

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/349822798>

Ranking of VGI contributor reputation using an evaluation-based weighted pagerank

Article in *Transactions in GIS* · March 2021

DOI: 10.1111/tgis.12735

CITATIONS

0

READS

87

4 authors:



Die Zhang

Chinese Academy of Sciences

2 PUBLICATIONS 21 CITATIONS

[SEE PROFILE](#)



Yong Ge

State Key Laboratory of Resources and Environmental Information System

146 PUBLICATIONS 2,356 CITATIONS

[SEE PROFILE](#)



Alfred Stein

University of Twente

438 PUBLICATIONS 10,896 CITATIONS

[SEE PROFILE](#)



Wenbin Zhang

Institute of Geographic Sciences and Natural Resources Research

12 PUBLICATIONS 17 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Spatial poverty modeling [View project](#)



Quality aspects of distributed scatterers in polarimetric SAR interferometry [View project](#)

Ranking of VGI contributor reputation using an evaluation-based weighted pagerank

Die Zhang^{1,2}  | Yong Ge^{1,2} | Alfred Stein³ | Wen-Bin Zhang^{1,2}

¹State Key Laboratory of Resources & Environmental Information System, Institute of Geographic Sciences & Natural Resources Research, Chinese Academy of Sciences, Beijing, China

²University of Chinese Academy of Sciences, Beijing, China

³Faculty of Geoinformation Science and Earth Observation (ITC), University of Twente, Enschede, The Netherlands

Correspondence

Yong Ge, State Key Laboratory of Resources & Environmental Information System, Institute of Geographic Sciences & Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China.
Email: gey@lreis.ac.cn

Funding information

Key Programs of the National Science Foundation of China, Grant/Award Number: 41531179; National Key Research and Development Program, Grant/Award Number: 2017YFB0503501; National Science Foundation of China, Grant/Award Number: 41421001; National Science Foundation for Distinguished Young Scholars of China, Grant/Award Number: 41725006

Abstract

The quality of volunteered geographic information (VGI) is questionable as it emerges through a diversity of contributors. The reputation of a contributor is increasingly applied to VGI quality assessment and its assurance. Research on how to measure and validate reputation, however, is still required. This study proposes an evaluation-based weighted PageRank (EWPR) algorithm to provide a ranking metric of reputation. Ranking is established on the basis of the assumptions that: (a) there is an evaluation relationship between VGI contributors; (b) the reputation of a contributor is movable within the VGI community; and (c) highly active contributors are more likely to have high reputation. By means of case studies using OpenStreetMap, two existing methods are compared with the EWPR algorithm. The results show that the algorithms used for web pages and social network participants are applicable to understand the reputation of VGI contributors. This study also applies credibility, based on reputation, as a popular indicator of VGI quality. The high correlation between credibility and VGI characteristics indicates that reputation is useful for dealing with quality variability in VGI.

1 | INTRODUCTION

Volunteered geographic information (VGI) provides platforms with knowledge sharing and collaborative work. Data sources are provided by the contributing community, which is composed of individuals where everyone is welcome to participate in a VGI project. However, this raises concerns that VGI quality cannot always be guaranteed because there are no standards and specifications for the contribution process of data providers.

OpenStreetMap (OSM) is one of the most popular VGI communities. A frequently applied method of assessing OSM data quality is a comparison against reference datasets (Fan, Zipf, Fu, & Neis, 2014; Haklay, 2010; Jackson et al., 2013; Zhang, Leung, & Ma, 2019). Such a data comparison method, however, is restricted by the inaccessibility of reference datasets of higher quality (Antoniou & Skopeliti, 2015). Flanagan and Metzger (2008) state that indicators constructed by means of collective or community efforts, such as ratings and reputation systems or social networking applications, may in many cases serve as assessments of VGI quality. As a consequence, studies on the intrinsic indicators of VGI quality have increased in recent years, wherein reputation is widely used as a factor in determining these intrinsic indicators. For instance, Zhou and Zhao (2016) argued that data trustworthiness was related to the reputation of the current contributor and to the reputation of the previous contributor. The average reputation has been used as the expertise of OSM data (Muttaqien, Ostermann, & Lemmens 2018). Additionally, Severinsen, de Roiste, Reitsma, and Hartato (2019) measured the trustworthiness of OSM data using a linear weighting of reputation and spatial and temporal factors. These studies showed that reputation is helpful in VGI quality assessment.

Reputation is also beneficial when it comes to quality assurance. Goodchild and Li (2012) proposed a social approach of assuring VGI quality, which refers to the ability of highly reliable contributors as moderators within a VGI community. In this way, OSM is able to assign high-reputation contributors as gatekeepers and develop a bunch of administrative records in the future. Hence, the reputation influences the produced data and can be used as a basis for VGI quality assessment and its assurance (Degrossi, de Albuquerque, Rocha, & Zipf, 2018). Before this can happen, though, effective methods of measuring reputation are necessary.

Different formulations of reputation have been associated with diverse perspectives. Lodigiani and Melchiori (2016) modeled contributors as nodes whose reputations have been assessed by a personalization vector according to the contributors' actions. Fogliaroni, D'Antonio, and Clementini (2018) suggested that the reputation of a contributor was the average of the trustworthiness of all the contributed data, while Muttaqien et al. (2018) utilized one characteristics-based method involving mapping days, longevity, number of edits, software skill, and local knowledge. Other characteristics, such as their preferred subjects and years in school, were used to determine the contributors' reputation (Severinsen et al., 2019). Nevertheless, the relationships among VGI contributors are rarely investigated for reputation measurement. In practice, reputation can be built on the history of past interactions between members within a single community (Maué, 2007). Reputation also indicates agreement of a community on the perception of a person's reliability (Bishr & Janowicz, 2010) and reflects the collective trust vested in a person by the community (Bishr & Kuhn, 2013). Accordingly, the relationships among the members within a VGI community are needed for reputation measurement.

VGI contributors have established large communities that allow millions of people to participate in and edit the existing data. Therefore, there are vast amounts of implicit information available about their relationships because of successive interactions and collaborations between contributors. Several studies have used graph/network theory to analyze the relationships among contributors (Mooney & Corcoran, 2014; Stein, Kremer, & Schlieder, 2015; Truong, de Runz, & Touya, 2018). These studies have been devoted to identifying the optimal social networks that describe the various forms of interactions and collaborations among VGI contributors. However, they have a limited ability in providing a reputation-related metric based on these interactions. Moreover, the interactions among VGI contributors indirectly reflect the evaluation relationships (i.e., peer feedback) among the contributors. Particularly, Zhou and Zhao (2016) concluded that there is an evaluation relationship between the subsequent contributors and the previous contributors who yielded different versions of the same OSM object. Such a relationship provides useful information for reflecting on reputation. For example, online shopping sites such as eBay, Amazon, and Newegg have user reputation systems based on relevant feedback and ratings (Adler, de Alfaro, Kulshreshtha, & Pyeand, 2011), and various types of peer feedback were aggregated to obtain reputation of OSM contributors (Gusmini et al., 2017).

Network-based technology has been used effectively to capture a range of relationships among VGI contributors, while rarely to address evaluation relationships. It has also rarely been investigated how to incorporate the characteristics of contributors and their relationships as a reputation measurement. To fill these gaps, this study

first analyzes contributors' past interactions and extracts contributor-to-contributor (C2C) evaluation relationships between VGI contributors. A contributor network is subsequently built to illustrate the C2C evaluation. We find that the links among contributors are similar to the links among website pages. Thus, we modify a link-based ranking algorithm—the weighted PageRank (WPR)—into an evaluation-based weighted PageRank (EWPR) algorithm. WPR distributes the rank score of a page based on the popularity (i.e., weights in WPR) of the page and the quality of the links to it. In EWPR, the weights reflect the activity levels of VGI contributors, and the links characterize the C2C evaluation. Similarly, EWPR ranks the reputation of a contributor based on the activity level and the C2C evaluations of the links to the contributor. Based on this, the reputation ranking of VGI contributors is obtained using an iterative calculation. To verify the effectiveness of the EWPR algorithm, EWPR-based reputation has been obtained using OSM datasets for the cities of Hong Kong and Beijing, and two published methods are implemented for comparison. The first method is proposed by Muttaqien et al. (2018) (MOL), using the characteristics of VGI contributors, and the second method is by Zhou and Zhao (2016) (ZZ), who also used the evaluation relationships by means of a similarity-based method. The credibility of OSM data was further explored using these two methods, which helps us to understand the usefulness of reputation in VGI quality assessment and indirectly validates the present ranking of reputation.

The remainder of the article is as follows. Section 2 introduces the methodology for the EWPR algorithm to identify reputation based on C2C evaluation and activity level. Section 3 describes the case studies through the proposed method and two published methods for comparison using OSM datasets. Section 4 discusses the advantages, limitations, and future research. Finally, Section 5 concludes.

2 | METHODOLOGY

2.1 | C2C evaluation within the OSM community

This section describes a C2C evaluation that characterizes the evaluation relationships among contributors in the OSM community. OSM uses a data structure with four core objects: nodes, ways, relations, and tags. Here, two kinds of objects represent geographic features: nodes, which refer to points with geographic locations; and ways, which are ordered lists of nodes that represent polylines or possibly polygons if they form a closed loop. Relations and tags represent the attributes attached to the geographic features.

Let there be N objects, $\{O_i\}$ ($1 \leq i \leq N$), and M contributors, $\{u_m\}$ ($1 \leq m \leq M$), at a specific time. The object O_i has the following characteristics:

- **Versions.** $\{v_{ij}\}$ ($j \geq 1$), where v_{ij} is the j th version of the object O_i .
- **Object contributors.** $\{u_{ij}\}$ ($u_{ij} \in \{u_m\}, j \geq 1$), where u_{ij} is the contributor that generates version v_{ij} .
- **Adjacent contributors.** If the object O_i has more than one version, there are adjacent contributors, denoted $\{u_{ij}, u_{ij+1}\}$; u_{ij} and u_{ij+1} that generate versions v_{ij} and v_{ij+1} , respectively. It should be noted that u_{ij} and u_{ij+1} can be the same user.

An evaluation relationship between contributors can be reflected by the editing process; that is, a contributor implicitly produces an evaluation of the previous version. For instance, if u_{ij} generated version v_{ij} (e.g., one polyline) for object O_i , as shown in Figure 1, and then u_{ij+1} edited the version v_{ij} by adding more nodes (as shown in orange), it means that u_{ij+1} evaluated the quality of v_{ij} and indirectly evaluated u_{ij} 's reputation. Similarly, u_{ij+2} further modified v_{ij+1} to get version v_{ij+2} and, therefore, also evaluated u_{ij+1} 's reputation.

Intuitively, v_{ij+1} was more similar to v_{ij+2} than to v_{ij} . That is, the change of O_i made by u_{ij+2} was small compared with that by u_{ij+1} . This implies that u_{ij+2} mostly agrees with the quality of v_{ij+1} and, thus, implicitly confirmed the reputation of u_{ij+1} . However, most of u_{ij} 's edits were modified by u_{ij+1} (i.e., a large change from v_{ij} to v_{ij+1}), and u_{ij} 's

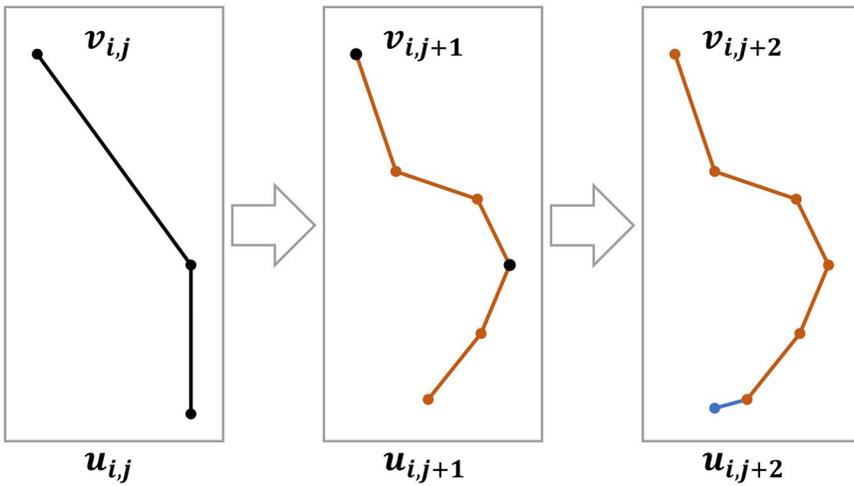


FIGURE 1 Example of the evaluation relationship between adjacent contributors

reputation did not receive good feedback from $u_{i,j+1}$. Hence, the evaluation depends upon the changes made from one OSM version to the next, where, for example, a minor change from $v_{i,j+1}$ to $v_{i,j+2}$ corresponds with a highly rated reputation for $u_{i,j+1}$ as evaluated by $u_{i,j+2}$. This kind of evaluation may exist between any two contributors within the OSM community, and for the remainder of the current study will be called C2C evaluation.

2.2 | Establishment of a contributor network

For a better illustration of the C2C evaluation, a directed and weighted contributor network $G = (U, L, W)$ is established, where $U = \{u_m\}_{m=1, \dots, M}$ is the set of nodes (contributors), $L \subseteq U^2$ is the set of directed links $l_{p,q} = (u_p, u_q)$ (the evaluation relationship between contributors, from u_q to u_p), and $W = \{w_{p,q}\}$ is the set of evaluations $w_{p,q}$ attached to each link, denoting the evaluation of u_p by u_q .

Traversing historical OSM data, all contributors are depicted as nodes, and the directed links are generated between adjacent contributors. Suppose that there are $N = 3$ objects, $\{O_i\}_{i=1, \dots, 3}$, and $M = 6$ contributors, then $\{u_m\}_{m=1, \dots, 6}$. The relevant contributions and generated contributor network are illustrated in Figure 2. Adjacent contributors are linked according to the order of their contributions, which here is shown by the links of u_4 to u_4 , u_6 to u_4 , u_2 to u_6 , and u_1 to u_2 . If two adjacent contributors appear many times for different objects, then only one link is generated between them (e.g., u_2 to u_6). When the object has only one version, it is isolated in the network (e.g., u_5 for O_2). For convenience, the links pointing to one node are named the inlinks of the node, and the associated nodes on the inlinks are called the inlink nodes of the node. For example, $l_{2,1}$ is the inlink of u_2 , and u_1 is the inlink node of u_2 . In turn, regarding the inlink nodes, these links are called outlinks, and the node is the outlink node of these inlink nodes (e.g., $l_{2,1}$ is the outlink of u_1 , and u_2 is the outlink node of u_1).

The evaluation $w_{p,q}$ is the extent to which contributor u_q agrees with contributor u_p , which is reflected in the extent of changes made between the corresponding OSM versions. The current study uses the similarity between two versions to describe the extent of the changes. In addition, ZZ pointed out that a contributor can operate an object in OSM when one of two cases applies: there is a change in the state of the real-world entity represented by the object (i.e., whether the entity exists or not); or a modification of the object without a change in the state of the geographic entity. When the contributors edit the object because of a change of state, there is virtually no corresponding C2C evaluation for the previous contributor. Thus, we use a time threshold T and a similarity threshold V to identify the circumstance. If the time interval between two adjacent versions extends T and the

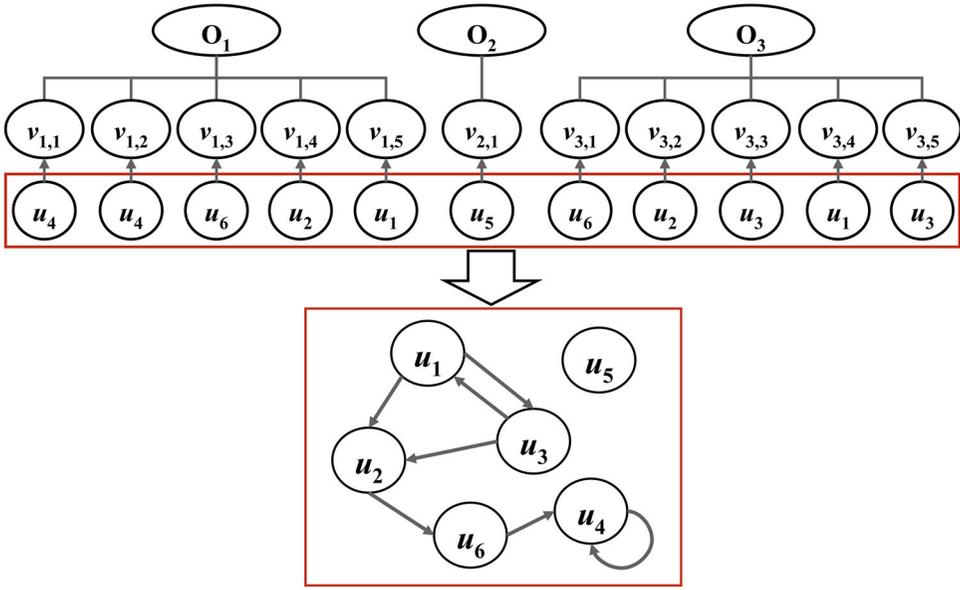


FIGURE 2 Example of how complete past edits are turned into a contributor network

version similarity is smaller than V , the change is recognized as a change of state, and the corresponding adjacent versions are not considered in the C2C evaluation.

In fact, the adjacent versions between two contributors may exist more than once. Thus, the evaluation $w_{p,q}$ should be rewritten as:

$$w_{p,q} = \sum_{(v_{ij}, v_{ij+1}) \in Q_{p,q}^{T,V}} VerSim(v_{ij}, v_{ij+1}) / N_{Q_{p,q}} \tag{1}$$

where $w_{p,q}$ denotes the evaluation of u_p by u_q , $Q_{p,q}^{T,V}$ is the set of adjacent version pairs from u_q to u_p , meeting the time threshold T and the similarity threshold V (e.g., $Q_{6,2} = \{(v_{1,3}, v_{1,4}), (v_{3,1}, v_{3,2})\}$), and $N_{Q_{p,q}}$ is the number of version pairs (e.g., $N_{Q_{6,2}} = 2$). $VerSim(v_{ij}, v_{ij+1})$ represents the similarity between adjacent versions generated by u_p and u_q . $VerSim(v_{ij}, v_{ij+1})$ may be computed in terms of spatial and thematic similarity of OSM versions.

Similarity measurement of geographical features is another important research topic in GIS. Many scholars have studied similarity computation methods for polylines and polygons. The current study does not attempt to suggest a new method of similarity measurement. Instead, the simple characteristics of OSM objects are initially selected to compute the version similarity to implement a complete methodology. The following describes the spatial similarity and thematic similarity for a point, polyline, and polygon in OSM.

The spatial similarity of point features is computed from the overlapping rate of the buffer ($SimBuff$). Length similarity ($SimLen$), the overlapping rate of the buffer, and the distance between turning functions ($SimShape$) are used for the spatial similarity of polylines. Likewise, spatial similarity is measured from the area similarity ($SimArea$), the overlapping rate of the buffer, and the distance between the turning functions for polygons.

$$SimBuff = \frac{Intersect(Buff(v_{ij}), Buff(v_{ij+1}))}{\max(Buff(v_{ij}), Buff(v_{ij+1}))} \tag{2}$$

$$SimLen = 1 - \frac{|Len(v_{ij}) - Len(v_{ij+1})|}{\max(Len(v_{ij}), Len(v_{ij+1}))} \tag{3}$$

$$SimShape = 1 - \frac{|TF(v_{ij}) - TF(v_{ij+1})|}{2\pi} \quad (4)$$

$$SimArea = 1 - \frac{|Area(v_{ij}) - Area(v_{ij+1})|}{\max(Area(v_{ij}), Area(v_{ij+1}))} \quad (5)$$

where $Buff(v_{ij})$ denotes the area of the buffer of version v_{ij} and $Intersect(Buff(v_{ij}), Buff(v_{ij+1}))$ denotes the overlapping area of two buffers. $Len(v_{ij})$ and $Area(v_{ij})$ are, respectively, the length of a polyline and the area of a polygon. $TF(v_{ij})$ is the turning function introduced by Arkin, Chew, Huttenlocher, Kedem, and Mitchell (1991). Fan et al. (2014) also used it to measure the shape similarity of two polygons.

The thematic similarity of a point, polyline, and polygon is computed from the tag similarity ($SimTag$):

$$SimTag = \frac{Intersect(Tag(v_{ij}), Tag(v_{ij+1}))}{\max(Tag(v_{ij}), Tag(v_{ij+1}))} \quad (6)$$

where $Intersect(Tag(v_{ij}), Tag(v_{ij+1}))$ is the number of intersections of two tag sets. For example, "railway = rail" is the tag for one version, and "railway = rail" and "layer = 1" are the tags for the next version. Their intersection is "railway = rail" and tag similarity $1/2 = 0.5$.

These metrics of spatial and thematic similarities can be averaged as the similarity of the adjacent versions, that is, $VerSim(v_{ij}, v_{ij+1})$ in Equation (1).

2.3 | Ranking metric of contributor reputation

The current study applied a modified WPR algorithm to rank contributor reputation. The WPR (Xing & Ghorbani, 2004) is a page-ranking algorithm to measure the importance of website pages. It distributes the rank scores of the pages based on the popularity of the pages and the quality of their inlink pages. Different from the classical PageRank algorithm, each page also receives a weight proportional to its popularity (its number of inlinks and outlinks). The formula for the WPR is as follows:

$$PR(A_p) = \alpha \sum_{A_q \in M_{A_p}} PR(A_q) \cdot W_{pq}^{in} \cdot W_{pq}^{out} + \frac{1 - \alpha}{H} \quad (7)$$

where $PR(A_p)$ is the rank score of page A_p ; α is the damping factor that is generally set at 0.85 (Brin & Page, 1998); H is the total number of the pages; M_{A_p} is the set of A_p 's inlink pages; and $W_{pq}^{in} \cdot W_{pq}^{out}$ represents the popularity from link $l_{p,q}$.

The popularity from the number of inlinks and outlinks is calculated as follows:

$$W_{pq}^{in} = \frac{I(A_p)}{\sum_{A_t \in R_{A_q}} I(A_t)} \quad (8)$$

$$W_{pq}^{out} = \frac{O(A_p)}{\sum_{A_t \in R_{A_q}} O(A_t)} \quad (9)$$

where $I(A_p)$ and $I(A_t)$ represent the number of inlinks of A_p and A_t , respectively; R_{A_q} denotes the set of A_q 's outlink pages; and $O(A_p)$ and $O(A_t)$ represent the number of outlinks of A_p and A_t , respectively.

Based on Equation (7), the WPR was modified into the EWPR for the reputation ranking of VGI contributors:

$$R(u_p) = \alpha \sum_{u_q \in M_{u_p}} w_{p,q} \cdot R(u_q) \cdot W_{pq}^{in} \cdot W_{pq}^{out} + \frac{1-\alpha}{H} \quad (10)$$

where $R(u_p)$ is the EWPR score of contributor u_p and $w_{p,q} \in [0, 1]$ is the C2C evaluation of u_p by u_q . The other parameters have the same meanings as Equations (7)–(9).

The WPR is used to give each page a relative score of importance by evaluating the popularity and quality of its links. That is, the rank score of a page, $PR(A_p)$, is positively correlated with the popularity ($W_{pq}^{in} \cdot W_{pq}^{out}$) and the rank scores of their inlink pages, $PR(A_q)$. Highly active contributors within the OSM community are supposed to generate more edits (outlinks), and the edits are more likely to be modified by other contributors (inlinks) over time. Thus, $W_{pq}^{in} \cdot W_{pq}^{out}$ in the contributor network can represent the activity level of a contributor within the community. As in Equation (7), the contributor u_p will receive a higher EWPR score (i.e., $R(u_p)$) when the activity level is higher. Moreover, $R(u_p)$ is positively correlated with the weighted EWPR scores of the inlink pages, meaning $w_{p,q} \cdot R(u_q)$. This means that a contributor who receives higher C2C evaluations from the contributors with higher EWPR scores is assumed to have a higher EWPR score.

The EWPR can also address special contributors based on some assumptions. When calculating the EWPR, the pages with no outlinks are assumed to link out to all other pages. Therefore, their rank scores are divided evenly among all other pages. As an example, in Figure 2, u_5 , which never modified the edits in OSM, is seen as the inlink page of all pages. This indicates that u_5 may at any time randomly modify an OSM object. Additionally, PageRank uses a damping factor, namely α in Equation (10), to deal with the ranks for self-link nodes (e.g., u_4) and interlink nodes (e.g., u_1 and u_3). u_4 's reputation depends not only on itself, but also on the average reputation of other contributors. Essentially, there is a certain probability that any contributor will modify u_4 's edits.

The above describes the proposed ranking metric for reputation. To validate the metric, two existing methods were selected to compare with our method. In the works of MOL, mapping days, longevity, number of edits, software skill, and local knowledge were combined in a weighted sum as the integrated expertise score of VGI contributors. These are described as characteristics-based reputation in the current study. In addition, ZZ calculated reputation using the average of the evaluations from the contributors that co-edited the same OSM objects. These evaluations were also calculated by the similarity degree among multiple versions. We describe their results as similarity-based reputation.

MOL and ZZ also proposed methods for calculating the indicators of VGI quality. The former employed the weighted sum of characteristics-based reputation of corresponding contributors as the expertise of OSM data. The latter suggested the trustworthiness of the current version (e.g., v_{ij}) for an OSM object through the composition of the trustworthiness of the previous version (v_{ij-1}) and the reputation of the current contributor (u_{ij}). That is, the trustworthiness of the modification proportion of v_{ij-1} to v_{ij} was equal to u_{ij} 's reputation, and the trustworthiness of the remainder proportion was the max value between the trustworthiness of v_{ij-1} and u_{ij} 's reputation. Using the above methods, the expertise and trustworthiness of OSM data can be obtained based on EWPR-based reputation. The credibility of OSM data can be understood in general as a perceived quality that comprises trustworthiness and expertise (Hovland, Janis, & Kelley, 1953). Thus, the average of expertise and trustworthiness was used as the credibility of OSM data.

3 | CASE STUDIES

To show the general principles, feasibility, and transferability of our methodology, three different methods were implemented using real OSM datasets for Hong Kong (HK) and Beijing (BJ), China. Section 3.2 details the reputation ranking from the proposed EWPR (denoted as the R-level) and explores data credibility in respect to the R-level. The comparisons among the R-level, characteristics-based reputation (E-level), and similarity-based reputation (S-level) are examined in Section 3.3.

3.1 | Study areas and data

Historical data for OSM are available on the Geofabrik server (download.geofabrik.de/), which is the provider of region-wise extracts of OSM data with up-to-date worldwide coverage. The historical OSM data for China were downloaded and last modified on Geofabrik on May 19, 2020. The original data containing the full OSM history for China were downloaded in PBF format, and experimental data in HK and BJ were extracted from it using data processing.

The contributors active in an area may also contribute data in the vicinity of the area they are active in, and may even contribute data globally. That is, the contributor network should be established based on the past edits of OSM data on a global scale. Considering the large amount of data, the case defined a bounding box of 112.5°E, 21.5°N to 115.5°E, 24°N (WGS 84) for extracting data that contained every version of all objects in HK and a bounding box of 114°E, 38°N to 119°E, 43°N to extract the data covering past edits in BJ. After that, we projected these data onto the Krasovsky_1940_Albers coordinate system (see Figure 3). Using Python scripts, the experimental data provided a series of tables for the objects of the nodes and ways. Each object had the attributes of OSM ID, user ID, username, timestamp, version, changeset, and visible. In addition, the characteristics of geographical features, including point, polyline, and polygon, were extracted.

3.2 | EWPR-based reputation

3.2.1 | C2C evaluation and activity level

To obtain the C2C evaluation in Equation (1), we initially set the time threshold T to 365 days (used by ZZ) and the similarity threshold V to 0.2, here using the 80/20 rule (Box & Meyer, 1986). Figure 4 shows the adjacent versions of a point, polyline, and polygon feature in HK, and their version similarities are calculated in Table 1. As shown in Figure 4a, spatial location and the attributes of the point were modified from version 5 to version 6. The corresponding spatial and thematic similarity were 0.80 and 0.78, respectively, and they were averaged into the version similarity equal to 0.79. The polyline in Figure 4b and the polygon in Figure 4c only changed in space, not in attributes; thus, their thematic similarities were equal to 1.00. Meanwhile, the spatial change of the polyline

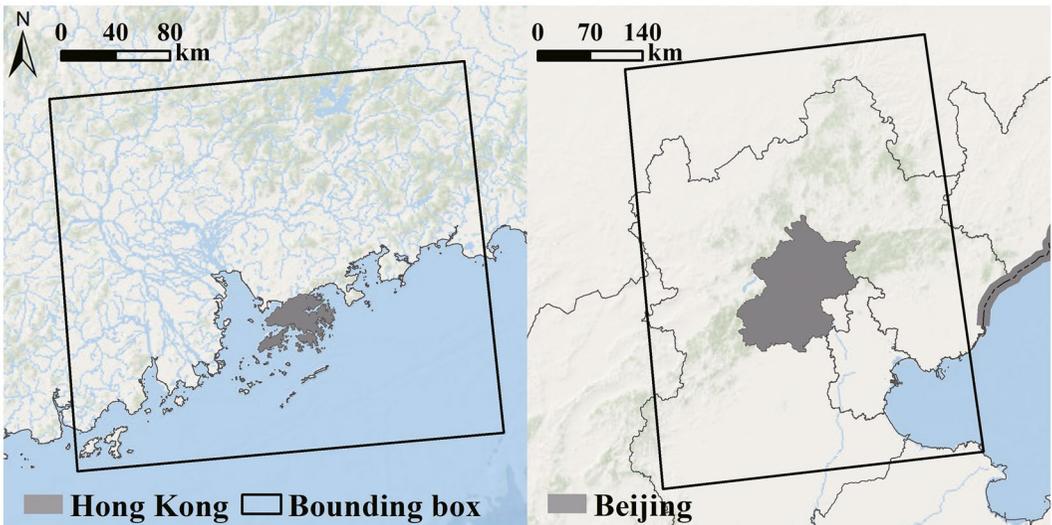


FIGURE 3 Study areas and bounding boxes for China

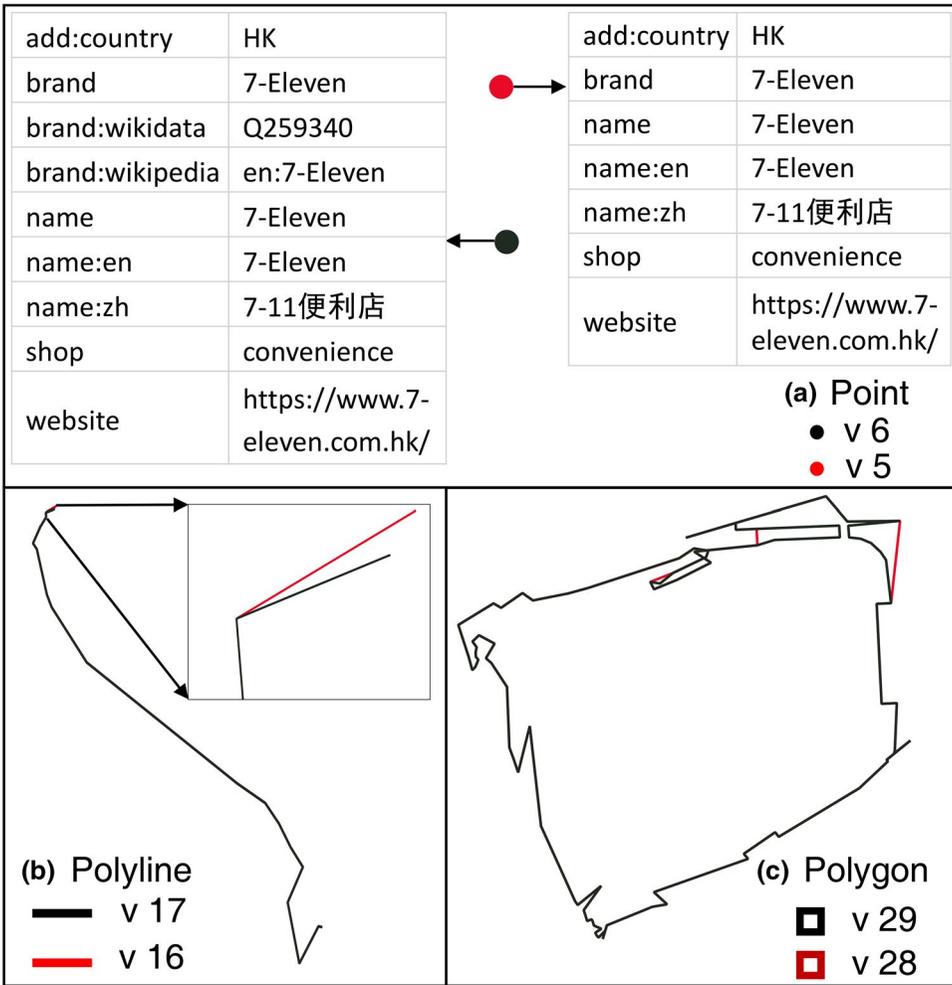


FIGURE 4 Changes on adjacent versions of a point, polyline, and polygon feature in Hong Kong

TABLE 1 Similarity metrics and version similarity for the geographical features in Figure 5

Metrics	Feature type		
	Point	Polyline	Polygon
<i>SimBuff</i>	0.80	0.99	0.63
<i>SimLen</i>	-	0.99	-
<i>SimShape</i>	-	0.99	0.99
<i>SimArea</i>	-	-	0.97
<i>SimTag</i>	0.78	1.00	1.00
<i>VerSim</i>	0.79	0.99	0.89

was small in terms of the whole feature, and its version similarity was close to 1.00. For the polygon, its similarity was equal to the average of the area similarity (0.97), the overlapping rate of the buffer (0.63), and the distance between turning functions (0.99).

The contributors who generated six versions in Figure 4 are introduced in Table 2. Contributor ID 1,724,771 generated the fifth version of the point feature, and contributor 10,130,685 modified it to generate the sixth version. A corresponding link was established for contributor 10,130,685 to point to contributor 1,724,771, denoted as 10,130,685 \rightarrow 1,724,771. The version similarity of the point feature was 0.79, which represented a single evaluation of contributor 1,724,771 by 10,130,685. In the contributor network for HK, however, 10,130,685 modified the edits of 1,724,771 much more than once: 7,561 times (see the frequency in Table 2). According to Equation (1), the evaluation $w_{p,q}$ ($p = 1,724,771, q = 10,130,685$) was the average of multiple evaluations equal to 0.72. In addition, the activity level from 10,130,685 \rightarrow 1,724,771 ($W_{pq}^{in} \cdot W_{pq}^{out}$) was 7.97×10^{-6} in terms of Equations (8) and (9). The evaluations and activity levels of the links for involved contributors in Figures 4b and c are also shown in Table 2. The single evaluation of 1,787,380 \rightarrow 3,691,120 through the polyline was close to 1.00, whereas the average evaluation in HK was 0.56. Moreover, contributor 2,620,510 presented a higher activity level through the evaluation of 8,162,358 \rightarrow 2,620,510.

3.2.2 | Ranking of reputation

There were 5,767 contributors extracted in HK and 4,291 in BJ. The EWPR scores of the contributors were obtained through the proposed method and were divided into 10 classes, 1–10, by the classification of natural breaks (Jenks), giving an R-level that could indicate a ranking from low to high ratings of reputation. The proportion of each class is shown in Figure 5. In both HK and BJ, contributors providing less than 60% were estimated as 1 R-level, showing that most contributors create few edits and are rarely active in the OSM. We observed that the proportion of contributors in each class from 4 to 10 was similar, about 2 or 3%, and the sum of 7–10 R-levels in HK and BJ was about 11%, meaning that a few contributors had high-level reputations in VGI projects.

TABLE 2 C2C evaluations and activity levels among the involved contributors in Figure 5

Feature type	Link of contributors	Frequency	$w_{p,q}$	$W_{pq}^{in} \cdot W_{pq}^{out} (\times 10^{-6})$
Point	10130685 1724771	7,561	0.72	7.97
Polyline	1787380 3691120	24	0.56	9.19
Polygon	8162358 2620510	154	0.84	151.76

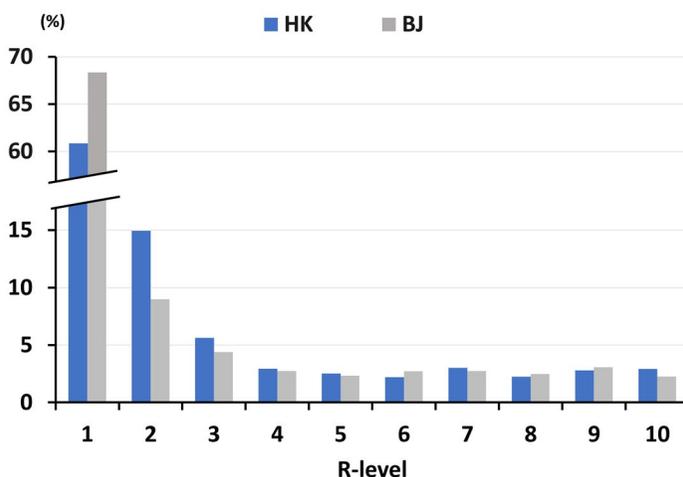


FIGURE 5 The proportion of contributors at the different R-levels

Two study areas were selected to establish two respective networks such that the local reputations of the contributors in HK and BJ were evaluated. There were 785 contributors involved in both HK and BJ. To determine whether the reputation results for the same contributor were the same in different regions, the differences (absolute value) of their corresponding R-levels were computed; the results are shown in Table 3. It was discovered that 335 contributors (43%) were estimated to have the same R-level in both locations, and 60% of contributors had an R-level difference of less than 3. The number of contributors with R-level differences of more than 4 was 167, and among those, 21 contributors reached the largest difference of 9. We found that these large differences were mainly caused by the number of contributors' edits and C2C evaluations in the regions. Table 4 details the information for five contributors whose R-level differences were more than 4. Contributor ID 13,721 had a 3 R-level in HK and 10 in BJ because the contributor was highly active in BJ (18,012 edits) but had only 686 edits in HK. Similarly, the contributors could be evaluated as having a highly rated reputation when they were extremely active in a certain region, such as contributor 2,237,091 with a 10 R-level in HK and contributor 6,149,968 with a 9 R-level in BJ. Conversely, contributor 22,998 participated in only a few edits in HK and BJ but was assigned, respectively, a 10 R-level and a 3 R-level, and contributor 1,666,292 had more edits but a lower R-level in BJ. These demonstrate that when the activity level of a contributor has no obvious advantage, their R-level was more affected by the C2C evaluation. Contributor 22,998's edits were modified by one contributor with 10 R-level in HK and by three contributors with an average R-level of 10 in BJ. The corresponding C2C evaluations were 0.95 and 0.78, respectively. Similarly, contributor 1,666,292's edits in HK were modified by only one contributor whose reputation was assigned a 10 R-level. Meanwhile, contributor 1,666,292 received a high C2C evaluation equal to 0.96. The edits of contributor 1,666,292 in BJ, however, were modified by 15 contributors, with their average R-level equal to 7 and average C2C evaluation equal to 0.84.

3.2.3 | Credibility of OSM data

Based on the R-level, we obtained the expertise and trustworthiness of OSM features using the methods mentioned in Section 2.3. There were 426,401 OSM polylines in HK and 321,335 in BJ. Figure 6 illustrates the correlation between the two kinds of indicators. Here, expertise was highly correlated with trustworthiness, with a Pearson correlation coefficient equal to 0.92 in HK and 0.93 in BJ.

TABLE 3 The proportion of the difference (0, ..., 9) between R-levels in Hong Kong (HK) and Beijing (BJ) for the same contributors

Difference in R-levels	0	1	2	3	4	5	6	7	8	9	Total
Quantity	335	134	65	48	36	44	37	37	28	21	785
Proportion (%)	42.68	17.07	8.28	6.11	4.59	5.61	4.71	4.71	3.57	2.68	100

TABLE 4 Examples of the R-levels and the number of edits for five contributors

User ID	R-level in HK	R-level in BJ	Number of edits in HK	Number of edits in BJ
13721	3	10	686	18,012
22998	10	3	32	15
2237091	10	1	24,608	316
6149968	1	9	1,018	412,911
1666292	10	5	12	203

Abbreviations: BJ, Beijing; HK, Hong Kong.

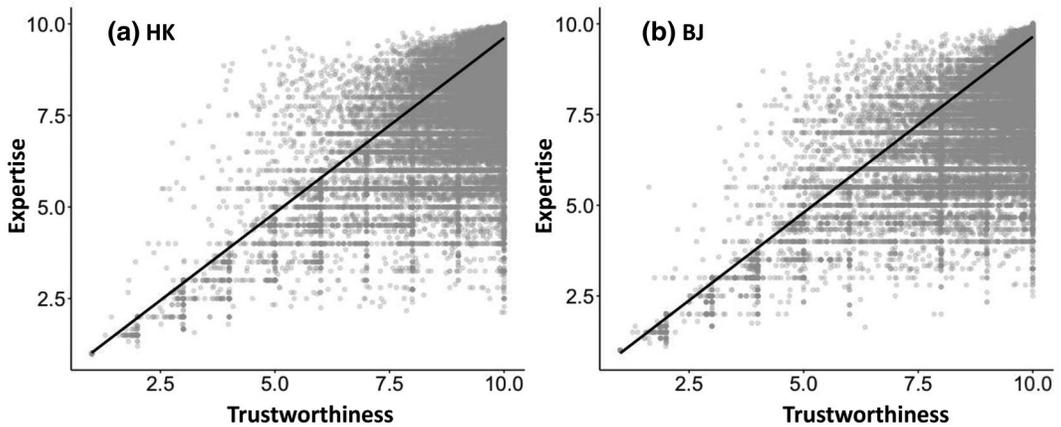


FIGURE 6 A scatterplot of the comparison between expertise and trustworthiness of the polyline features in: (a) HK; and (b) BJ

TABLE 5 The comparison between credibility and quality level for the random samples in HK and BJ

Credibility	HK				BJ			
	Low	Medium	High	Total	Low	Medium	High	Total
1-5	30	2	0	32	62	20	3	85
5-7	7	16	2	25	5	25	2	32
7-9	8	68	35	111	10	88	30	128
9-10	15	170	190	375	26	140	268	434
Total	60	256	227	543	103	273	303	679

Abbreviations: BJ, Beijing; HK, Hong Kong.

Expertise and trustworthiness were averaged as the evaluation of data credibility. To validate the credibility, we randomly selected 1% of OSM roads in HK and BJ, respectively, as the samples. These samples were assigned a quality level that included three categories: low, medium, and high. The quality was mainly judged according to the consistency of the location and the shape of the sample with Google images. Whether the sample had a road name was also considered. A small number of samples were removed because they were difficult to recognize from the images. The comparison between credibility and quality level is summarized in Table 5, where the larger numerical number denotes the higher credibility. Here, 94% (HK) and 73% (BJ) of the roads has less than a 5 in credibility and were evaluated as low quality, and the rest were mainly assigned a ranking of medium quality. Most of the roads that had a credibility between 5 and 9 were in the category of medium quality. The roads with greater than 9 credibility were mainly evaluated as medium or high quality. Generally, the proportion of high-quality roads was increasing as the increase of credibility and corresponding proportion of low-quality roads decreased.

The spatial distribution of the credibility of OSM roads was shown to further validate the credibility. As shown in Figure 7, the study areas were clipped into $1 \times 1 \text{ km}^2$ grids. Based on the roads intersecting with each grid, the credibility and reputation attached to the roads were averaged for each grid. A few grids in HK were colored red, showing that most of the roads in HK were credible. We labeled five regions with clustered red grids representing the low-credible regions [see labels (1)–(5) in Figure 7a]. It was observed that the reputation in these regions was also on the low side (grids in red color; see Figure 7b). In addition, we found some red grids for reputation to appear near the boundaries of the districts in HK, and the districts located in the central areas of HK had positive credibility and reputation. Under the larger regional scope of BJ, there were more low-credibility grids and

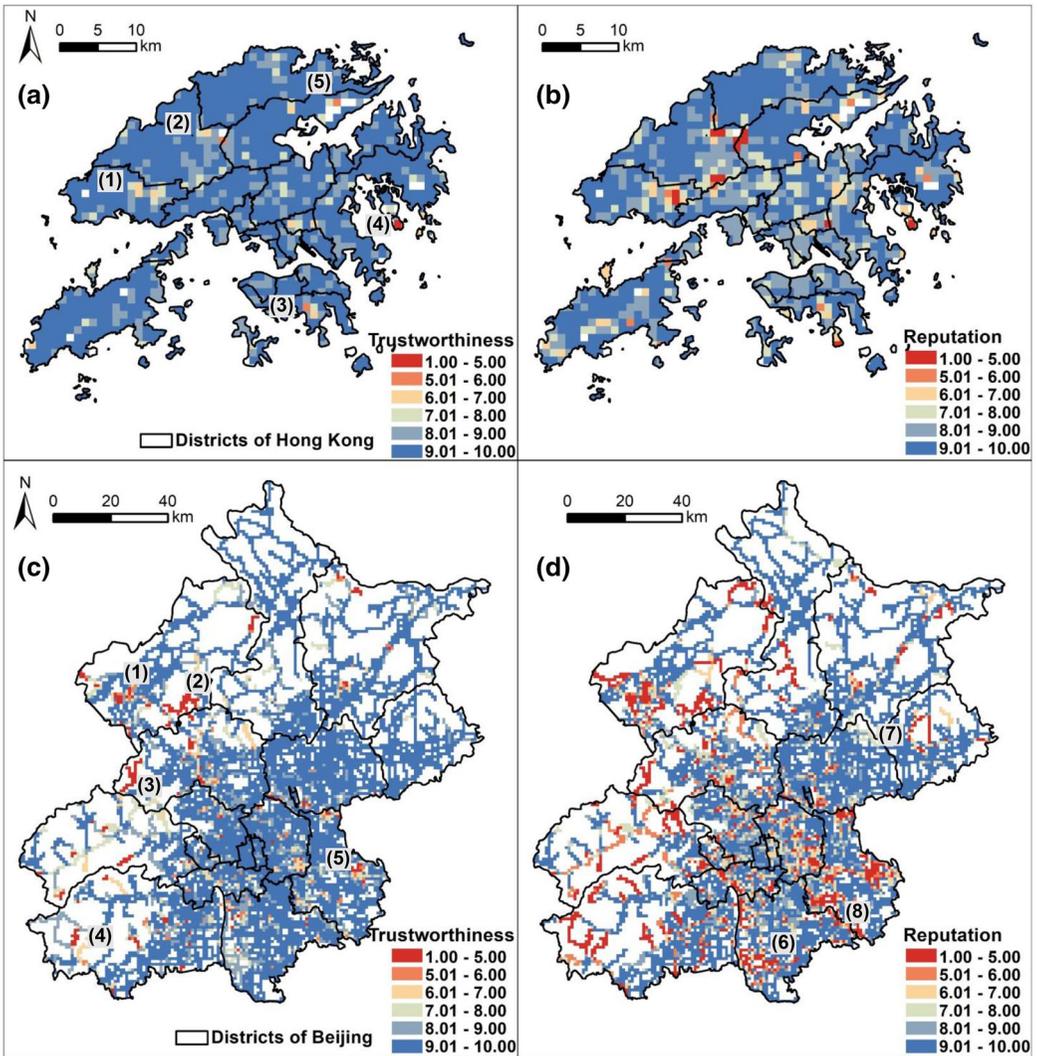


FIGURE 7 The spatial distribution of the average credibility assigned to the contributed roads for the grids in (a) HK, (c) BJ and the spatial distribution of average reputation of the contributors of the roads for the grids in (b) HK, and (d) BJ

more roads that were contributed by low-reputation contributors. Five regions [(1)–(5) in Figure 7c] that had low credibility were also focused on, and the corresponding grids in these regions were evaluated as having a low reputation (less than 6). In addition, the five regions in Figure 7c were located in the suburbs, and the credibility and reputation were generally positive in the northeast of BJ. However, we found some regions, such as regions (6)–(8) in Figure 7d, to have clustered low reputation but acceptable credibility.

To find out what causes the heterogeneity of credibility, we also explored some characteristics of OSM roads. OSM applies a classification system for all roads, and Figure 8 summarizes the average credibility of the different types of roads with an ordered list from the most important (motorway) to the least important (cycleway). Except for the brideway, the credibility of the other roads in HK is greater than 9.00. With the reduced importance of roads, the corresponding credibility tended to decline (see the blue trend line in Figure 8a). In BJ, there was a greater difference in the credibility of the different types of roads than in HK. Trunk, link, primary, secondary, and

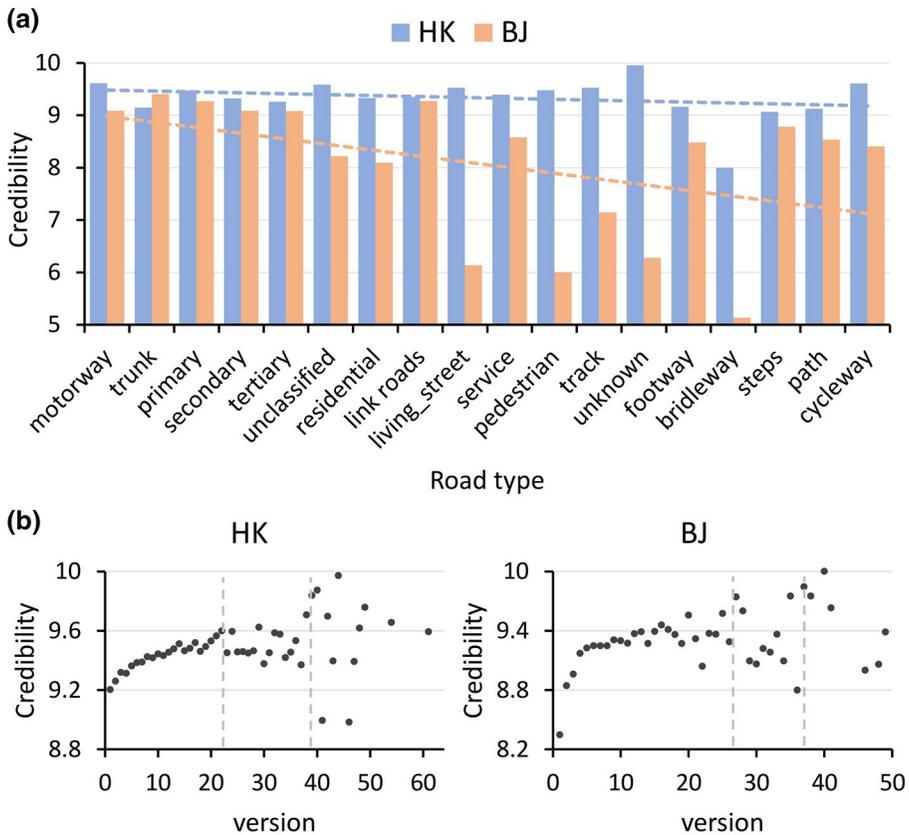


FIGURE 8 The average credibility of the roads for: (a) different road types; and (b) different version numbers

tertiary roads were assigned greater than a 9.00 credibility, and the remainder had a credibility less than 9.00. It was a downward trend for credibility as the importance of roads decreased (see the orange line). The association between credibility and the version number was also analyzed (see Figure 8b). We calculated the average credibility of the roads with the same version number. In HK, credibility enhancement emerged with an increase in the version number when the version number was smaller than about 20. Also, a rough improvement of credibility from version number 20 to 40 was found. Similarly, there was increased credibility from version 1 to about 28 and from 28 to 40 in BJ. However, the rise in the credibility of the roads might disappear when the version number was larger, such as over 40. Another characteristic involved was the road's name. We observed that many OSM roads were not provided with their name. The average credibility of the roads without names in HK was 9.23, and those with names 9.38. Moreover, the roads with names in BJ were assigned an average credibility equal to 9.06, and the rest without names were assigned 8.51 trustworthiness. That is, the credibility of the roads with road names attached was higher than for those with no name.

3.3 | Comparisons with existing methods

3.3.1 | Characteristics-based reputation

The characteristics of 5,688 contributors in HK and 4,219 contributors in BJ were extracted from the following website: <https://hdyc.neis-one.org>. The corresponding characteristics are outlined in Table 6 in terms of practice,

TABLE 6 Descriptive statistics of contributors' characteristics in Hong Kong

Characteristics	Min	Q1	Median	Mean	Q3	Max
Number of mapping days	1	1	3	156.63	51	4,433
Longevity of activity (days)	0	209	382	1,054.19	1,422	5,746
Number of edits	1	6	51.5	268,881	6,642.5	185,176,420
Software skill (level)	0	1	1	1.9	3	8
Local knowledge (ratio)	0	0.01	0.91	0.57	1	1.45

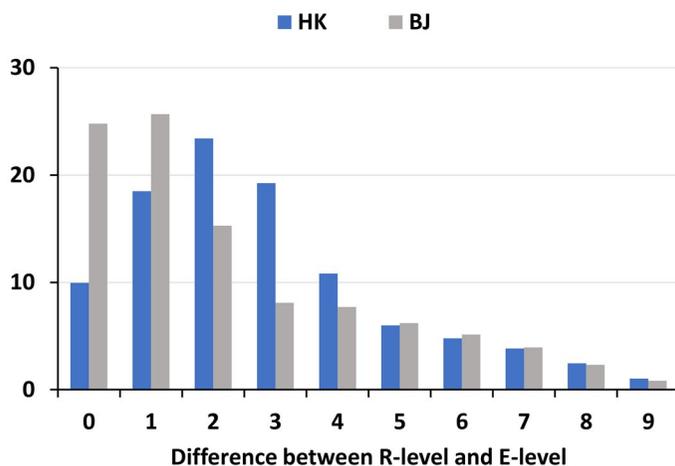


FIGURE 9 The proportion of the difference (0, ..., 9) between the R-level and the E-level for the same contributor. BJ, Beijing; HK, Hong Kong

skill, and local knowledge. For all characteristics, the median and mean values differed considerably, indicating that the data were unevenly distributed. These characteristics were normalized and combined in a weighted sum as the characteristics-based reputation of VGI contributors, wherein the derived weights were calculated based on the correlation between the characteristics and OSM quality (shown by MOL).

The characteristics-based reputation was similarly classified into 10 classes (E-level). The differences between the R-levels and the E-levels were determined for individual contributors (i.e., 0, ..., 9), and their frequency distribution is provided in Figure 9. We hypothesized that the R-level would be consistent with the E-level if their difference were 0 or 1, and 28% of 5,688 contributors in HK and 51% of 4,219 contributors in BJ met this criterion. In addition, a difference greater than 4 was considered inconsistent between the R-level and the E-level, and the contributors of about 18% in both HK and BJ had this result. This suggests there were more contributors rated with consistent ranking for the R-level and the E-level. The contributors in HK and BJ with a 2–4 difference between the R-level and the E-level accounted for a proportion of 54% and 31%, respectively.

To further explore their consistency, we computed the average R-level for the average E-level in each class (see Figure 10). Overall, the R-level and the E-level showed a highly positive correlation, and the R-squared was about 0.86 in both HK and BJ under a linear regression.

As shown in Table 6, there was one contributor with approximately 200 million edits (contributor 7,168). To demonstrate the difference between his evaluations at the E-level and the R-level, we summarize the characteristics of this contributor (Table 7). He registered in 2007 and mainly contributed to OSM in the USA. Owing to the large number of edits, he had a high E-level equal to 9, though other characteristics like poor local knowledge gave no obvious advantages. Using the EWPR algorithm in HK, he had an acceptable activity level equal to 0.75 and received a medium-level C2C evaluation (0.56) from two contributors at the 1 R-level. In EWPR, the evaluations

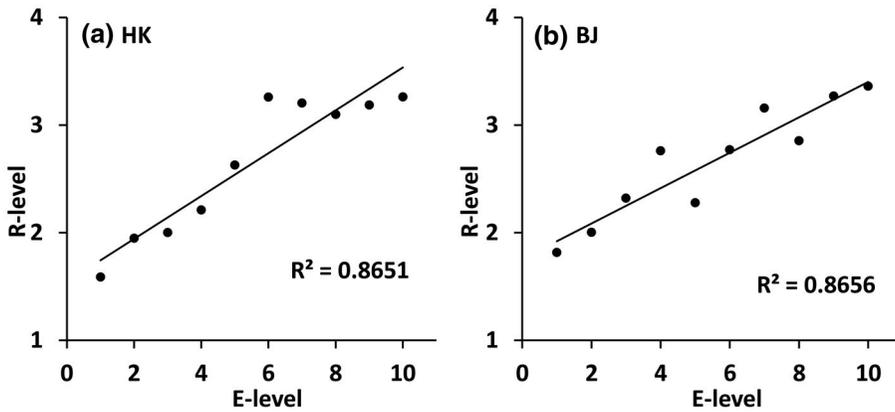


FIGURE 10 A scatterplot of the comparison of the average R-level and average E-level in: (a) HK; and (b) BJ

TABLE 7 The characteristics of the contributor who contributed approximately 200 million edits

Characteristic	Value
Number of mapping days	240
Longevity of activity (days)	999
Number of edits	185,176,420
Software skill (level)	5
Local knowledge (ratio)	1.62×10^{-8}
E-level	9
Average activity level	0.56
Average C2C evaluation	0.75
R-level	7

from low-reputation contributors have little impact on reputation ranking. Thus, his reputation mainly depended on activity level and was equal to 7 R-level.

3.3.2 | Similarity-based reputation

There were 3,228 contributors who were assigned similarity-based reputations in HK and 2,254 contributors in BJ. Their reputations were divided into 10 classes, labeled as the S-level. The differences among R-level, E-level, and S-level for the contributors are shown in Figure 11. In HK, 6% of contributors were rated as having the same R-level and S-level (blue segment), and 11% of contributors' R-levels and S-levels differed by 1 (see orange). There were consistent evaluations of the R-level and the E-level for 22% of contributors, and of the R-level and the S-level for 31% of contributors. In BJ, 18% of contributors were rated as having R-level and S-level with a 0 or 1 difference, whereas 41% of contributors' R-levels and S-levels differed by more than 4. The difference between the E-level and the S-level for 30% of contributors was also greater than 4. This means that the comparisons of the E-level and the S-level as evaluated by published methods showed certain inconsistencies.

The scatterplots of the comparisons among the average S-, R-, and E-levels are similarly drawn in Figure 12. Although their evaluations in each class did not show strong consistency, the R-level showed a highly positive correlation with the S-level on the whole, and the R^2 was 0.90 in HK and 0.81 in BJ. Conversely, it is unexpected

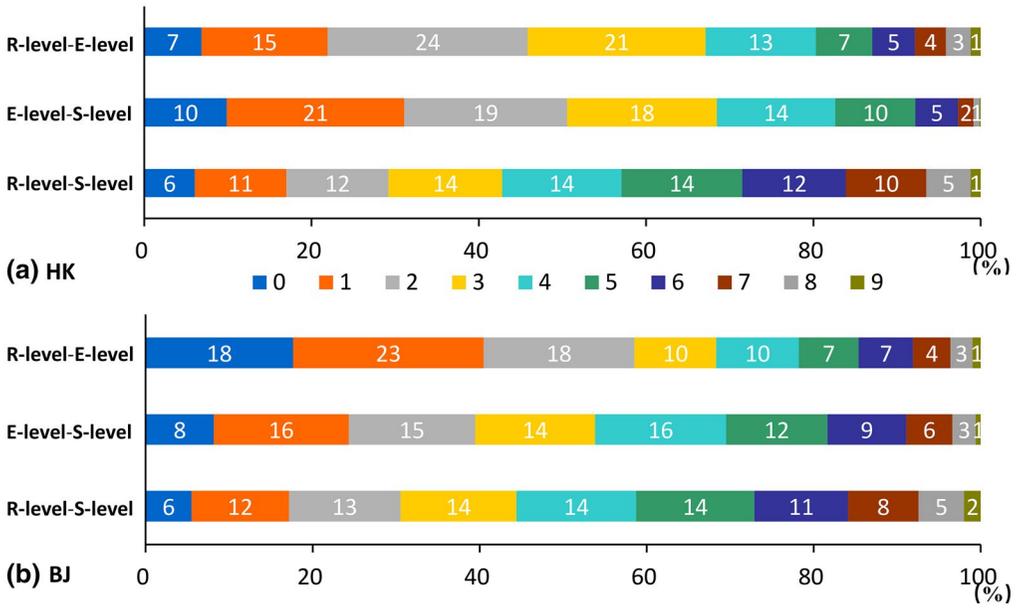


FIGURE 11 The proportion distribution of the differences between the E-level and the S-level, the R-level and the E-level, and the R-level and the S-level in (a) HK and (b) BJ; different color segments represent each difference (0, ..., 9), and the numbers on them represent the corresponding proportion (%) of each difference

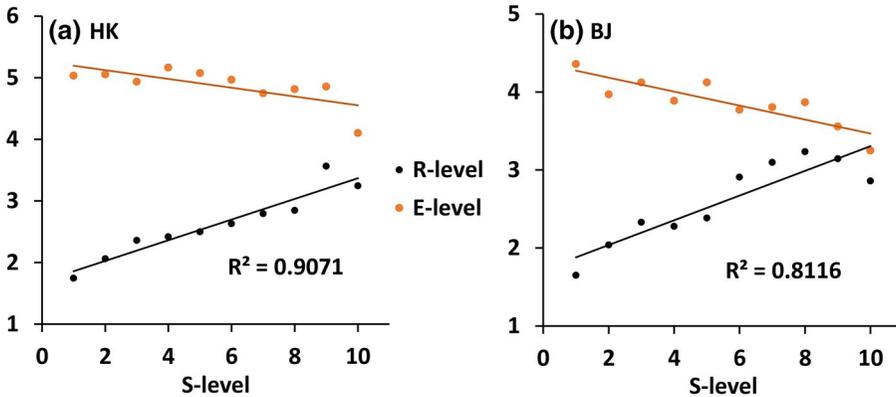


FIGURE 12 A scatterplot of the comparisons among the average S-level, R-level, and E-level in: (a) HK; and (b) BJ

for the E-level and the S-level to be inversely correlated. That is, the average E-level presented the opposite trend, with an increase in the average S-level of each class.

4 | DISCUSSION

4.1 | Measurement of reputation

The novelty of our study is that it adopted a computer algorithm to deal with the activity level of OSM contributors and their evaluation relationships such that a ranking metric for contributor reputation is presented.

Two published methods (the E-level and the S-level) were realized to compare with our results (the R-level), and there are certain consistencies and differences among them. Various methods cannot replace each other; indeed, each has its own advantages and limitations. There are several possibilities for the inconsistencies among the evaluations of the R-, E-, and S-levels. First, reputation is a qualitative indicator without a numerical definition. Accordingly, the three types of levels have uncertainties as the proxy of reputation, and they differ in calculation errors. Moreover, the E-level is computed using the characteristics on a global scale, whereas the R-level and the S-level are on a local scale in the case studies. The S-level only employs the evaluation relationship, and the R-level also uses the activity level. Despite all this, the R-level shows a high degree of consistency and correlation with the E-level and the S-level, confirming that the evaluation of the R-level tends to be credible.

The E-level is obtained based on the characteristics of VGI contributors. It is a reasonable way to measure contributor reputation, because a contributor's characteristics can help in understanding the contributor's behavior and further indicate reputation. However, the kind of method cannot be implemented without available data about the contributor's characteristics. For example, there are 5,767 total contributors in HK, whereas 5,688 contributors have access to characteristics information. Likewise, the distribution of these characteristics is uneven, as shown by the number of edits in Table 6, where the minimum is 1, the maximum is 185,176,420, and the median is 51.5. In this case, data normalization and weight setting are sensitive when various characteristics are incorporated into the reputation.

The S-level can be extracted for fewer contributors (3,228 in HK) and is not applicable to contributors who are less active in the community. In our study, the C2C evaluation describes the evaluation relationship between adjacent contributors who co-edit an OSM object. When calculating the S-level, however, there is an evaluation relationship between multiple continuous contributors co-editing an object. For instance, in addition to the C2C evaluation among the three contributors in Figure 1, the S-level also involves the evaluation of u_{ij} by u_{ij+2} . In essence, the evaluation of u_{ij} by u_{ij+2} partially contains the evaluation of u_{ij} by u_{ij+1} . This reuses the information. Furthermore, the model for calculating the S-level failed to consider contributors' activity level.

The EWPR algorithm has four advantages in measuring reputation. First, it can deal with all contributors within the OSM community, including the contributors who have only edited once. Second, the evaluation relationship and the activity level can simultaneously be used to understand contributor reputation. Within the community, the contributors with a higher activity level tend to have more practice and skills. Third, as of July 31, 2020, the number of OSM-registered users has exceeded seven million, and the number of changesets in 2013 has exceeded five million. A computer algorithm is automatic and efficient in dealing with a large number of contributors within a VGI community. Finally, although the proposed method was validated using case studies in two metropolises—HK and BJ—PageRank also works for the contributors in other areas, even rural regions. The algorithm was mainly achieved using the co-edits between VGI contributors; this means that it can work when there are collaborations among the contributors in a region. It is essential for VGI contributors to correct and complete data together, even in rural regions. Thus, the proposed method can be transferred to other projects maintained through the collaborative editing of the public.

The works on reputation measurement are beneficial to the assurance and assessment of VGI quality, and they mainly include the following points. (a) If the evaluations (feedbacks) from others are exposed to the VGI community, it is useful for the contributors to raise awareness of quality assurance to receive more positive feedbacks. (b) Reputation ranking helps pick out the reliable contributors and the questionable contributors, such that the former can be the moderators or gatekeepers within the community and the latter will attract more attention and corrections. These also help promote the quality and development of VGI projects. (c) Contributors with high-level reputations are believed to contribute more reliable data. Thus, reputation is helpful in identifying data quality and can be used as a factor to assess VGI quality.

The EWPR algorithm is based on an indirect evaluation relationship between contributors and activity level. This helps to distinguish different types of contributors. For instance, corporate editors are more likely to have a high activity level and to receive positive evaluations within a VGI community than novices. Novices, in contrast, are less active and may receive less positive evaluations owing to limited practice and skills. Hobbyists may also

be active but they will receive middle-level evaluations. The users who have similar activity levels and conduct similar evaluations to others may be the bots. Nevertheless, the contributors edit OSM data in complex ways and for various reasons (Rehrl et al., 2013). Our contributor network was established based on the change of adjacent edits and has a limited ability in classifying the different types of editing. Therefore, the network is not able to distinguish whether a change belongs to completing, correcting, or reusing a contribution. Several studies mentioned in the Introduction have established the networks depicting different types of editing. Our study aims to explore efficient, network-based algorithms to address the assessments of the users within a VGI community. The PageRank algorithm—as one of the most popular node-ranking algorithms—is used for this. Future work will apply other, optimized node-ranking algorithms, and focus on the networks that consist of various editing types or multiple relationships between VGI contributors.

Future research may also find a more comprehensive method to visualize the evaluation relationship between VGI contributors. In our study, this relationship is quantified by the version similarity of geographical features, using two thresholds. There may be better ways to measure this relationship. For example, it may be affected by a calculation error in the spatial similarity measure if adjacent versions suffer image registration problems; also, use of similarity and temporal thresholds may miss some state changes of geographical features. As different contributor networks are established in different areas, reputation rankings may be diverse within those networks: a local network emphasizes local knowledge and experience. For instance, contributor 13,721, mentioned in Section 4.2, may be more reliable in BJ than in HK. Future studies will create a network of VGI contributors in China or globally to yield a contributor reputation database for real-time updating.

4.2 | Use of reputation to assess VGI quality

In previous studies, reputation has been used to obtain the indicators of VGI quality, such as expertise, trustworthiness, and credibility. This shows that reputation is highly correlated with VGI quality. Using two published methods, we also explored the credibility of OSM data based on the proposed reputation, and validated the credibility through visual judgment and analyses of the characteristics of OSM data. The results show that our ranking metric can successfully be used to assess VGI quality.

The poor-credibility regions tend to have low reputation (see Figure 7), whereas VGI quality was not entirely dependent on reputation [see regions (6)–(8) in Figure 7d]. In other words, VGI quality is also related to other factors. As Figure 8a shows, the credibility showed a slight correlation with road type, that is, the main roads had higher credibility than the minor roads. In addition, credibility increased with the increase of the version number in a certain range (see Figure 8b), and the roads assigned their names were more credible than those without names. Another related characteristic was the type of land crossed by the roads; here, the roads in regions (1)–(5) in Figure 7c mainly passed through the areas of mountains, forests, and arable land.

Based upon the above characteristics, we next discuss the possibility of inconsistency between reputation and credibility. In central districts in BJ, there may be a larger variety of contributors to edit OSM data; thus, there is likely to be more heterogeneity in their reputations (see Figure 7d). Covering more main roads, data in the central districts attracted more attention and had greater version numbers, such that their credibility becomes acceptable or high. In rural areas [see regions (1)–(5) in Figure 7c], there is a body of mixed use of arable land, woodland, and built-up land, meaning that a road may pass through mountains, forests, arable land, and buildings at the same time. This increased the difficulty in mapping the road in OSM. In addition, fewer contributors were dedicated to editing the roads in rural areas, and some roads may have been contributed several years ago and may come with less attribute information (e.g., road type and name). These indicate that the roads in rural areas are more likely to be of a low quality than those in urban areas. In some rural areas, however, the roads were mainly contributed by a few high-reputation contributors; that is, the high-reputation contributors generated the roads without messy contributors to edit. Thus, the roads in these areas had high credibility (see the northeast of BJ).

VGI quality is related to a range of factors in addition to reputation, and how to combine reputation and other characteristics for quality assessment remains challenging. Some studies have proposed various mathematical relations between the indicators of VGI quality and reputation. In addition (see Table 5), the visual quality was mainly evaluated based on positional accuracy and shape accuracy, whereas the quantitative relationships between the two extrinsic indicators and credibility/reputation are unknown. Hence, future studies need to figure out more definite relationships between reputation and quality. The methods for reputation measurement are worth studying before this research on VGI quality can take place.

5 | CONCLUSIONS

In this study, we established a network that depicts evaluation relationships among VGI contributors. It includes past edits and the activity levels of the contributors through the number of produced edits. Based on this network, we developed an evaluation-based weighted PageRank (EWPR) algorithm to measure reputation, with the aim of providing an available indicator of VGI quality assessment or assurance. Comparing our algorithm with two existing methods on two OSM datasets, we show that the algorithm is effective. Based on reputation, we obtained the credibility of OSM data, demonstrating the successful implementation of using reputation to assess aspects of VGI quality. The analyses on data credibility also indirectly validate the proposed reputation and show that reputation is useful for dealing with quality variability in VGI.

Reputation is a qualitative indicator of VGI quality and in principle can be used widely as a tool to ensure the validity of VGI providers. Before doing so, the study shows that it is useful to develop an algorithm like in our study. Its effectiveness may indicate several new perspectives on measuring reputation. Such an algorithm, often used for understanding relationships between web pages and social network participants, has the potential for reputation ranking among VGI contributors. This study therefore fills the gap that network-based technology rarely addresses evaluation relationships and contributor characteristics at the same time.

ORCID

Die Zhang  <https://orcid.org/0000-0003-0255-6814>

REFERENCES

- Adler, B. T., de Alfaro, L., Kulshreshtha, A., & Pye, I. (2011). Reputation systems for open collaboration. *Communications of the ACM*, 54(8), 81–87. <https://doi.org/10.1145/1978542.1978560>
- Antoniou, V., & Skopeliti, A. (2015). Measures and indicators of VGI quality: An overview. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, 1, 345–351. <https://doi.org/10.5194/isprsannals-II-3-W5-345-2015>
- Arkin, E. M., Chew, L. P., Huttenlocher, D. P., Kedem, K., & Mitchell, J. S. (1991). An efficiently computable metric for comparing polygonal shapes. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 13(3), 209–216. <https://doi.org/10.1109/34.75509>
- Bishr, M., & Janowicz, K. (2010). Can we trust information? The case of volunteered geographic information. In *Proceedings of the Towards Digital Earth: Search, Discover and Share Geospatial Data Workshop at the Future Internet Symposium*, Berlin, Germany.
- Bishr, M., & Kuhn, W. (2013). Trust and reputation models for quality assessment of human sensor observations. In T. Tenbrink, J. Stell, A. Galton, & Z. Wood (Eds.), *Spatial information theory* (Lecture Notes in Computer Science (Vol. 8116, pp. 53–73). Cham, Switzerland: Springer.
- Box, G. E., & Meyer, R. D. (1986). An analysis for unreplicated fractional factorials. *Technometrics*, 28(1), 11–18. <https://doi.org/10.1080/00401706.1986.10488093>
- Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer Networks & ISDN Systems*, 30(1), 107–117. [https://doi.org/10.1016/S0169-7552\(98\)00110-X](https://doi.org/10.1016/S0169-7552(98)00110-X)
- Degrossi, L. C., de Albuquerque, J. P., Rocha, R. D., & Zipf, A. (2018). A taxonomy of quality assessment methods for volunteered and crowdsourced geographic information. *Transactions in GIS*, 22, 542–560. <https://doi.org/10.1111/tgis.12329>

- Fan, H. C., Zipf, A., Fu, Q., & Neis, P. (2014). Quality assessment for building footprints data on OpenStreetMap. *International Journal of Geographical Information Science*, 28, 700–719. <https://doi.org/10.1080/13658816.2013.867495>
- Flanagin, A. J., & Metzger, M. J. (2008). The credibility of volunteered geographic information. *GeoJournal*, 72(3/4), 137–148. <https://doi.org/10.1007/s10708-008-9188-y>
- Fogliaroni, P., D'Antonio, F., & Clementini, E. (2018). Data trustworthiness and user reputation as indicators of VGI quality. *Geo-Spatial Information Science*, 21(3), 213–233. <https://doi.org/10.1080/10095020.2018.1496556>
- Goodchild, M. F., & Li, L. N. (2012). Assuring the quality of volunteered geographic information. *Spatial Statistics*, 1, 110–120. <https://doi.org/10.1016/j.jspasta.2012.03.002>
- Gusmini, M., Jabeur, N., Karam, R., Melchiori, M., & Renso, C. (2017). Reputation evaluation of georeferenced data for crowd-sensed applications. *Procedia Computer Science*, 109, 656–663. <https://doi.org/10.1016/j.procs.2017.05.372>
- Haklay, M. (2010). How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment & Planning B*, 37, 682–703. <https://doi.org/10.1068/b35097>
- Hovland, C. I., Janis, I. L., & Kelley, H. H. (1953). *Communication and persuasion: Psychological studies of opinion change*. New Haven, CT: Yale University Press.
- Jackson, S. P., Mullen, W., Agouris, P., Crooks, A., Croitoru, A., & Stefanidis, A. (2013). Assessing completeness and spatial error of features in volunteered geographic information. *ISPRS International Journal of Geo-Information*, 2, 507–530. <https://doi.org/10.3390/ijgi2020507>
- Lodigiani, C., & Melchiori, M. (2016). A PageRank-based reputation model for VGI data. *Procedia Computer Science*, 98, 566–571. <https://doi.org/10.1016/j.procs.2016.09.088>
- Maué, P. (2007). Reputation as tool to ensure validity of VGI. *Paper presented at the VGI Specialist Meeting*, Santa Barbara, CA.
- Mooney, P., & Corcoran, P. (2014). Analysis of interaction and co-editing patterns amongst OpenStreetMap contributors. *Transactions in GIS*, 18, 633–659. <https://doi.org/10.1111/tgis.12051>
- Muttaqien, B. I., Ostermann, F. O., & Lemmens, R. L. G. (2018). Modeling aggregated expertise of user contributions to assess the credibility of OpenStreetMap features. *Transactions in GIS*, 22, 823–841. <https://doi.org/10.1111/tgis.12454>
- Rehrl, K., Gröechenig, S., Hochmair, H., Leitinger, S., Steinmann, R., & Wagner, A. (2013). A conceptual model for analyzing contribution patterns in the context of VGI. In J. Krisp (Ed.), *Progress in location-based services* (Lecture Notes in Geoinformation and Cartography, pp. 373–388). Berlin, Germany: Springer.
- Severinsen, J., de Roiste, M., Reitsma, F., & Hartato, E. (2019). VGTrust: Measuring trust for volunteered geographic information. *International Journal of Geographical Information Science*, 33, 1683–1701. <https://doi.org/10.1080/13658816.2019.1572893>
- Stein, K., Kremer, D., & Schlieder, C. (2015). Spatial collaboration networks of OpenStreetMap. In J. A. Jamal, Z. Alexander, M. Peter, & H. Marco (Eds.), *OpenStreetMap in GIScience* (pp. 167–186). Cham, Switzerland: Springer.
- Truong, Q. T., de Runz, C., & Touya, G. (2018). Analysis of collaboration networks in OpenStreetMap through weighted social multigraph mining. *International Journal of Geographical Information Science*, 33, 1651–1682.
- Xing, W., & Ghorbani, A. (2004). Weighted PageRank algorithm. In *Proceedings of the Second Annual Conference on Communication Networks and Services Research*, Fredericton, NB, Canada (pp. 305–314). Piscataway, NJ: IEEE.
- Zhang, W.-B., Leung, Y., & Ma, J.-H. (2019). Analysis of positional uncertainty of road networks in volunteered geographic information with a statistically defined buffer-zone method. *International Journal of Geographical Information Science*, 33, 1807–1828. <https://doi.org/10.1080/13658816.2019.1606430>
- Zhou, X., & Zhao, Y. (2016). A version-similarity based trust degree computation model for crowdsourcing geographic data. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 41, 327–333.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Zhang D, Ge Y, Stein A, Zhang W-B. Ranking of VGI contributor reputation using an evaluation-based weighted pagerank. *Transactions in GIS*. 2021;00:1–21. <https://doi.org/10.1111/tgis.12735>