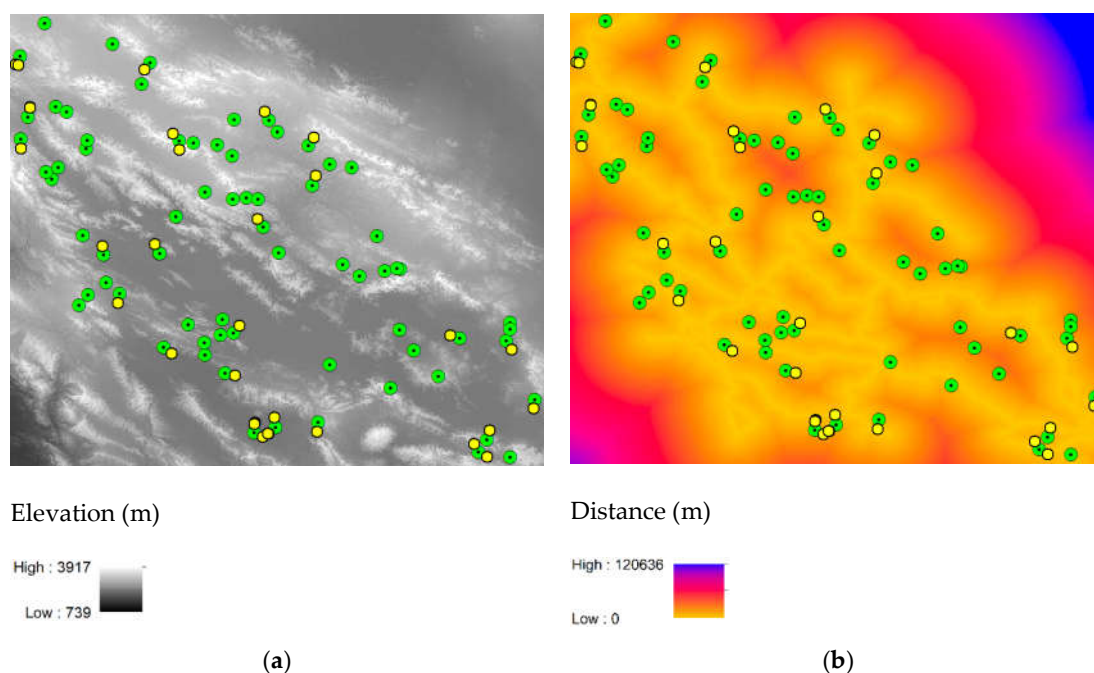
Figure 11. Final results map of error correlation and environmental effects: (**a**) digital elevation model (DEM) of the study area. (**b**) Distance from roads map.

Table 4 includes the statistics results of the analysis of the map. As the 10-30 (meters) error is the most frequent error category, we tried to analyze the effects of different elevation and road distance ranges on the abundance of this error. The first column mentions the elevation range and the second column shows the number of errors in the range. The third column illustrates the distance to roads ranges and the frequency of errors in the ranges. It indicates the highest frequency of errors where the elevation exceeds 1800 m while no meaningful relationship between the distance from roads and the error location is observed.

Table 4. Statistics of DEM and road map distance results.

Elevations (m)	Frequency of Errors (20–30 m) for DEM	Distance from Roads (km)	Frequency of Errors (20–30 m) for Distance from Roads
1400–1500	19	0–2	41
1500–1600	9	2–4	16
1600–1700	19	4–6	9
1700–1800	9	6–8	13
1800<	46	8<	21

Figure 12 illustrates the error map of our method for assessing the attribute accuracy. Further analysis showed that more than 57% of calculated attribute accuracy had less than 5% error, 35% of them have errors from 5% to 10%, 7% have errors between 10% and 15%, and 1% have errors more than 20%.
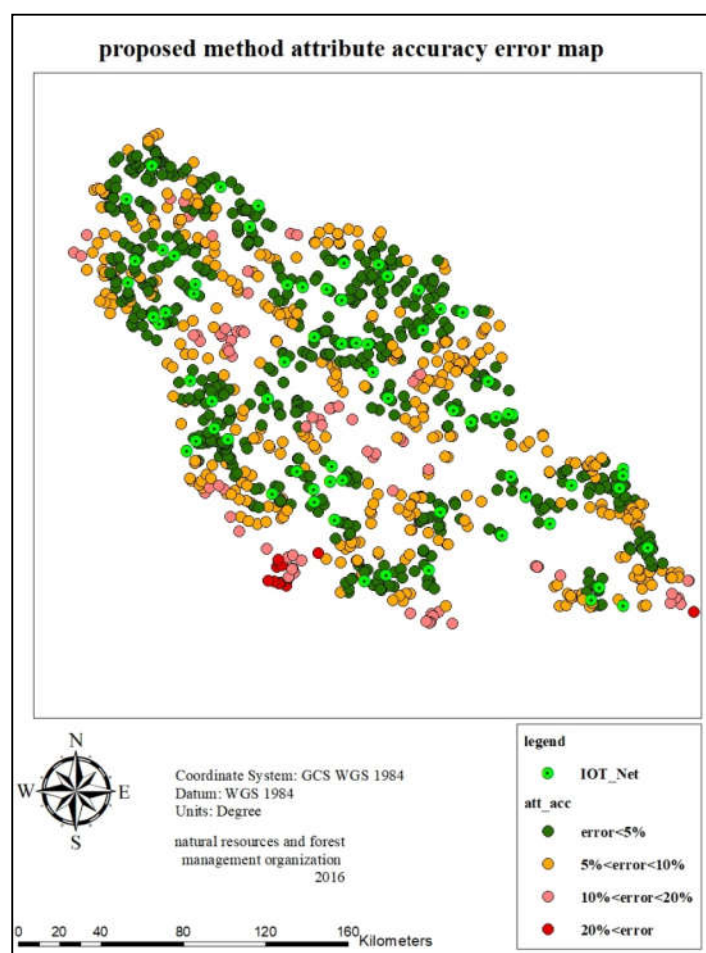
Figure 12. Attribute error map of the proposed method.

Regarding the error map in Figure 10, most of the big errors (errors > 30 m) lied in areas far from sensor locations. On the other hand, the most frequent errors, which are in the 10–20 m range, demonstrated different spatial patterns. In some cases, errors were found over sensors nearby. To explore the reason for this phenomenon, locations of the cases were pinpointed and investigated. Of these errors 88% look far from roads and close to the high elevated area. Therefore, the digital elevation model (DEM) and road distance map of the study area are generated to evaluate the correlation between them [82–84].

After considering the results in Table 4, we found that most of the errors were in areas with a steep slope, which means the high elevation difference between two close points. Therefore, when a volunteer is trying to observe and report water quality in such areas, he/she sometimes is matched with a sensor with a close planar distance with a much higher elevation. Consequently, if the water quality in the sensor location is completely different from the volunteer's report, the algorithm incorrectly reports low quality for the volunteer's report while the volunteer's report is true. For example, in Figure 13 two volunteers are sharing data with a similar planar distance to the sensor while the left volunteer's real distance to the sensor is 80 m. This causes incorrect matching and, consequently, incorrect results.
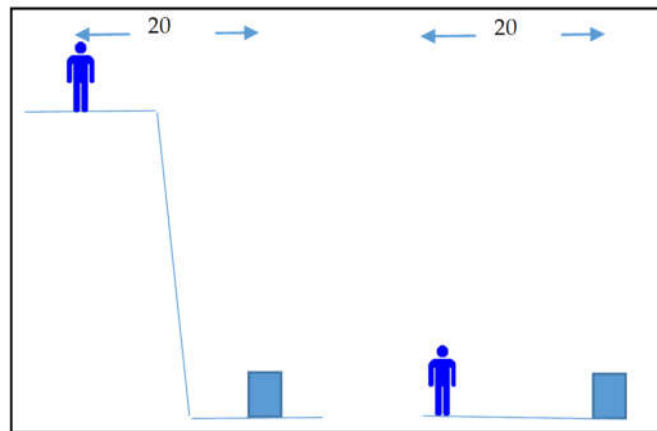
Figure 13. Effect of elevation to give wrong results for the proposed VGI quality assessment method.

Having considered attribute accuracy in Figure 12, the proposed algorithm provides 90% accuracy for assessing attribute accuracy for more than 92% of existing data. It confirms the high accuracy of the proposed method. A comparison between positional and attribute error maps, it revealed an obvious similar error distribution. To find the meaningful relationship between these two maps, we used the correlation between error values after normalizing them to an equal range. The correlation value was 87% for VGI datasets, which demonstrates the high similarity between positional and attribute accuracy errors of the proposed method.

5. Conclusions

Volunteers are free sensors who are able to collect various types of information everywhere. However, the VGI quality assessment is a serious concern when precise information is required. Moreover, the state of the art of data quality assessment methods cannot be used in real-time applications such as incidence management or real-time environmental monitoring. Real-time assessment of VGI quality has been a crucial problem as existing methods broadly rely on the intrinsic quality assessment that brings approximate qualitative values. IoT networks are widely used to collect various real-time precise quantitative information from the environment. This paper tried to employ IoT observations to assess VGI quality to keep reliable information in the network and store them in the observations database.

Conducted evaluations in this paper using the proposed method shows that 76% of the results (i.e., output VGI quality of the proposed method) had less than 20 m of positional accuracy. Moreover, 92% of the calculated attribute accuracy had more than 90% accuracy. It is concluded that the proposed method could lead IoT to assessing VGI quality in real-time. As temperature, color, and volume of water did not dramatically change in short distances, the results were reliable and might soothe the application expectations. The 87% for correlation between positional and attribute accuracy errors means robustness of the matching method as it provides similar error behavior for different matching criteria. Therefore, we rarely find low attribute accuracy error for a VGI data with high positional accuracy error. Moreover, the significant effect of elevation on the results accuracy supports the notion that considering heights for the tessellation and matching would increase the accuracy of the method. This method can be applied in any case (e.g., fire, flood, or accident report), which needs real-time VGI quality assessment using sensor observations. The other case is environment monitoring reports when it is essential to score volunteers based on their contribution quality.

Future works may consider relocating sensors based on the proposed quality assessment method to optimize VGI quality assessment using volunteers' information. Additionally, we are going to use high-quality VGI data to activate sensors in a sensor network to submit more information about the environment. For example, if a volunteer reports the smoke smell, nearby temperature sensors are activated to give more information about the incidence area. VGI filtering based on the real-time quality assessment method is helpful for the automation of sensor activation based on credible information. However, considering VGI intrinsic quality and determining volunteers' relationships may increase the proposed method accuracy and extending the matching process to the 3D space [85]. Moreover, the automation of fake report detection using the proposed method in a social network would be of interest for future works. Finally, providing a comprehensive ontology of all available observed properties of different sensors can help to deliver a more precise attribute matching in the matching step of the proposed method.

Author Contributions: Research conceptualization, Sepehr Honarparvar, Mohammad Reza Malek and Sara Saeedi; methodology and software Sepehr Honarparvar and Mohammad Reza Malek; validation, Sepehr Honarparvar, Mohammad Reza Malek and Sara Saeedi; writing—original draft preparation, Sepehr Honarparvar, Mohammad Reza Malek, Sara Saeedi, and Steve Liang; writing—review and editing, Sepehr Honarparvar, Mohammad Reza Malek, Sara Saeedi, and Steve Liang. All authors have read and agreed to the published version of the manuscript.

References

1. Goodchild, M.F. Citizens as voluntary sensors: Spatial data infrastructure in the world of Web 2.0. *Int. J. Spat. Data Infrastruct. Res.* **2007**, *2*, 24–32.
2. Turner, A. *Introduction to Neogeography*; O'Reilly Media, Inc.: Sevastopol, CA, USA, 2006.
3. Howe, J. The rise of crowdsourcing. *Wired Mag.* **2006**, *14*, 1–4.
4. Bonney, R.; Cooper, C.B.; Dickinson, J.; Kelling, S.; Phillips, T.; Rosenberg, K.V.; Shirk, J. Citizen science: A developing tool for expanding science knowledge and scientific literacy. *BioScience* **2009**, *59*, 977–984.
5. Krumm, J.; Davies, N.; Narayanaswami, C. User-generated content. *IEEE Pervasive Comput.* **2008**, *7*, 10–11.
6. Cooper, A.K.; Coetzee, S.; Kourie, D.G. Volunteered geographical information, crowdsourcing, citizen science and neogeography are not the same. In Proceedings of the International Cartographic Conference 2017, Washington, DC, USA, 3–7 July 2017; doi:10.5194/ica-proc-1-131-2018.
7. See, L.; Mooney, P.; Foody, G.; Bastin, L.; Comber, A.; Estima, J.; Fritz, S.; Kerle, N.; Jiang, B.; Laakso, M. Crowdsourcing, citizen science or volunteered geographic information? The current state of crowdsourced geographic information. *ISPRS Int. J. Geo-Inf.* **2016**, *5*, 55.
8. van Exel, M.; Dias, E.; Fruijtier, S. Proposing a redefinition of the social geographic information domain—Why perpetuating the use of 'VGI' will lead to misconceptions and information clutter. Position Papers on Virtual Globes or Virtual Geographical Reality: How Much Detail Does A Digital Earth. In Proceedings of the ASPRS/CaGIS 2010 Workshop, Orlando, FL, USA, 2011; pp. 29–36.
9. Capineri, C. *European Handbook of Crowdsourced Geographic Information*; Ubiquity Press: Lomdon, UK, 2016.
10. Haklay, M. Citizen science and volunteered geographic information: Overview and typology of participation. *Crowdsourcing Geogr. Knowl.* **2013**, 105–122, doi:10.1007/978-94-007-4587-2_7.
11. Goodchild, M.F.; Li, L. Assuring the quality of volunteered geographic information. *Spat. Stat.* **2012**, *1*, 110–120.
12. Mohammadi, N.; Malek, M. Artificial intelligence-based solution to estimate the spatial accuracy of volunteered geographic data. *J. Spat. Sci.* **2015**, *60*, 119–135.
13. See, L.; Fonte, C.C.; Antoniou, V.; Minghini, M. *Volunteered Geographic Information: Looking towards the Next 10 Years*; Springer: Berlin/Heidelberg, Germany, 2019.
14. Chrisman, N. Development in the treatment of spatial data quality. *Fundam. Spat. Data Qual.* **2006**, 21–30, doi:10.1002/9780470612156.ch1.
15. Koukoletsos, T. *A Framework for Quality Evaluation of VGI Linear Datasets*; UCL (University College London): London, UK, 2012.

16. Agumya, A.; Hunter, G.J. Responding to the consequences of uncertainty in geographical data. *Int. J. Geogr. Inf. Sci.* **2002**, *16*, 405–417.

17. Honarparvar, S.; Forouzandeh Jonaghani, R.; Alesheikh, A.A.; Atazadeh, B. Improvement of a location-aware recommender system using volunteered geographic information. *Geocarto Int.* **2019**, *34*, 1496–1513.

18. Koukoletsos, T.; Haklay, M.; Ellul, C. Assessing data completeness of VGI through an automated matching procedure for linear data. *Trans. Gis* **2012**, *16*, 477–498.

19. Zhang, H.; Malczewski, J. Quality evaluation of volunteered geographic information: The case of OpenStreetMap. *Crowdsourcing Concepts Methodol. Tools Appl.* **2019**, 1173–1201, doi:10.4018/978-1-5225-8362-2.ch058.

20. Devillers, R.; Bédard, Y.; Jeansoulin, R.; Moulin, B. Towards spatial data quality information analysis tools for experts assessing the fitness for use of spatial data. *Int. J. Geogr. Inf. Sci.* **2007**, *21*, 261–282.

21. ISO, EN. *8402: Quality Management and Quality Assurance-Vocabulary*; *London*; The International Organization for Standardization, Geneva, Switzerland: 1994.

22. Haklay, M. How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environ. Plan. B Plan. Des.* **2010**, *37*, 682–703.

23. Zielstra, D.; Zipf, A. A comparative study of proprietary geodata and volunteered geographic information for Germany. In Proceedings of the 13th AGILE International Conference on Geographic Information Science, Guimarães, Portugal, 11–14 May 2010.

24. Girres, J.F.; Touya, G. Quality assessment of the French OpenStreetMap dataset. *Trans. Gis* **2010**, *14*, 435–459.

25. Poser, K.; Dransch, D. Volunteered geographic information for disaster management with application to rapid flood damage estimation. *Geomatica* **2010**, *64*, 89–98.

26. Van Exel, M.; Dias, E.; Fruijtier, S. The impact of crowdsourcing on spatial data quality indicators. In Proceedings of the GIScience 2010 Doctoral Colloquium, Zurich, Switzerland, 14–17 September 2010; pp. 14–17.

27. Barron, C.; Neis, P.; Zipf, A. A comprehensive framework for intrinsic OpenStreetMap quality analysis. *Trans. Gis* **2014**, *18*, 877–895.

28. Senaratne, H.; Bröring, A.; Schreck, T. Using reverse viewshed analysis to assess the location correctness of visually generated VGI. *Trans. Gis* **2013**, *17*, 369–386.

29. Jilani, M.; Corcoran, P.; Bertolotto, M. Automated highway tag assessment of OpenStreetMap road networks. In Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Dallas, TX, USA, 4–7 November 2014; pp. 449–452.

30. de Albuquerque, J.P.; Fan, H.; Zipf, A. A conceptual model for quality assessment of VGI for the purpose of flood management. In Proceedings of the 19th AGILE Conference on Geographic Information Science, Helsinki, Finland, 14–17 June 2016; pp. 14–17.

31. Hung, K.-C.; Kalantari, M.; Rajabifard, A. Methods for assessing the credibility of volunteered geographic information in flood response: A case study in Brisbane, Australia. *Appl. Geogr.* **2016**, *68*, 37–47.

32. Qiu, T.; Qiao, R.; Wu, D.O. EABS: An event-aware backpressure scheduling scheme for emergency Internet of Things. *IEEE Trans. Mob. Comput.* **2017**, *17*, 72–84.

33. Stankovic, J.A. Research directions for the internet of things. *IEEE Internet Things J.* **2014**, *1*, 3–9.

34. Fang, S.; Da Xu, L.; Zhu, Y.; Ahati, J.; Pei, H.; Yan, J.; Liu, Z. An integrated system for regional environmental monitoring and management based on internet of things. *IEEE Trans. Ind. Inform.* **2014**, *10*, 1596–1605.

35. Fang, S.; Xu, L.; Zhu, Y.; Liu, Y.; Liu, Z.; Pei, H.; Yan, J.; Zhang, H. An integrated information system for snowmelt flood early-warning based on internet of things. *Inf. Syst. Front.* **2015**, *17*, 321–335.

36. Spalazzi, L.; Taccari, G.; Bernardini, A. An Internet of Things ontology for earthquake emergency evaluation and response. In Proceedings of the 2014 International Conference on Collaboration Technologies and Systems (CTS), Minneapolis, MN, USA, 19–23 May 2014; pp. 528–534.

37. Zambrano, A.M.; Perez, I.; Palau, C.; Esteve, M. Technologies of internet of things applied to an earthquake early warning system. *Future Gener. Comput. Syst.* **2017**, *75*, 206–215.

38. Kamaludin, K.H.; Ismail, W. Water quality monitoring with internet of things (IoT). In Proceedings of the 2017 IEEE Conference on Systems, Process and Control (ICSPC), Malacca, Malaysia, 15–17 December 2017; pp. 18–23.

39. Ojagh, S.; Malek, M.R.; Saeedi, S.; Liang, S. A location-based orientation-aware recommender system using IoT smart devices and Social Networks. *Future Gener. Comput. Syst.* **2020**, *108*, 97–118.

40. Zavazava, C. ITU work on Internet of Things. In Proceedings of the ICTP Workshop, Geneva, Italy, 26 March 2015.

41. Manrique, J.A.; Rueda-Rueda, J.S.; Portocarrero, J.M. Contrasting internet of things and wireless sensor network from a conceptual overview. In Proceedings of the 2016 IEEE International Conference on Internet of Things

(iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), Chengdu, China, 15–18 December 2016; pp. 252–257.

42. Morillo, P.; Orduña, J.M.; Fernández, M.; García-Pereira, I. Comparison of WSN and IoT approaches for a real-time monitoring system of meal distribution trolleys: A case study. *Future Gener. Comput. Syst.* **2018**, *87*, 242–250.

43. Senaratne, H.; Mobasheri, A.; Ali, A.L.; Capineri, C.; Haklay, M. A review of volunteered geographic information quality assessment methods. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 139–167.

44. Bakillah, M.; Liang, S.H.; Zipf, A.; Arsanjani, J.J. Semantic interoperability of sensor data with Volunteered Geographic Information: A unified model. *ISPRS Int. J. Geo-Inf.* **2013**, *2*, 766–796.

45. Gruenerbl, A.; Bahle, G.; Oehler, S.; Banzer, R.; Haring, C.; Lukowicz, P. Sensors vs. human: Comparing sensor based state monitoring with questionnaire based self-assessment in bipolar disorder patients. In Proceedings of the 2014 ACM International Symposium on Wearable Computers, Seattle, WA, USA, 13–17 September 2014; pp. 133–134.

46. Gouveia, C.; Fonseca, A. New approaches to environmental monitoring: The use of ICT to explore volunteered geographic information. *GeoJournal* **2008**, *72*, 185–197.

47. Hast, I. Quality assessment of spatial data: Positional uncertainties of the national shoreline data of Sweden. University of Galve, Galve, Sweden. 2014. Available online: https://www.semanticscholar.org/paper/Quality-Assessment-of-Spatial-Data%3A-Positional-of-Hast/bf40383e9bbd86e58cfd41be3a70fce017da29c0 (accessed on 5 March 2021)

48. Vassiliadis, P.; Simitsis, A. Near real time ETL. In *New Trends in Data Warehousing and Data Analysis*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 1–31.

49. Bakillah, M.; Liang, S.H.; Zipf, A. Toward coupling sensor data and volunteered geographic information (VGI) with agent-based transport simulation in the context of smart cities. In Proceedings of the First ACM SIGSPATIAL Workshop on Sensor Web Enablement, Redondo Beach, CA, USA, 6 November 2012; pp. 17–23.

50. Horita, F.E.; de Albuquerque, J.P.; Degrossi, L.C.; Mendiondo, E.M.; Ueyama, J. Development of a spatial decision support system for flood risk management in Brazil that combines volunteered geographic information with wireless sensor networks. *Comput. Geosci.* **2015**, *80*, 84–94.

51. Fontes, D.; Fonte, C.; Cardoso, A. Integration of VGI and sensor data in a Web GIS-based platform to support emergency response. In Proceedings of the 2017 4th Experiment@ International Conference (exp. At'17), Faro, Portugal, 6–8 June 2017; pp. 214–219.

52. Liang, S.; Huang, C.-Y.; Khalafbeigi, T. OGC SensorThings API Part 1: Sensing, Version 1.0. *Open Geospat. Consort.* **2016**, doi:10.25607/OBP-455.

53. Cuenca-Jara, J.; Terroso-Saenz, F.; Valdes-Vela, M.; Skarmeta, A.F. Classification of spatio-temporal trajectories from Volunteer Geographic Information through fuzzy rules. *Appl. Soft Comput.* **2020**, *86*, 105916.

54. Wu, O.; Gao, J.; Hu, W.; Li, B.; Zhu, M. Identifying multi-instance outliers. In Proceedings of the 2010 SIAM International Conference on Data Mining, Columbus, OH, USA, 29 April–1 May 2010; pp. 430–441.

55. Shcherbakov, M.; Brebels, A.; Shcherbakova, N.; Kamaev, V.; Gerget, O.M.; Devyatykh, D. Outlier detection and classification in sensor data streams for proactive decision support systems. In *Proceedings of Journal of Physics: Conference Series*; IOP Publishing: Tomsk, Russia, 2017; p. 012143.

56. Hawkins, D.M. *Identification of Outliers*; Springer: Berlin/Heidelberg, Germany, 1980; Volume 11.

57. Buzzi-Ferraris, G.; Manenti, F. Outlier detection in large data sets. *Comput. Chem. Eng.* **2011**, *35*, 388–390.

58. Rodrigues, L.M.; Montez, C.; Budke, G.; Vasques, F.; Portugal, P. Estimating the lifetime of wireless sensor network nodes through the use of embedded analytical battery models. *J. Sens. Actuator Netw.* **2017**, *6*, 8.

59. Bernhard, H.-P.; Springer, A.; Berger, A.; Priller, P. Life cycle of wireless sensor nodes in industrial environments. In Proceedings of the 2017 IEEE 13th International Workshop on Factory Communication Systems (WFCS), Trondheim, Norway, 31 May–2 June 2017; pp. 1–9.

60. Costa, J.J.; Maniruzzaman, M. Detection of Arsenic Contamination in Drinking Water using Color Sensor. In Proceedings of the 2018 International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE), Gazipur, Bangladesh, 22–24 November 2018; pp. 1–4.

61. Lévy, B.; Liu, Y. L p centroidal voronoi tessellation and its applications. *Acm Trans. Graph.* **2010**, *29*, 1–11.

62. Smith, R.D.; Maltrud, M.E.; Bryan, F.O.; Hecht, M.W. Numerical simulation of the North Atlantic Ocean at 1/10. *J. Phys. Oceanogr.* **2000**, *30*, 1532–1561.

63. Cacciagrano, D.; Culmone, R.; Micheletti, M.; Mostarda, L. Energy-efficient clustering for wireless sensor devices in internet of things. In *Performability in Internet of Things*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 59–80.

64. Argany, M.; Mostafavi, M.A.; Karimipour, F.; Gagné, C. A GIS based wireless sensor network coverage estimation and optimization: A Voronoi approach. In *Transactions on Computational Science XIV*, Springer: Berlin/Heidelberg, Germany, 2011; pp. 151–172.

65. Zhao, W.B.; Zhao, Z.X. Voronoi Diagram Based Retrieval Method for the Internet of Things. In *Proceedings of Advanced Materials Research*; Trans Tech Publications Ltd.: Bach, Switzerland. 2012; pp. 3420–3424.

66. Lovellette, E.; Hexmoor, H. Voronoi diagrams for automated argumentations among Internet of Things. *Multiagent Grid Syst.* **2016**, *12*, 303–318.

67. Okabe, A. *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams, with a Foreword by DG Kendall*; Wiley Series in Probability and Statistics; Wiley: Chichester, England, 2000.

68. Abdolmajidi, E.; Mansourian, A.; Will, J.; Harrie, L. Matching authority and VGI road networks using an extended node-based matching algorithm. *Geo-Spat. Inf. Sci.* **2015**, *18*, 65–80.

69. Niwattanakul, S.; Singthongchai, J.; Naenudorn, E.; Wanapu, S. Using of Jaccard coefficient for keywords similarity. In Proceedings of the International Multiconference of Engineers and Computer Scientists, Hong Kong, China, 21–23 March 2007; pp. 380–384.

70. Novack, T.; Peters, R.; Zipf, A. Graph-based strategies for matching points-of-interests from different VGI sources. In Proceedings of the 20th AGILE Conference, Wageningen, The Netherlands, 9–12 May 2017; pp. 9–12.

71. Chaidee, S. Interactive land-use optimization using laguerre voronoi diagram with dynamic generating point allocation. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2017**, *42, 1091–1098.*

72. Schmid, S.; Wattenhofer, R.; Boukerche, A. Modeling Sensor Networks. *Algorithms Protoc. Wirel. Sens. Netw.* **2008**, *62*, 77.

73. Li, M.M.; Peters, C. Reconceptualizing service systems–Introducing service system graphs. In Proceedings of the Thirty Ninth International Conference on Information Systems, San Francisco, CA, USA, 13–16 December 2018.

74. Hossmann, T.; Spyropoulos, T.; Legendre, F. Putting contacts into context: Mobility modeling beyond inter-contact times. In Proceedings of the Twelfth ACM International Symposium on Mobile Ad Hoc Networking and Computing, Paris, France, 16–19 May 2011; pp. 1–11.

75. Yao, L.; Sheng, Q.Z.; Ngu, A.H.; Li, X. Things of interest recommendation by leveraging heterogeneous relations in the internet of things. *ACM Trans. Internet Technol.* **2016**, *16*, 1–25.

76. Jung, J.; Chun, S.; Lee, K.-H. Hypergraph-based overlay network model for the Internet of Things. In Proceedings of the 2015 IEEE 2nd World Forum on Internet of Things (WF-IoT), Milan, Italy, 14–16 December 2015; pp. 104–109.

77. Zhou, D.; Huang, J.; Schölkopf, B. Learning with hypergraphs: Clustering, classification, and embedding. *Adv. Neural Inf. Process. Syst.* **2006**, *19*, 1601–1608.

78. Agarwal, S.; Branson, K.; Belongie, S. Higher order learning with graphs. In Proceedings of the 23rd International Conference on Machine Learning, Pittsburgh, PA, USA, 25–29 June 2006; pp. 17–24.

79. Leordeanu, M.; Zanfir, A.; Sminchisescu, C. Semi-supervised learning and optimization for hypergraph matching. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 7 November 2011; pp. 2274–2281.

80. Honarparvar, S.; Malek, M.R. Updating information on water resources drought using volunteered geographic information. *Sci. Res. Q. Geogr. Data* **2019**, *28*, 123–135.

81. Zhu, X.J. *Semi-Supervised Learning LIterature Survey*; University of Wisconsin-Madison, Department of Computer Sciences: Madison, WI, USA, 2005

82. Varga, A.; Hornig, R. An overview of the OMNeT++ simulation environment. In Proceedings of the 1st International Conference on Simulation Tools and Techniques for Communications, Networks and Systems & Workshops, Marseille, France, 3–7 March 2008; pp. 1–10.

83. Castro, D.; Jofo, P.; Dos, S.; Alexander, Z. A taxonomy of quality assessment methods for volunteered and crowdsourced geographic information. *Trans. Gis* **2018**, *22*, 542–560.

84. Saeedi S., Integrating macro and micro scale approaches in the agent-based modeling of residential dynamics, *Int J Appl Earth Obs Geoinf* **2018**, pp. 214-229, 10.1016/j.jag.2018.02.012

85. Saeedi S., Liang S., Graham D., Lokuta M.F., Mostafavi M., Overview of the OGC CDB Standard for 3D Synthetic Environment *Modeling* and Simulation, Computer Science. *ISPRS Int. J. Geo Inf.* **2017**, *6*, 306, doi:10.3390/ijgi6100306